ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XVIII, 4 (1985), p.507-515

R. L. DISNEY and D. C. McNICKLE (Blacksburg, Virginia)

SEMI-MARKOV SWITCHES*

In this paper we consider some properties of a large class of network decomposition switches. Expressions are given for the semi-Markov matrices of the output Markov renewal processes they produce, and a classification theorem for the states in these processes is proved.

1. Introduction. The limited range of results available for multiserver queueing systems and networks of service facilities means that we must often consider methods by which the network may be split up or decomposed into units which can be analyzed separately. A natural place to split up a network is at a point where a single stream of arrivals or customers is itself split up into several output streams. The mechanism governing the allocation of customers is called a switching rule. Usually the only such rule which has been considered is that of random assignment, where the probability of assigning a customer from the (usually Poisson) arrival stream to a particular output is assumed to remain constant. Such a switch amounts to little more than random deletions from a Poisson stream and may well be unsatisfactory since it allows no dependence on properties of the arrival stream such as customer type. Further, the Poisson arrival assumption on which this switch works best is rarely preserved in non-Markov networks.

The review paper by Disney [5] emphasizes the importance of Markov renewal processes in such aspects of networks as superposition, feedback, departure streams from service facilities, and overflow. We define a class of switches which will produce semi-Markov output streams. With the

* This research was supported in part by the Office of Naval Research, under contract N00014-75-C-0492 NR 042-292.

additional assumption of a semi-Markov input stream special cases of this class have been considered by Cinlar [2], [3] and Hall and Disney [8]. An application to the allocation of medical emergency services can be found in Hall [7]. By imposing a further property we can show that a subclass containing these special cases, in fact requires that the input to the switch be a semi-Markov process.

A broader class of switches based on the imbedded Markov chain was suggested in Disney and Cherry [6]. It should be mentioned that such switches may produce output streams which are not Markov renewal processes because of dependences transmitted through the continuous component of the output process.

2. Semi-Markov switches. We consider an arrival stream containing a number of different customer types, with J_n the type of the nth arrival. Customers arrive at times $0 = T_0 < T_1 < \dots$, we write $X_n = T_n - T_{n-1}$, and assume that J_n takes values in some countable set I. The switch is defined by the random variable Y_n, which takes a finite number of values, R. Thus, if $Y_n = r$, then the nth arrival is directed to the rth output stream. The behavior of the process $\{J_n, Y_n, X_n | n \geq 0\}$ describes both the switch and the arrival process. If the event $\{Y_n = r\}$ occurs on a subset of the arrival times given by $0 \leq T_{0'} < T_{1'} < \dots$ then the process $\{J_{n'}, Y_{n'} = r, T_{n'} | n' \geq 0\}$ describes the rth output stream. We assume that the initial state of the system is determined by some vector of probabilities for (J_0, Y_0). We consider the following class.

DEFINITION. A switch $\{Y_n\}$ is called a *semi-Markov switch* if

$$\text{(a)} \quad P(J_n = j, Y_n = r, X_n \leq t | J_{n-1}, \dots, J_0, Y_{n-1}, \dots, Y_0, X_{n-1}, \dots, X_1)$$
$$= P(J_n = j, Y_n = r, X_n \leq t | J_{n-1}, Y_{n-1}), \quad \text{for } j \in I, t > 0, r \in \{1, \dots, R\}.$$

We define as *stationary semi-Markov switches*, the subclass of semi-Markov switches which have the additional property:

$$\text{(b)} \quad P(J_n = j, X_n \leq t | J_{n-1}, \dots, J_0, X_{n-1}, \dots, X_0, Y_{n-1}, \dots, Y_0)$$
$$= P(J_n = j, X_n \leq t | J_{n-1}).$$

3. Lumpability. Let $\{Z_n, X_n\}$ be a Markov renewal process with semi-Markov kernel $G(t) = \{g_{rs}(t)\}$ defined on a countable state space I', and f a function from I' onto another countable set I. $\{Z_n, X_n\}$ is said to be *lumpable* with respect to f if, for any initial distribution p on $\{Z_n\}$, the process $\{J_n, X_n\}$, where $J_n \doteq f(Z_n)$, is a Markov renewal process whose transition matrix does not depend on p.

Serfozo [10] shows that necessary and sufficient conditions for lumpability are that for any two sets S_a, S_b in the partition of I' induced by f

$$\sum_{s \in S_b} g_{rs}(t) = \sum_{s \in S_b} g_{qs}(t), \quad \text{for each } r, q \in S_a \text{ and } t \geqslant 0.$$

We will call the $\{Z_n, X_n\}$ process a *lumpable expansion* of the $\{J_n, X_n\}$ process if there exists a lumping function f from I' onto I such that $J_n = f(Z_n)$.

With these definitions, we can characterize the behavior of the switch, the type of the arrival process, and their interaction.

THEOREM 1. *The conditions* (a) *and* (b) *are equivalent to the single condition*: $\{J_n, Y_n, X_n | n \geqslant 0\}$ *is a lumpable expansion of the* $\{J_n, X_n\}$ *process.*

Proof. Assume that $\{Y_n\}$ satisfies the conditions for a semi-Markov switch. Condition (a) shows that $\{J_n, Y_n, X_n\}$ is a Markov renewal process defined on the space $I \times \boldsymbol{R}$, where $\boldsymbol{R} = (1, 2, \ldots, R)$, and I is the state space of the input process.

If we consider the function f: $(j, r) \to r$, which defines a partition of $I \times \boldsymbol{R}$ consisting of sets of the form $\{(j, r) | r = 1, \ldots, R\}$, then we can show that f is a lumping function, since:

$$\begin{aligned}
&\sum_{r=1}^{R} P(J_n = j, Y_n = r, X_n \leqslant t | J_{n-1}, Y_{n-1} = q) \\
&\quad = \sum_{r=1}^{R} P(Y_n = r | J_n = j, X_n \leqslant t, J_{n-1}, Y_{n-1} = q) P(J_n = j, X_n \leqslant t | J_{n-1}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad Y_{n-1} = q) \\
&\quad = P(J_n = j, X_n \leqslant t | J_{n-1}, Y_{n-1} = q) \\
&\quad = P(J_n = j, X_n \leqslant t | J_{n-1}) \quad \text{from condition (b).}
\end{aligned}$$

Hence $\{J_n, Y_n, X_n\}$ is a lumpable expansion of $\{J_n, X_n\}$.

If, on the other hand, $\{J_n, Y_n, X_n\}$ is a lumpable expansion of $\{J_n, X_n\}$, with f a lumping function defined as above, then (b) follows immediately, since $\{J_n, Y_n, X_n\}$ is a Markov renewal process.

For (a):

$$\begin{aligned}
&P(J_n = j, X_n \leqslant t | J_{n-1}, Y_{n-1}, X_{n-1}, \ldots, J_0, Y_0) \\
&\qquad = \sum_{r=1}^{R} P(J_n = j, Y_n = r, X_n \leqslant t | J_{n-1}, Y_{n-1}, X_{n-1}, \ldots, J_0, Y_0), \\
&\qquad = \sum_{r=1}^{R} P(J_n = j, Y_n = r, X_n \leqslant t | J_{n-1}, Y_{n-1}), \\
&\qquad = P(J_n = j, X_n \leqslant t | J_{n-1}) \quad \text{from Serfozo's result.}
\end{aligned}$$

In the next section, we show that a semi-Markov switch determines the type of the output streams. The result of Theorem 1 implies that the two properties of a stationary semi-Markov switch also characterize the arrival process.

From the definition of lumpability and Theorem 1 we immediately get

THEOREM 2. *For a stationary semi-Markov switch, the arrival process, $\{J_n, X_n\}$, must be a Markov renewal process whose kernel does not depend on the initial distribution of* (J_0, Y_0).

4. Output streams. To show that each of the output streams from a semi-Markov switch forms a Markov renewal process we need a result originally due to Anderson [1], which is discussed in Cinlar [4]. Consider a subset, K, of the states of a countable state Markov renewal process, and define a decomposition switch, giving rise to two output streams, by:

$$P(Y_n = 1 | J_n \in K) = 1, \qquad P(Y_n = 2 | J_n \notin K) = 1.$$

Anderson shows that each of the output streams of this switch (or filter) is a (possibly delayed) Markov renewal process. Since each of the output streams is of this type it must form a Markov renewal process.

Although all of the published examples of switches producing Markov renewal outputs are semi-Markov it should be noted that this is not a necessary condition for the production of Markov renewal outputs, as the following example shows. Consider a renewal process with distribution function $F(t)$, which forms the input to a switch producing three output streams. Customers are assigned deterministically in the order $\ldots 1, 2, 3, 2, 1, 2, 3, \ldots$ Now

$$P(Y_n = 3, X_n < t | Y_{n-1} = 2, Y_{n-2}, \ldots, Y_0) = \begin{cases} 0, & Y_{n-2} = 3, \\ F(t), & Y_{n-2} = 1, \end{cases}$$

and so the switch is not semi-Markov, yet clearly streams 1 and 3 are renewal processes with common distribution function $F^4(t)$, and stream 2 is a renewal process with distribution function $F^2(t)$. Here $F^n(t)$ stands for the n-fold convolution of $F(t)$. Note that this switch can, however, be considered to be semi-Markov if we assume the input process has an imbedded Markov chain of order 2. We have the following general result.

THEOREM 3. *Any stationary switch that acts on a Markov renewal process to produce two output streams which are Markov renewal processes defined on the same state space is semi-Markov.*

Proof. Let us assume for convenience that the last customer was assigned to output stream 1. Then

$$P(J_n = j,\ Y_n = 1,\ X_n \leqslant t \mid J_{n-1},\ Y_{n-1} = 1,\ X_{n-1}, \ldots, J_0,\ Y_0)$$
$$= P(J_n = j,\ Y_n = 1,\ X_n \leqslant t \mid J_{n-1},\ Y_{n-1} = 1),$$

since the first output stream is Markov renewal.

$$\begin{aligned} P(J_n = j,\ Y_n = 2,\ X_n \leqslant t \mid J_{n-1},\ Y_{n-1} = 1,\ X_{n-1}, \ldots, J_0,\ Y_0) \\ = P(J_n = j,\ X_n \leqslant t \mid J_{n-1},\ Y_{n-1} = 1,\ X_{n-1}, \ldots, J_0,\ Y_0) - \\ - P(J_n = j,\ Y_n = 1,\ X_n \leqslant t \mid J_{n-1},\ Y_{n-1} = 1, \ldots, J_0,\ Y_0) \\ = P(J_n = j,\ Y_n = 2,\ X_n \leqslant t \mid J_{n-1},\ Y_{n-1} = 1), \end{aligned}$$

since the input and the first output stream are Markov renewal.

5. The semi-Markov matrices of the output streams. We assume for convenience that $\{J_n, Y_n, X_n\}$ is an irreducible Markov renewal process and that the initial arrival was assigned to the rth output stream. Events in the filter set $K = I \times \{r\}$ of the state space of the Markov renewal process form the rth output stream.

In particular, let $n_0 = 0$,

$$n_{k+1} = \inf(i > n_k\colon (J_i,\ Y_i) \in K), \qquad k = 0, 1, 2, \ldots,$$

and define the process $\{Z_k, \tau_k\}$ by:

$$Z_k = J_{n_k}, \qquad \tau_k = T_{n_k}, \qquad \text{for} \qquad k = 0, 1, 2, \ldots$$

We have shown previously that $\{Z_k, \tau_k\}$ is a Markov renewal process which describes the behavior of the rth output stream. Let

$$G(t) = \{G_{ij}(t)\} = \{P(Z_{k+1} = j,\ W_{k+1} \leqslant t \mid Z_k = j)\}, \qquad i, j = 1, \ldots, m,$$

where $W_{k+1} = \tau_{k+1} - \tau_k$. A formal expression for the semi-Markov matrix of the tth output stream, $G(t)$, may be found from Cinlar [4].

We write the semi-Markov matrix of the $\{J_n, Y_n, X_n\}$ process in block form as $A = \{A_{pq}\}$, where the i, j-th element of the $m \times m$ matrix $A_{pq}(t)$ is

$$P(J_n = j,\ Y_n = q,\ X_n \leqslant t \mid J_{n-1} = i,\ Y_{n-1} = p).$$

Then on relabelling the output streams if necessary, A may be partitioned as

$$A = \begin{bmatrix} A_{rr} & B \\ C & D \end{bmatrix}.$$

Then $G(t)$ satisfies:

$$G(t) = A_{rr}(t) + B * (\sum_n D^n) * C(t),$$

where $*$ stands for the usual matrix convolution operation and D^n for the n-fold convolution of D with itself.

If the first arrival was assigned to the sth output stream, where $s \neq r$, then partitioning A as:

$$A = \begin{bmatrix} A_{sr} & B' \\ C' & D' \end{bmatrix}$$

leads to a similar expression for the matrix of transition functions for the time until the first event in the rth output stream.

It appears, however, that no simplification of these expressions can be made unless the class of switches is restricted further.

6. Switches depending only on the input Markov renewal process. Consider the class of stationary semi-Markov switches in which the switch no longer depends explicitly on the assignment of the last customer. Thus,

$$P(Y_n = r | J_n, X_n, J_{n-1}, Y_{n-1}) = P(Y_n = r | J_n, X_n, J_{n-1}), \quad \text{for} \quad n = 1, 2, \ldots$$

The semi-Markov matrices for the output streams of a special case of this class where arrivals are assigned according to type alone were derived directly in Cinlar [2]. By considering a suitable Markov renewal equation, expressions for those matrices of all switches of this class can be found.

Since the switch is stationary, for any $s \in \{1, 2, \ldots, R\}$

$$P(J_n = j, Y_n = r, X_n \leqslant t | J_{n-1} = i, Y_{n-1} = s) = f_{ijr}(t) = f_{ij}(t) q_r(i, j, t),$$

where $F(t) = \{f_{ij}(t)\}$ is the semi-Markov matrix of the arrival process, and

$$q_r(i, j, t) = P(Y_n = r | J_n = j, X_n \leqslant t, J_{n-1} = i).$$

Let $G(t) = \{G_{ij}(t)\}$ be the semi-Markov matrix for the rth output stream. Then since the condition on the switch implies that all the higher order transition probabilities of the $\{J_n, Y_n, X_n\}$ process are independent of the initial state of the switch

$$G_{ij}(t) = \sum_{l=1}^{\infty} P(J_{n_{k+1}} = j, n_{k+1} - n_k = l, T_{n_{k+1}} - T_{n_k} \leqslant t | J_{n_k} = i)$$

$$= f_{ijr}(t) + \sum_{l=2}^{\infty} \sum_{h=1}^{m} \sum_{\substack{s=1 \\ s \neq r}}^{R} \int_0^t P(J_{n_k+1} = h, Y_{n_k+1} = s,$$

$$X_{n_k+1} \in (x, x+dx) | J_{n_k} = i)$$

$$P(J_{n_{k+1}}=j,\ n_{k+1}-n_k-1=l-1,\ T_{n_{k+1}}-T_{n_k+1}\leqslant(t-x)\,|\,J_{n_k+1}=h,\ Y_{n_k+1}=s)$$
$$=f_{ijr}(t)+\sum_{l=2}^{\infty}\sum_{h=1}^{m}\int_0^t\big(f_{ih}(dx)-f_{ihr}(dx)\big)\times$$
$$\times P(J_{n_{k+1}}=j,\ n_{k+1}-n_k-1=l-1,\ T_{n_{k+1}}-T_{n_k+1}\leqslant(t-x)\,|\,J_{n_k+1}=h).$$

Hence

$$G_{ij}(t)=f_{ijr}(t)+\sum_{h=1}^{m}\int_0^t\big(f_{ih}(dx)-f_{ihr}(dx)\big)G_{hj}(t-x),\qquad\text{for all } i,j\in I. \tag{1}$$

If the first arrival to the switch was not assigned to the rth output stream then it can be seen that the conditions on this class of switch means that the distributions of the delay until the first event in the rth output stream also satisfy the equations (1). These equations can be solved under a further weak assumption to give:

THEOREM 4. *If all the states of the input Markov renewal process are conservative then the r-th output stream from the switch forms a Markov renewal process with kernel and matrix-valued distribution function of the time until the first event both given by*

$$G(t)=\int_0^t R_r(dx)F_r(t-x), \tag{2}$$

where $F_r(t)=\{f_{ijr}(t)\}$, *and* $R_r(t)$ *is the Markov renewal matrix corresponding to the semi-Markov matrix* $\{f_{ij}(t)-f_{ijr}(t)\}$.

Proof. $F_r(t)$ and $\{f_{ij}(t)-f_{ijr}(t)\}$ are both semi-Markov matrices, so (1) is a Markov renewal equation. If all states in I are conservative then Cinlar [4] shows that a unique solution to equations of this type exists, and in this case is given by (2).

If I is finite, then at least for $\mathrm{Re}(s)>0$

$$\boldsymbol{G}(s)=\sum_{n=0}^{\infty}\big(\boldsymbol{F}(s)-\boldsymbol{F}_r(s)\big)^n\boldsymbol{F}_r(s)=\big(I-\boldsymbol{F}(s)+\boldsymbol{F}_r(s)\big)^{-1}\boldsymbol{F}_r(s), \tag{3}$$

where

$$\boldsymbol{F}_r(s)=\int_0^{\infty}e^{-st}F_r(dt),\qquad \boldsymbol{G}(s)=\int_0^{\infty}e^{-st}G(dt).$$

We can further note that if $f_{ijr}(t)=f_{ij}(t)\cdot q_r(j)$, so that the switch depends on the type of the arrival only, then with $Q=\{\delta_{ij}q_r(j)\}$

$$\boldsymbol{G}(s)=\big(I-\boldsymbol{F}(s)+\boldsymbol{F}(s)Q\big)^{-1}\boldsymbol{F}(s)Q=\boldsymbol{F}(s)\big(I-(I-Q)\boldsymbol{F}(s)\big)^{-1}Q,$$

which was Cinlar's result [3].

One example of the application of (3) is to a generalization of Palm's overflow problem to the case of Markov renewal input. Customers arrive according to a Markov renewal process at a system of N servers with mean service time $1/\mu$. If there is no waiting room and service times are exponential then the streams of customers who join each server and those who overflow to the next can be considered as the results of a semi-Markov switches of this type. Thus if $\boldsymbol{F}^n(s)$ represents the input to the nth server then $\boldsymbol{F}_r^n(s) = \boldsymbol{F}^n(s+\mu)$, and so

$$\boldsymbol{F}^{n+1}(s) = \left(I - \boldsymbol{F}^n(s) + \boldsymbol{F}^n(s+\mu)\right)^{-1} \boldsymbol{F}^n(s+\mu), \qquad n = 1, 2, \ldots, N-1$$

is the transform of the kernel of the overflow processes. The renewal case of this result is well known ([9]).

7. Classification of states in the output streams. Although states which are transient in the arrival process will clearly also be transient in any output stream, the same is not true for recurrent states.

Let $f_{ijrs} = P(J_n = j,\ Y_n = s \mid J_{n-1} = i,\ Y_{n-1} = r)$, with $G = \{g_{ij}\}$ the imbedded Markov chain for the rth output stream. We write (j, r) for the state $(J_n = j,\ Y_n = r)$.

The following result completes the characterization of the output streams of a semi-Markov switch.

THEOREM 5. *For a stationary semi-Markov switch state j in the r-th output stream will be recurrent if and only if there exists some recurrent state k in the arrival process such that $j \to k$ and if $(j, r) \to (k, s)$ for any s then $(k, s) \to (j, r)$.*

Proof. Clearly

$$\sum_{p=1}^{\infty} g_{jj}^{p} = \sum_{l=1}^{\infty} f_{jjrr}^{l},$$

where g_{jj}^p and f_{jjrr}^l are the p-step and the l-step transition probabilities of the respective Markov chains.

Let $R' = \{s \mid (j, r) \to (k, s)\}$, with $N(s) = \min(n \mid f_{kjsr}^{n} > 0)$, $n = \max_{R'}(N(s))$, and $\delta = \min_{R'}(f_{kjsr}^{N(s)})$

$$\sum_{p=1}^{\infty} g_{jj}^{p} \geqslant \sum_{l=N}^{\infty} \sum_{k=1}^{m} \sum_{s=1}^{R} f_{jkrs}^{l-N(s)} f_{kjsr}^{N(s)} \geqslant \sum_{l=N}^{\infty} \sum_{R'} f_{jkrs}^{l-N(s)} f_{kjrs}^{N(s)}$$
$$\geqslant \delta \sum_{l=N}^{\infty} \sum_{R'} f_{jkrs}^{l-N(s)} = \delta \sum_{l=N}^{\infty} \sum_{s=1}^{R} f_{jkrs}^{l-N(s)}$$

which is divergent, since k is recurrent and $j \to k$.

Necessity of the first condition follows since if (j, r) is recurrent then at least one of the terms $f_{jkrs}^{l-N(s)}$, $f_{kjsr}^{N(s)}$ must be positive, and hence the recurrent state k exists. The second condition is a standard result for recurrent states.

If the switch is of the type described in Section 6 then since

$$\sum_{p=1}^{\infty} g_{ij}^{p} = \sum_{l=1}^{\infty} \sum_{k=1}^{m} f_{kjr}^{l-1}, \quad \text{where} \quad f_{kjr} = f_{kjr}(\infty),$$

the conditions of the theorem may be simplified.

COROLLARY. *For a semi-Markov switch in which the assignment of an arrival is independent of the previous assignment, state j in the r-th output stream is recurrent if and only if there exists some recurrent state k in the arrival process such that $k \to j$ and $f_{kjr} > 0$.*

The proof of this is similar to that of the theorem and is ommitted.

Acknowledgement. We wish to thank Dr. Peter Cherry who was instrumental in the development of early drafts of this paper. This final version was completed while Dr. McNickle was Visiting Professor at Virginia Polytechnic Institute and State University.

References

[1] L. B. Anderson, *Filtered semi-Markov processes*, Masters Thesis, Northwestern University 1967.

[2] E. Cinlar, *Decomposition of a semi-Markov process under a Markovian rule*, Austral. J. Statist. 8 (1966), p. 163-170.

[3] –, *Decomposition of a semi-Markov process under a state dependent rule*, SIAM J. Appl. Math. 15 (1967), p. 252-263.

[4] –, *Markov renewal theory*, Adv. in Appl. Probab. 1 (1969), p. 123-187.

[5] R. L. Disney, *Random flow in queueing networks*, AIIE Trans. 7 (1975), p. 268-288.

[6] – and W. P. Cherry, *Mathematical methods in queueing theory*, Lecture Notes in Economics and Mathematical Systems, no. 98, Springer-Verlag, Berlin 1974.

[7] W. K. Hall, *A queueing theoretic approach to the allocation and distribution of ambulances in an urban area*, Doctoral Thesis, University of Michigan 1969.

[8] – and R. L. Disney, *Finite queues in parallel with a generalized channel selection rule*, J. Appl. Probab. 8 (1971), p. 413-416.

[9] A. Y. Khinchin, *Mathematical methods in the theory of queues*, Methuen, New York 1960.

[10] R. R. Serfozo, *Functions of semi-Markov processes*, SIAM J. Appl. Math. 20 (1971), p. 530-535.

DEPARTMENT OF INDUSTRIAL ENGINEERING AND OPERATIONS RESEARCH
VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
BLACKSBURG, VIRGINIA

Received on 2. 9. 1983;
revised version on 13. 3. 1984