stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 56/75 OCTOBER

I. MEILIJSON

MULTIPLE FEEDBACK AT A SINGLE SERVER STATION

2e boerhaavestraat 49 amsterdam

*Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.*

*The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.*

---

AMS(MOS) subject classification scheme (1970): 60K25, 90B35

---

# Multiple feedback at a single server station

by

I. Meilijson
Tel Aviv University

ABSTRACT

A queue model is studied in which the arrival stream depends on the type of the customer being served. A priority ordering of the types is defined, and the service policy it generates is proved optimal. This note surveys and generalizes the work of BRUNO, KLIMOV, and MEILIJSON and WEISS.

KEY WORDS & PHRASES: M|GI|1 *queue, single server, priority, service policy.*

# Multiple feedback at a single server station

by

I. Meilijson*
Tel Aviv University.

## 1. INTRODUCTION

KLIMOV [2] dealt with an M|GI|1 queue model with several types of customers, in which a customer of type i, upon leaving the server, rejoins the queue as a customer of type j with probability Q(i,j) and leaves the facility with probability $1 - \Sigma_j Q(i,j)$. He defined an ordering of the types of customers and proved the priority rule it generates to minimize expected rates of cost under steady state. The ordering depends on expected service lengths, on holding costs and on the matrix Q but not on the intensities of the Poisson streams. The priority rule agrees with the so called "cμ" priority rule when $Q(i,j) \equiv 0$.

BRUNO [1] and MEILIJSON & WEISS [3] proved Klimov's rule to be optimal for the service of a batch of customers.

These models may be unified and generalized as follows. Let the non-negative random variables $v_i$, $c_i$, $n_{ij}$ be the length of service of a customer of type i, its holding cost per unit time and the number of customers of type j that arrived during its service. Assume that, given the types of the customers, the vectors $(v_i; c_i; n_{i1}, n_{i2}, \ldots, n_{ir})$ corresponding to different customers are independent, and those corresponding to customers of the same type are, in addition, identically distributed.

The server's problem is to find a service policy that will minimize the expected total cost during the current busy period.

The purpose of this note is to define an ordering of the types

---

* Research conducted during the author's stay at the Mathematisch Centrum, Amsterdam, in the Summer of 1975.

for the model defined above, to prove the priority rule it generates to be an optimal service policy, and to prove that the ordering is unaffected by time-homogeneous modifications of the stream.

## 2. ASSUMPTIONS AND RESULTS.

(i) $v(i) = E(v_i)$ and $c(i) = E(c_i)$ are positive and finite, $n(i,j) = E(n_{ij})$ are non-negative and finite.

(ii) All eigenvalues of $((n(i,j)))$ are strictly less than 1 in absolute value.

(iii) The length of a busy period and the number of customers served during it possess finite second moments.

REMARK. Assumption (iii) deals with variables that do not depend on service policy.

Let r be the number of types of customers and denote $R = \{1, 2, \ldots, r\}$.
For a matrix M on $R \times R$ and sets $A \subseteq B \subseteq R$, the matrix $M_A$ on $A \times A$ has $M_A(i,j) = M(i,j)$ and the matrix $M_{A,B}$ on $A \times B$ has $M_{A,B}(i,j) = M(i,j)\ \phi_{B-A}(j)$, where $\phi_K$ is the indicator function of the set K.
For a vector w on R and sets $A \subseteq B \subseteq R$, the vector $w_A$ on A has $w_A(i) = w(i)$ and the vector $w_{A,B}$ on B has $w_{A,B}(i) = w(i)\ \phi_{B-A}(i)$. The *direct product* $w_1 * w_2$ of the vectors $w_1$ and $w_2$ on R is the vector on R with $w_1 * w_2(i) = w_1(i)\ w_2(i)$, and their *direct ratio* $w_1 \overset{*}{*} w_2$ is the vector on R with $w_1 \overset{*}{*} w_2(i) = w_1(i)/w_2(i)$ if $w_2(i) \neq 0$, $= 0$ if $w_2(i) = 0$.
The vector with coordinates v(i) (c(i)) is v(c). The matrix with coordinates n(i,j) is N. The vector all of whose coordinates are 1 is 1. *Vectors* are column vectors unless transposed by (').

For a non-empty subset A of R, define the following vectors on A.

(1) $d(A) = c_A - (I_A - N_A)^{-1} N_{A,R} c$

(2) $\gamma(A) = (I_A - N_A)^{-1} v_A$

(3) $H(A) = d(A) \overset{*}{*} \gamma(A)$

Define a vector H on R by

(4) $H(i) = \max_{A \subseteq R} H(A \cup \{i\})\ (i)$

A service policy is H-*monotone* if at every decision moment it chooses almost surely to serve one of the customers whose type has the highest value of H(i) among those in the queue.

THEOREM 1. *A service policy minimizes the expected total cost during a whole busy period if and only if it is* H-*monotone.*

Let the non-empty sets $R_1^*, R_2^*, \ldots, R_\ell^*$ be the partition of R with $(i,j) \in R_k^* \Rightarrow H(i) = H(j)$ and $i \in R_k^*$, $j \in R_{k+1}^* \Rightarrow H(i) < H(j)$. $(R_1^*, \ldots, R_k^*)$ is called the *optimal priority partition* of R. To compute it it is not necessary to perform the maximizations in (4):

(5) THEOREM 2. $R_1^* = \{i \in R \mid H(R)(i) = \min_{j \in R} H(R)(j)\}$. *Let* $R_2 = R - R_1^*$. *If* $R_2 = \emptyset$ *then* $\ell = 1$. *Otherwise, inductively,* $R_k^* = \{i \in R_k \mid H(R_k)(i) = \min_{j \in k} H(R_k)(j)\}$. *Let* $R_{k+1} = R_k - R_k^*$. *If* $R_{k+1} = \emptyset$ *then* $\ell = k$.

THEOREM 3. *Assume that for some non-negative matrix* M *on* R×R *and some non-negative vector* $\lambda$ *on* R, $N = M + v\lambda'$. *Denote by* $H^{(M)}$, $\ell^{(M)}$, $R_k^{*(M)}$ *the corresponding expressions missing* (M) *computed as if* N *was* M. *Then,*

(i) For every non-empty subset A of R,

$$H(A) = \frac{1}{1 + \lambda_A' \gamma(A)} \left(H^{(M)}(A) - (\lambda' c - \lambda_A' d(A)) 1_A\right) \tag{6}$$

(ii) $\ell^{(M)} = \ell$ and for each $1 \le k \le \ell$, $R_k^{*(M)} = R_k^*$.

REMARK. $v\lambda'$ is a *time homogeneous* factor to the stream.

A word about d, $\gamma$ and H.

By ([4], theorem 1.1) assumption (ii) implies that $I_A - N_A$ is non-singular for every non-empty subset A of R. Express $(I_A - N_A)^{-1} = \Sigma_{k=0}^{\infty} N_A^k$ to infer that $(I_A - N_A)^{-1} v_A(i)$ is the expected time it will take to serve customers to exhaustion, when there is originally one customer only, its type is i, and only customers whose types belong to A are provided service. At the conclusion of that time, the expected number of customers of type j in the queue is $(I_A - N_A)^{-1} N_{A,R}(i,j)$.

Imagine John and Bob are the only customers in the queue, John's type is i and Bob's type is j. Let $i \in A \subseteq R$, $j \in B \subseteq R$. Under the policy JB start by serving John, then serve to exhaustion customers with types in A but do not serve Bob. Now serve Bob, then serve to exhaustion customers with types in B but do not serve those in the queue at the moment Bob's service started. Proceed in some arbitrary manner $\Pi$. Define a policy BJ in the same way, using the same $\Pi$ as before. To compare the performances of JB and BJ we may disregard the common tail $\Pi$. The relevant waiting costs to compare are, then,

$$c(j)\,\gamma(A)(i) + (I_A - N_A)^{-1} N_{A,R}\, c(i)\,\gamma(B)(j)$$

and

$$c(i)\,\gamma(B)\,(j) + (I_B - N_B)^{-1} N_{B,R}\, c(j)\,\gamma(A)(i).$$

In other words, H(B)(j) and H(A)(i) are to be compared. If H(i) > H(j) then for some set A containing i and all sets B containing j, we would rather use JB than BJ. This is an intuitive reason why the customer of type i should be preferred.

## 3. PROOFS

We will prove below formula (12), which is the same as ([3], formula (7)), for the present more general H. Beyond that, there is nothing else to prove to obtain theorems 1 and 2. We will just point out how does everything follow from results in [3]. Theorem 2 follows from (12) just as ([3], theorems 1 and 2) follow from ([3], (7)). To obtain theorem 1, observe that H as defined in (4) is already the H corresponding to *vector customers* in ([3], section 6), and that the lexicographic order on the *vector stages* as defined in ([3], section 6) applied to the present set-up makes H monotone, so ([3], assumption 2) is satisfied and every H-monotone policy is excessive. Assumption (iii), which corresponds to ([3], assumption 4(ii)), permits us, then, to infer optimality from excessivity.

LEMMA 1. *For* $i \in S \subseteq B \subseteq R$,

$$d(B)(i) - d(S)(i) = (I_S - N_S)^{-1} N_{S,B}\, d(B)(i) \tag{7}$$

and

(8) $\gamma(B)(i) - \gamma(S)(i) = (I_S - N_S)^{-1} N_{S,B}\, \gamma(B)(i)$

PROOF. Use the probabilistic interpretation of d and $\gamma$ following (6) to express, for $i \in S$, $j \in B$,

(9) $$(I_B - N_B)^{-1} N_{B,R}(i,j) = (I_S - N_S)^{-1} N_{S,R}(i,j) + (I_S - N_S)^{-1} N_{S,B} (I_B - N_B)^{-1} N_{B,R}(i,j)$$

and for $i \in S$, $j \in B$,

(10) $$(I_B - N_B)^{-1} N_{B,R}(i,j) = (I_S - N_S)^{-1} N_{S,R}(i,j) - (I_S - N_S)^{-1} N_{S,B}(i,j) + (I_S - N_S)^{-1} N_{S,B} (I_B - N_B)^{-1} N_{B,R}(i,j).$$

So, combining (9) and (10),

$$d(B)(i) - d(S)(i) = (I_S - N_S)^{-1} N_{S,R}\, c(i) - (I_B - N_B)^{-1} N_{B,R}\, c(i) =$$

$$= (I_S - N_S)^{-1} N_{S,B} c_B(i) - (I_S - N_S)^{-1} N_{S,B} (I_B - N_B)^{-1} N_{B,R}\, c(i) =$$

$$= (I_S - N_S)^{-1} N_{S,B} d(B)\,(i)$$

thus proving (7). (8) is immediate. □

Define a vector $H_S(B)$ on S by

(11) $$H_S(B) = ((I_S - N_S)^{-1} N_{S,B} d(B)) \overset{*}{*} ((I_S - N_S)^{-1} N_{S,B} \gamma(B)).$$

$H_S(B)$ is a convex combination of the values of H(B) (j) for $j \in B-S$. The following expression, (12), follows immediately from Lemma 1. For $i \in S \subseteq B \subseteq R$,

(12) $$H(B)(i) = (\gamma(S)(i) / \gamma(B)(i))\, H(S)(i) + (1 - (\gamma(S)(i) / \gamma(B)(i)))\, H_S(B)(i).$$

Proof of theorem 3.

Check that for every vector w on A,

(13) $(I_A - N_A)^{-1} w = (I_A - M_A)^{-1} w + (\lambda_A'(I_A - N_A)^{-1} w).\ \gamma^{(M)}(A)$

Substitute $N_{A,R}c$ as w in (13) to express

(14) $d(A) = d^{(M)}(A) - (\lambda' c - \lambda'_A d(A))\ \gamma^{(M)}(A)$

Substitute $v_A$ for w in (13) to express

(15) $\gamma(A) = (1 + \lambda'_A \gamma(A))\ \gamma^{(M)}(A)$

The direct ratio of (14) and (15) yields (6). (ii) follows easily from (i), using theorem 2. □

ACKNOWLEDGEMENT.

I wish to thank Avi Federgrun for some fruitful conversations. My thanks are also due to everybody at the Mathematisch Centrum in Amsterdam for making my stay so enjoyable.

REFERENCES.

[1] BRUNO, J.L. (1975) *Task Scheduling with a (usually) finite planning horizon.* Technical Report No. 163, Computer Science Department, The Pennsylvania State University.

[2] KLIMOV, G.P. (1974) Time-Sharing Service Systems, I. *Theory of Probability and its Applications,* 19 532-551.

[3] MEILIJSON, I & WEISS, G. (1975) Time-Sharing via Dynamic Programming. Technical Report No. 68, Department of Statistics, Tel Aviv University.

[4] SENETA, E. (1973). *Non-negative matrices.* George Allen & Unwin Ltd., London,