OPTIMAL DECENTRALIZED CONTROL IN A MULTI-ACCESS CHANNEL WITH PARTIAL INFORMATION

by

Zvi Rosberg

Technical Report #219

September 1981
OPTIMAL DECENTRALIZED CONTROL IN A MULTI-ACCESS CHANNEL WITH PARTIAL INFORMATION

by

Zvi Rosberg

Abstract

We consider two transmission stations sharing a single communication channel. For different values of the input message rates $r_i$, $i = 1, 2$, a simple open-loop control policy is shown to be optimal for the long-run average throughput criterion.
1. INTRODUCTION

We consider I transmission stations sharing a single communication channel, subject to the following dynamic system:

\[
X_i(t+1) = V_i(t) + X_i(t)(1 - V_i(t))(1 - u_i(t)), \quad 1 \leq i \leq I, \quad (1.1)
\]

where \( \{X_i(0), V_i(t) \mid 1 \leq i \leq I, t = 0, 1, 2, \ldots \} \) are independent Bernoulli r.v.'s and

\[
P(V_i(t) = 1) = P(X_i(0) = 1) = r_i, \quad 0 \leq r_i \leq 1. \quad (1.2)
\]

The channel is assumed to be slotted; that is, the channel time is divided into equal segments called slots. The length of each slot equals that of a packet. Furthermore, it is assumed that the stations are synchronized with the channel in the sense that attempts to transmit packets are exactly time-aligned with the slots.

Every station \( i \) contains a buffer which can store one packet; \( X_i(t) = 1 \) or \( 0 \) depending on whether the buffer is full or empty at slot \( t \); \( V_i(t) = 1 \) or \( 0 \) depending on whether a new packet does or does not arrive at slot \( t \). The buffer content at time \( t \) is either transmitted or not, via the common channel, depending on whether \( u_i(t) = 1 \) or \( 0 \).

Let \( W_i(t) = 1 \) or \( 0 \), depending on whether a packet is transmitted from station \( i \) at slot \( t \), or not. We have

\[
W_i(t) = X_i(t)u_i(t).
\]

Every station is listening to the channel, so it can detect whether zero, one or more than one packets are transmitted at each slot. Let \( W(t) \) be equal to 0, 1 or 2 according to these three possibilities.
Furthermore, let

\[ W_t^{-1} = \{ W(s) | 0 \leq s \leq t-1 \} \]

and

\[ u_t^{-1} = \{ u_i(s) | 1 \leq i \leq I, 1 \leq s \leq t-1 \}. \]

The control action of station \( i \) at slot \( t \), \( u_i(t) \), can depend on the common information \( W_t^{-1} \) and \( u_t^{-1} \), that is

\[ u_i(t) = u_i(W_t^{-1}, u_t^{-1}). \]

Note that \( u_t^{-1} \) is a common information, since the control policy is known by all the stations.

Let the immediate reward at slot \( t \), \( r(t) \), be the number of packets which have been successfully transmitted during the slot. We have

\[ r(t) = r(X_1(t), \ldots, X_i(t), u_1(t), \ldots, u_i(t)) = \sum_{i=1}^I X_i(t)u_i(t) \prod_{j \neq i} (1 - X_j(t)u_j(t)). \quad (1.3) \]

For every admissible control policy \( \pi \), let \( V_T(\pi) \) be the total expected number of packets which were successfully transmitted during the first \( T \) slots (hereinafter \( T \) steps) using policy \( \pi \). We have

\[ V_T(\pi) = E_{\pi} \sum_{t=0}^{T-1} r(t) = E_{\pi} \sum_{t=0}^{T-1} E( r(t)/W_t^{-1}, u_t^{-1} ), \quad (1.4) \]

where \( E( r(t)/W_t^{-1}, u_t^{-1} ) \) is the expected immediate reward at step \( t \) under policy \( \pi \).

Also, let

\[ \bar{V}(\pi) = \lim_{T \to \infty} \inf V_T(\pi)/T. \quad (1.5) \]
When the limit in (1.5) exists, $\bar{V}(\pi)$ is the long-run average number of packets per slot which have been successfully transmitted using policy $\pi$; $\bar{V}(\pi)$ is also, by definition, the throughput of the channel. Finally, let

$$\bar{V} = \sup_{\pi} \bar{V}(\pi).$$

(1.6)

$\bar{V}$ is called the value function.

We say that a control policy $\pi^*$ is optimal for the long-run average reward criterion if $\bar{V}(\pi^*) = \bar{V}$. The goal of this paper is to find $\pi^*$.

The model given above is closely related to the model first presented and analyzed in Schoute [3] and later in Varaiya [4]. However, it differs from Schoute's model in three main aspects: (i) In Schoute's model the probabilities $r_i$ are all equal to some $r$. (ii) In our model, the control action of station $i$ at slot $t$ is a function of $(w_t^{t-1}, u_t^{t-1})$, whereas in Schoute's model it is a function of $\{X_j(s) | 0 \leq s \leq t-d, 1 \leq j \leq I\} \cup \{X_i(s) | t-d \leq s \leq t\}$, where $d$ is a fixed delay. (iii) In our model the system gains a fixed reward for every successful transmission. In Schoute's model the system pays two different fines at any slot depending on whether a collision occurs or a packet is blocked in the buffer.

For additional discussion on Schoute's model, the reader is referred to Schoute [3]. Note that in both models, packets which collide are lost or stored separately for later independent retransmission. So, no advantage is taken from the information that one might obtain from a collision. A similar model, which takes the information obtained from a collision into account, is given in Behan [1]. Based on a conjecture there, the set of all policies are reduced to several types of policies.
Then, numerically solving the optimality equations of the dynamic programming problem, for different values of $r_i$, $i = 1,2$ (using a computer program), the optimal policies are obtained.

In Schoute [3], it is shown that two open-loop policies are of special interest: (i) The policy which permits all the stations to transmit every slot ("all together"). (ii) The policy which permits a station $i$, $1 \leq i \leq I$, to transmit at slot $t$, if and only if, $I$ divides $|t-i|$ ("round-robin"). In Varaiya [4] it is shown that when $I = 2$, the "all together" ("round-robin") is optimal if $r < c$ $(r \geq c)$, where $c$ is a given constant. For $I > 2$, it is shown that these two particular policies are optimal for certain intervals of $r$.

In this paper we consider only the case $I = 2$. In Section 2 we formulate the dynamic programming model and in Section 3 we derive the optimal control policy for the long-run average throughput. We show that the optimal control does not permit the two stations to transmit at the same time. Moreover, the optimal policy permits one of the stations to transmit every $k$ slots and the other to transmit at the rest of the slots.
2. A DYNAMIC PROGRAMMING PROBLEM

First note that for $I = 2$, the information $\mathbf{W}^{t-1}$ is equivalent to the information $\{\mathbf{W}_1(s), \mathbf{W}_2(s) | 1 \leq s \leq t-1\}$.

Let $k_i(t) - 1$, $i = 1, 2$, be the number of steps elapsed until step $t$ since station $i$ has been last permitted to transmit. Define $k_i(0) = 1$, $i = 1, 2$.

We have

$$k_i(t+1) = 1 + k_i(t)(1-u_i(t)), \quad i = 1, 2. \quad (2.1)$$

**Lemma 2.1:** Given $\mathbf{W}^{t-1}$, $u^{t-1}$ and $u(t) = (u_1(t), u_2(t))$; $X_1(t)$ and $X_2(t)$ are independent r.v.'s.

**Proof:** The lemma follows from (1.1) and (1.2) by a standard induction on $t$. \(\square\)

**Lemma 2.2:** $P(X_i(t) = 1|\mathbf{W}^{t-1}, u^{t-1}) = 1 - (1-r_i)^{k_i(t)}$, $i = 1, 2$.

**Proof:** From (1.1) and (1.2) it follows that $X_i(t) = 1$ if and only if at least one new packet has been arrived since station $i$ was last permitted to transmit. From Lemma 2.1 it follows that given $(\mathbf{W}^{t-1}, u^{t-1})$, the probability of this event is $1 - (1-r_i)^{k_i(t)}$. \(\square\)

Let

$$p_i(k) = 1 - (1-r_i)^k. \quad (2.2)$$

From (1.3), (1.4), Lemmas 2.1 and 2.2 we have

$$E_\pi(r(t)|\mathbf{W}^{t-1}, u^{t-1}) = p_1(k_1(t))u_1(t) + p_2(k_2(t))u_2(t) - 2p_1(k_1(t))p_2(k_2(t))u_1(t)u_2(t), \quad (2.3)$$

where $u_i(t)$, $i = 1, 2$, are the control actions taken by policy $\pi$ given $(\mathbf{W}^{t-1}, u^{t-1})$. 
Now, in order to define a dynamic programming problem we have to specify the state space $S$, the action space $A = \bigtimes_{s \in S} A_s$, the law of motion $q$ and the reward function $\hat{r}$.

For every $t$, $X(t) = (X_1(t), X_2(t))$ is a r.v. whose probability distribution depends on the information $(W_{t-1}, u_{t-1})$. Had $X(t)$ been known, it could serve as a state. However, $X(t)$ is a r.v., so the best one can do is to take as the state its probability distribution given $(W_{t-1}, u_{t-1})$. From Lemmas 2.1 and 2.2, this distribution is completely defined by the parameters $k(t) = (k_1(t), k_2(t))$. Therefore, the state space is

$$S = \{k = (k_1, k_2) | k_i = 1, 2, \ldots, i = 1, 2\}$$

and the action space at state $s \in S$ is

$$A_s = \{u = (u_1, u_2) | u_i = 0, 1, i = 1, 2\}.$$ 

From (2.1), the law of motion becomes:

$$q((1,1) | (k_1, k_2), u = (1,1)) = 1,$$

$$q((k_1+1, k_2+1) | (k_1, k_2), u = (0,0)) = 1,$$

$$q((k_1+1, k_2) | (k_1, k_2), u = (0,1)) = 1,$$

$$q((k_1, k_2+1) | (k_1, k_2), u = (1,0)) = 1,$$

and

$$q(\cdot | (k_1, k_2), u) = 0 \text{ otherwise. }$$

(2.4)

Note that the law of motion is deterministic. Finally, from (2.3) the expected immediate reward is

$$\hat{r}(k, u) = p_1(k_1)u_1 + p_2(k_2)u_2 - 2p_1(k_1)p_2(k_2)u_1u_2.$$ 

(2.5)
Remark 2.1: One could take \((w^{t-1}, u^{t-1})\) as the state at step \(t\).

However, it can be shown that if the probability distribution of \(X(t)\) given \((w^{t-1}, u^{t-1})\) is also used as information for the decision makers, \((w^{t-1}, u^{t-1})\) becomes superfluous when the optimal policy is considered. (This subject is beyond the scope of this paper.)
3. THE OPTIMAL CONTROL POLICY

In this section, it will be first shown that an optimal stationary policy, \( \pi^* \), exists; then it will explicitly be found.

Since for \( r_i = 0 \) or \( 1 \) the optimal control is trivial, we assume that \( 0 < r_i < 1 \), \( i = 1,2 \).

Under any stationary control policy the process \( k(t) = (k_1(t), k_2(t)), \) \( t = 0,1,2, \ldots \), is a Markov chain and under any non-randomized stationary policy the transitions of the chain are deterministic.

Let \( \pi \) be a non-randomized stationary policy under which the chain \( k(t) \) is ergodic. Define the regenerative cycle of \( \pi \) as the shortest sequence of states \( k_0, k_1, \ldots, k_\ell \), such that \( k_0 = k_\ell \).

Clearly, \( \pi \) defines a regenerative cycle and vice versa.

Consider the following types of policies and regenerative cycles which will play a role in the proofs below.

(i) The policy \( \pi_1(k_1, k_2), k_1 \geq 2, k_2 \geq 1 \), which is defined by the regenerative cycle given in Figure 3.1.

\[
\begin{align*}
(1,1) & \xrightarrow{1,0} (1,2) \xrightarrow{1,0} \ldots \xrightarrow{1,0} (1,k_2) \\
1,1 & \uparrow \\
(1,1) & \xleftarrow{0,1} (3,1) \xleftarrow{0,1} (2,1)
\end{align*}
\]

Figure 3.1

The notation \((n_1, n_2) \xrightarrow{u_1, u_2} (n_1', n_2')\) in Figure 3.1 means that at state \((n_1, n_2)\), the control action is \(u_1, u_2\) after which the system moves on to state \((n_1', n_2')\).

(ii) The policy \( \pi_2(k_1, k_2), k_1 \geq 1, k_2 \geq 2 \), which is defined by the regenerative cycle given in Figure 3.2.
(iii) The policy $\pi(1,1)$, which is defined by Figure 3.3.

\[ \begin{array}{c}
(1,1) \rightarrow \ldots \rightarrow (1,3) \rightarrow (1,0) \\
1,0 \uparrow \\
(k_1,1) \leftarrow 0,1 \rightarrow (1,1)
\end{array} \]

**Figure 3.2**

(iv) The policy $\pi(k_1, k_2)$, $k_i \geq 2$, $i = 1, 2$, which is defined by Figure 3.4.

\[ \begin{array}{c}
(1,2) \rightarrow (1,0) \rightarrow (1,3) \rightarrow (1,0) \rightarrow \ldots \rightarrow (1,k_2) \\
1,0 \uparrow \\
(k_1,1) \leftarrow 0,1 \rightarrow \ldots \rightarrow (3,1) \rightarrow (2,1)
\end{array} \]

**Figure 3.4**

**Remark 3.1:** From the mean ergodic theorem it follows that for every policy $\pi$ above, $\bar{V}(\pi)$ is the average reward during one regenerative cycle.

In Theorem 3.1 below, we show that an optimal stationary policy exists.

**Lemma 3.1:** There exists a policy $\pi_0$, which takes a finite number of consecutive control actions $(0,u_2)$ and a finite number of consecutive control action $(u_1,0)$, $u_1 \neq 0,1$, such that $\bar{V}(\pi_0) > \bar{V}(\pi)$. 
for every $\pi$ which takes infinitely many consecutive control actions with $u_1 = 0$ ($u_2 = 0$).

**Proof:** Suppose $r_2 \geq r_1$. Let $k_1^0$ be the first integer such that $p_1(k_1^0) > r_2$ and let $\pi_0 = \pi(k_1^0, 2)$.

From Remark 3.1 and the Definition of $\pi(k_1, k_2)$ in Figure 3.4 we have

$$\bar{V}(\pi_0) = ((k_1^0 - 2)r_2 + p_1(k_1^0) + p_2(2))/k_1^0 > \max\{r_1, r_2\}.$$  \hfill (3.1)

Now, let $\pi$ be a policy which takes infinitely many consecutive control actions, say $(0, u_2)$. From (3.1), (1.5), (2.2) and (2.5) it follows that

$$\bar{V}(\pi) = \bar{r}_2 < \bar{V}(\pi_0).$$

For the control actions $(u_1, 0)$ we have

$$\bar{V}(\pi) \leq r_1 < \bar{V}(\pi_0).$$

\[ \Box \]

**Corollary 3.1:** For every $(r_1, r_2)$, there exists a dynamic programming problem (d.p.p.) with finite state and action spaces, such that its value function, $\bar{V}$, is not smaller than $\bar{V}$ - the value function of the d.p.p. in Section 2.

**Proof:** From Lemma 3.1 it follows that there exist integers $K_1, K_2$, such that any policy which takes more than $K_1 (K_2)$ consecutive control actions $(0, u_2) ((u_1, 0))$ is inferior to some policy $\pi_0$ which does not have this property. Therefore, for finding the optimal control policy, we can consider only policies which take at most $K_1 (K_2)$ consecutive control actions $(0, u_2) ((u_1, 0))$.

Since $k(0) = (1, 1)$, all the states $<(k_1, k_2)| k_1 > k_1 + 1$ or $k_2 > k_2 + 1>$ are transient under these policies.
Let

\[ S' = \{ (k_1, k_2) \mid 1 \leq k_i \leq K_i + 1, \ i = 1, 2 \}, \]

\[ A'_s = \{ (1,1), (0,1) \}, \text{ for } s = (k_1, K_2 + 1), \ k_1 < K_1 + 1, \]

\[ A'_s = \{ (1,1), (1,0) \}, \text{ for } s = (K_1 + 1, k_2), \ k_2 < K_2 + 1, \]

\[ A'_s = A_s \text{ otherwise}, \quad (3.2) \]

\[ q' = q, \]

and

\[ \hat{r}' = \hat{r}. \]

From Lemma 3.1, the value function of the d.p.p. given in (3.2), \( \bar{V}' \), is not smaller than \( \bar{V} \).

Henceforth, we consider only the d.p.p. given in (3.2).

**Theorem 3.1:** There exists a non-randomized stationary control policy \( \pi^* \) such that

\[ \bar{V} = \sup_{\pi} \bar{V}(\pi) = \bar{V}(\pi^*). \]

**Proof:** From Corollary 3.1 we can reduce the d.p.p. given in Section 2 to a d.p.p. with finite state and action spaces, in which a non-randomized stationary optimal control policy exists. (See e.g. Derman [2].) \( \square \)

Hereinafter, we consider only non-randomized stationary control policies. Let \( u^*(k_1, k_2) \) be the control action taken at state \( (k_1, k_2) \) by the optimal policy \( \pi^* \).

In the following lemmas and theorem we show that attention can be restricted to policies of type \( \pi(k_1, k_2), \ k_1 \geq 1. \)

**Lemma 3.2:** \( u^*(k_1, k_2) \neq (0,0) \) for every state \( (k_1, k_2) \).

**Proof:** Under the control action \( u = (0,0) \), the state of the process remains unchanged and the immediate reward is zero. Therefore,
Lemma 3.3: For every state \((k_1, k_2)\); if \(p_1(k_1) > \frac{1}{2}\) for some \(i\), then \(u^*(k_1, k_2) \neq (1,1)\).

Proof: Suppose in contradiction that under \(\pi^*\), \(u^*(k_1, k_2) = (1,1)\) for some state \((k_1, k_2)\) and \(p_1(k_1) > \frac{1}{2}\). (If \(p_2(k_2) > \frac{1}{2}\) then the proof is similar.)

From Lemma 3.2 it follows that \(\pi^*\) can only be a policy of type \(\pi_1(k_1, k_2)\), \(\pi_2(k_1, k_2)\) or \(\pi(1,1)\).

Define \(\tilde{\pi}\) as follows:

If \(\pi^* = \pi_2(1, k_2)\) or \(\pi(1,1)\), then let \(\tilde{\pi}\) be the policy which always does \(u = (1,0)\).

If \(\pi^* = \pi_1(k_1, k_2)\) or \(\pi_2(k_1, k_2)\), \(k_1 \geq 2\) then let \(\tilde{\pi} = \pi(k_1, k_2 + 1)\).

From Remark 3.1, it is easy to check that at any case

\[\bar{V}(\tilde{\pi}) > \bar{V}(\pi^*)\]

which is a contradiction. □

Corollary 3.2: If \(r_i > \frac{1}{2}\), for some \(i\), then \(u^*(k_1, k_2) \neq (1,1)\) for every state \((k_1, k_2)\).

Proof: The corollary follows from Lemma 3.3, since \(p_1(k)\) is increasing in \(k\). □

Lemma 3.4: If \(u^*(k_1, k_2) = (1,1)\) for some state \((k_1, k_2)\), then \(u^* = (1,1)\).

Proof: From Corollary 3.2 we may assume that \(r_i \leq \frac{1}{2}\), \(i = 1, 2\).

From Lemma 3.2 it follows that any non-randomized stationary policy is
of type $\pi_1(k_1,k_2)$, $\pi_2(k_1,k_2)$ or $\pi(k_1,k_2)$, which are given in Figures 3.1 - 3.4. It also follows that any non-transient state is of the form $(1,k)$ or $(k,1)$. Suppose $u^*(k,1) = (1,1)$. (For $u^*(1,k) = (1,1)$, the proof is similar.)

If $k = 1$, then $\pi^* = \pi(1,1)$ and the lemma is proved.

If $k > 1$, then the process $k(t)$, $t = 0,1,2,...$, under $\pi^*$, enters into state $(k,1)$ only from state $(k-1,1)$ and the control action at state $(k-1,1)$ is $u^*(k-1,1) = (0,1)$. From state $(k,1)$, the process under $\pi^*$, moves on to state $(1,1)$.

The reward under $\pi^*$ during the transitions from $(k-1,1)$ to $(k,1)$ and then to $(1,1)$ is

$$r^* = 2p_2(1) + p_1(k)(1 - 2p_2(1)). \quad (3.3)$$

Now, let $\tilde{\pi}$ be the non-stationary policy which does the same as $\pi^*$, except when the process enters into state $(k-1,1)$. At this state, $\tilde{\pi}$ takes the control action $u = (1,1)$. Then, the process moves on to state $(1,1)$ in which $\tilde{\pi}$ takes again the control action $u = (1,1)$.

At this point $\tilde{\pi}$ proceeds as $\pi^*$. The reward during these two consecutive $(1,1)$-control actions is

$$r = 2p_2(1) + (p_1(k-1) + p_1(1))(1 - 2p_2(1)). \quad (3.4)$$

Since $p_2(1) = r_2 \leq \frac{1}{2}$ and $p_1(k-1) + p_1(1) > p_1(k)$ it follows from (3.3) and (3.4) that

$$r > r^*. \quad (3.5)$$

All the other immediate rewards remain unchanged under $\tilde{\pi}$. Thus from the mean ergodic theorem we obtain $V(\tilde{\pi}) > V(\pi^*)$, which is a contradiction. \(\square\)
The following theorem is a direct consequence of Lemma 3.4 and Corollary 3.2.

**Theorem 3.2:** For every \( 0 < r_1, r_2 < 1 \), \( \pi^* = \pi(k_1, k_2) \) for some \( k_i \geq 1, i = 1,2 \).

Next, we shall find the values \((k_1, k_2)\) which maximizes \( \bar{V}(\pi(k_1, k_2)), k_i \geq 2, i = 1,2 \), and then we shall show that this maximum is larger than \( V(\pi(1,1)) \).

From (2.5), Figures 3.3, 3.4 and Remark 3.1, we have

\[
\bar{V}(\pi(1,1)) = p_1(1) + p_2(1) - 2p_1(1)p_2(1) = r_1 + r_2 - 2r_1r_2 \quad (3.6)
\]

and

\[
\bar{V}(\pi(k_1, k_2)) = ((k_2-2)p_1(1) + (k_1-2)p_2(1) + p_2(k_2) + p_1(k_1))/(k_1+k_2-2). \quad (3.7)
\]

From (3.7) we have

\[
\bar{V}(\pi(k_1, k_2+1)) = \frac{k_1+k_2-2}{k_1+k_2-1} \bar{V}(\pi(k_1, k_2)) + \frac{1}{k_1+k_2-1} (p_2(k_2+1)-p_2(k_2)+p_1(1))
\]

and

\[
\bar{V}(\pi(k_1+1, k_2)) = \frac{k_1+k_2-2}{k_1+k_2-1} \bar{V}(\pi(k_1, k_2)) + \frac{1}{k_1+k_2-1} (p_1(k_1+1)-p_1(k_1)+p_2(1)). \quad (3.8)
\]

Thus

\[
\bar{V}(\pi(k_1, k_2+1)) > \bar{V}(\pi(k_1, k_2)) \left( \bar{V}(\pi(k_1+1, k_2)) > \bar{V}(\pi(k_1, k_2)) \right) \quad (3.9)
\]

if and only if

\[
p_2(k_2+1)-p_2(k_2)+p_1(1) > \bar{V}(\pi(k_1, k_2)) \left( p_1(k_1+1)-p_1(k_1)+p_2(1) > \bar{V}(\pi(k_1, k_2)) \right).
\]

**Lemma 3.5:** (a) If \( r_2 \geq r_1 \), then \( \bar{V}(\pi(2,2)) > p_2(3) - p_2(2) + p_1(1) \).

(b) If \( r_1 \geq r_2 \), then \( \bar{V}(\pi(2,2)) > p_1(3) - p_1(2) + p_2(1) \).
Proof: Since \( p_1(k+1) - p_1(k) \) decreases in \( k \), the lemma follows by a straightforward computation using the definition of \( \bar{V}(\pi(2,2)) \) given in (3.7).

**Theorem 3.3:** For finding the optimal control policies when \( r_2 \geq r_1 \) \((r_1 \geq r_2)\), the only policies which have to be considered are \( \pi(1,1) \) and \( \pi(k,2) (\pi(2,k)) \), \( k = 2,3,\ldots \).

Proof: Let \( k_1 \geq 2 \). If \( \bar{V}(\pi(k,2)) > p_2(3) - p_2(2) + p_1(1) \), then from (3.7) and (3.9) it follows that

\[
\max_{k_2} \bar{V}(\pi(k_1,k_2)) = \bar{V}(\pi(k_1,2)). \tag{3.10}
\]

If \( \bar{V}(\pi(k_1,2)) \leq p_2(3) - p_2(2) + p_1(1) \) then - since \( p_1(k+1) - p_1(k) \) decreases in \( k \) - it follows from (3.7), (3.9) and Lemma 3.5 that

\[
\max_{k_2} \bar{V}(\pi(k_1,k_2)) \leq p_2(3) - p_2(2) + p_1(1) < \bar{V}(\pi(2,2)). \tag{3.11}
\]

Now, the theorem follows from (3.10) and (3.11).

In Theorem 3.4 below it will be shown that \( \pi(1,1) \) is not optimal.

**Lemma 3.6:** For every \( k = 0,1,2,\ldots \)

\[(1-r)^k \leq 1 - kr + r^2(k-1)^2.\]

Proof: By a standard induction on \( k \).

**Theorem 3.4:** For \( r_2 \geq r_1 \) \((r_1 \geq r_2)\) the policy \( \pi(k_0,2) \), where

\[
k_0 = \left\lfloor \frac{r_2}{r_1} \right\rfloor \pi(2,k_0), \quad \text{where} \quad k_0 = \left\lfloor \frac{r_1}{r_2} \right\rfloor \]

is better than \( \pi(1,1) \). (\( \lfloor a \rfloor \) is the smallest integer not smaller than \( a \).)

Proof: We consider only the case \( r_2 \geq r_1 \).
From (3.6) and (3.7) we have

\[ \bar{V}(\pi(k,2)) = p_2(1) + (p_1(k) - r_2^2)/k = r_2 + (1-(1-r_1)^k - r_2^2)/k \quad (3.12) \]

\[ \bar{V}(\pi(1,1)) = r_1 + r_2 - 2r_1r_2. \quad (3.13) \]

From (3.12) and (3.13) it is sufficient to show that

\[ g(k_o) = 1 - (1-r_1)^k - r_2^2 - kr_1 + 2kr_1r_2 \geq 0, \]

where \( k_o = \left\lfloor \frac{r_2}{r_1} \right\rfloor. \)

From Lemma 3.6 and the definition of \( k_o \) we obtain

\[ g(k_o) \geq 1 - (1-k_o r_1 + r^2(k_o - 1)^2) - r_2^2 - k_o r_1 + 2k_o r_1r_2 \]

\[ = 2k_o r_1r_2 - r_2^2 - (k_o - 1)^2 r_1 \geq 2r_2^2 - 2r_2^2 = 0. \quad \square \]

Finally, we present the conclusive theorem, which determines the optimal control policy.

Theorem 3.5: (a) If \( r_2 \geq r_1 \), then \( \pi^* = \pi(k^*_1,2) \) where \( k^* \) is the smallest integer satisfying

\[ r_2^2 \leq p_1(k^* + 1) - (k^* + 1)r_1(k-r_1)^{k^*}. \quad (3.14) \]

Moreover, \( \bar{V}(\pi^*) = r_2 + (p_1(k^*) - r_2^2)/k^* \).

(b) If \( r_1 \geq r_2 \), then \( \pi^* = \pi(2,k^*) \), where \( k^* \) is the smallest integer satisfying

\[ r_1^2 \leq p_2(k^* + 1) - (k^* + 1)r_2(1-r_2)^{k^*}. \quad (3.15) \]

Moreover, \( \bar{V}(\pi^*) = r_1 + (p_2(k^*) - r_1^2)/k^* \).
Proof: We consider only (a). From (3.12) it follows that

\[ \frac{V(r(k+1,2))}{V(n(k,2))} \leq 1 \quad \text{if and only if} \]

\[ r_2^2 \leq p_1(k) - k(1-r_1)^k r_1 = p_1(k+1) - (k+1)(1-r_1)^k r_1. \]  \hspace{1cm} (3.16)

The theorem now follows from (3.12) and (3.16).
4. CONCLUSIONS

(i) The optimal control policies $\pi(k^*, 2)$ and $\pi(2, k^*)$ are open loops and do not allow conflicts. Furthermore, they can easily be implemented as decentralized policies.

(ii) The probabilistic interpretation for the optimal value $k^*$ in condition (3.14) is as follows:
Let $x_k^i \sim B(k, r_i)$, $i = 1, 2$, $k = 1, 2, \ldots$. The integer $k^*$ is the smallest integer $k$ which satisfies

$$P(x_k^1 \geq 2) > P(x_k^2 \geq 2).$$

(iii) If $r_1 = r_2$, the optimal control policy is the "round-robin" policy.
REFERENCES


