Optimal Control of Markov Regenerative Processes

András Pfening, Miklós Telek
Department of Telecommunications
Technical University of Budapest, 1521 Budapest, Hungary
{pfening,telek}@hit.bme.hu

ABSTRACT

In the paper the integration of available results on Semi-Markov Decision Processes and on Markov Regenerative Processes is attempted, in order to define the mathematical framework for solving decision problems where the underlying structure state process is a Markov Regenerative Process, referred to as Markov Regenerative Decision Process.
The essential question investigated here is which description of Markov Regenerative Processes is needed to analyze the investigated decision model.

Key words: *Markov Decision Process, Stochastic Reward Models, Markov Regenerative Processes.*

1. INTRODUCTION

Decision processes have been widely recognized as a useful tool to provide optimal behaviour of stochastic systems. Optimal decisions are usually hard to predict by "simple engineering considerations" even for simple models, hence an accurate computer aided analysis of possible decisions is needed to obtain decision rules that lead to a sort of optimality.
Previous results on decisions in Discrete Time Markov Chains (DTMC) Continuous Time Markov Chains (CTMC) and Semi-Markov Processes (SMP) [8, 4] provide the analysis of the infinite time horizon problem with and without discounting. The case of finite time horizon was analyzed for DTMCs [8, 4], and for CTMCs [7]. The analysis of decision processes with memoryless (Markov) property at the decision instances is based on the reward analysis of the subprocesses, referred to as subordinated processes, between the consecutive decision instances. It turned out that the optimal decision in an infinite time horizon problem requires the analysis of the mean accumulated reward and the mean time between consecutive decision instances irrespective of discounting.

The most frequently applied stochastic modeling technique is the Markovian approach, which is based on the memoryless (Markov) property of the system behaviour. Nevertheless, this property and its consequence, the exponentially distributed event times in case of time homogeneous system behaviour, have been recognized as one of the main restrictions in the application of Markovian models [3]. An alternative non-Markovian modeling approach is based on the Markov renewal theory [2] and therefore the application of Markov Regenerative Processes (MRGP) in stochastic modeling received an increasing attention recently. The automated generation of such models by non-Markovian Stochastic Petri Nets [1] as well as new results on their transient and steady state analysis increased the applicability of this modeling framework.
The analysis of some reward measures of MRGPs was considered in [6], where the mean of the reward accumulated up to time t, the mean instantaneous reward rate and its limiting behaviour are evaluated without discounting based on the widely used kernels representation of MRGPs, i.e. based on the global and the local kernels [5]. But the kernels representation of MRGPs, does not contain enough information about the stochastic process to evaluate more sophisticated reward measures such as the higher moments or the distribution of the accumulated reward [9].
Hence the main problem investigated in this paper is to find the proper description of MRGPs which allows us to find the optimal decisions in the considered class of Markov Regenerative Decision Processes (MRDP).
The rest of the paper is organized as follows. Section 2 briefly summarize the result on MRGPs and decision models. Section 3 gives the analysis of MRDPs. An application example is studied in Section 4, and the paper is concluded in Section 5.

2. MODEL DESCRIPTION

Short Overview of Markov Regenerative Processes

For a detailed study of MRGPs we recommend [5], here

we only briefly summarize the main results that are referred to in the forthcoming sections.

A continuous time stochastic process $(Z(t))$ enjoys the *Markov property* (or Markov regenerative property) at time T if for any $0 < t_1 < t_2 < \ldots < t_k$ and $x_1, x_2, \ldots, x_k$:

$$P\{Z(T+t_1) \leq x_1, Z(T+t_2) \leq x_2, \ldots, Z(T+t_k) \leq x_k \mid Z(\tau), 0 \leq \tau \leq T\} =$$

$$P\{Z(T+t_1) \leq x_1, Z(T+t_2) \leq x_2, \ldots, Z(T+t_k) \leq x_k \mid Z(T)\}$$

The MRGP $\{Z(t) \in \Omega,\ t \geq 0\}$ does not have the Markov property in general, but there is a sequence of embedded (random) time points $(T_0 = 0, T_1, ..., T_n, ...)$ such that the process at these time points satisfies the Markov property. These time points are the Markov regeneration epochs.

The transient analysis of state probabilities of time homogeneous MRGPs is usually based on the following conditional probabilities:

$$K_{ij}(t) = \Pr\{Z(T_1) = j,\ T_1 \leq t \mid Z(0) = i\},$$
$$E_{ij}(t) = \Pr\{Z(t) = j, T_1 > t \mid Z(0) = i\}. \quad (1)$$

The matrix $\mathbf{K}(t) = [K_{ij}(t)]$ is termed the *global kernel* [1] and is the joint conditional probability of the time to the next Markov regeneration and the state right after the next Markov regeneration given the state at the current Markov regeneration. The matrix $\mathbf{E}(t) = [E_{ij}(t)]$ (called the *local kernel*) describes the state transition probabilities of the MRGP between two consecutive Markov regeneration epochs. The matrices $\mathbf{K}(t)$ and $\mathbf{E}(t)$ can be used in computing the transient probability: $V_{ij}(t) = P\{Z(t) = j \mid Z(0) = i\}$.

Let $\mathbf{K} * \mathbf{V}(t)$ denote a matrix whose (i,j)th element is

$$[\mathbf{K} * \mathbf{V}(t)]_{ij} = \sum_u \int_0^t dK_{iu}(x) V_{uj}(t-x).$$

Then the matrix of transient probabilities $\mathbf{V}(t) = [V_{ij}(t)]$ satisfies the Markov renewal equation [2]: $\mathbf{V}(t) = \mathbf{E}(t) + \mathbf{K} * \mathbf{V}(t)$. $\mathbf{V}(t)$ can be expressed in closed form in Laplace-Stieltjes (LST) transform domain: $\mathbf{V}^{\sim}(s) = [\mathbf{I} - \mathbf{K}^{\sim}(s)]^{-1} \mathbf{E}^{\sim}(s)$.

Note that several other simple measures of MRGPs, such as the sojourn time in a state or in a group of states etc., can not be evaluated based on $\mathbf{E}(t)$ and $\mathbf{K}(t)$, since the evolution of the process between the consecutive Markov regeneration epochs is not "completely" defined, only the transient state probabilities are described by $\mathbf{E}(t)$. The Markov regenerative property of the state transition probabilities is utilized in the above mentioned results. In the following section we propose an analysis approach based on the Markov regenerative property of the studied reward measures.

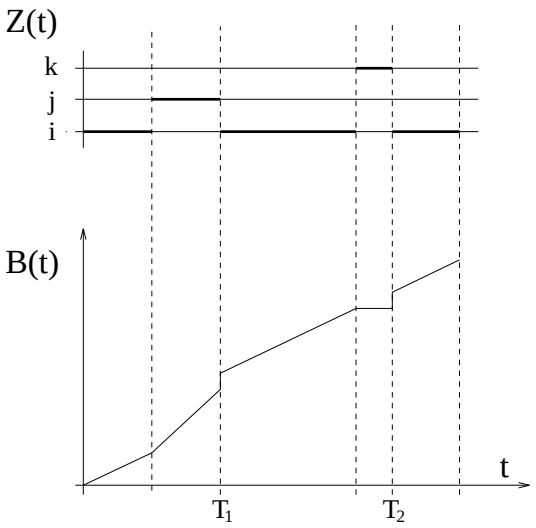

Figure 1. Cost accumulation of a Markov regenerative process

Decision Model

In our study the system is described by the stochastic process $Z(t)$, $t \geq 0$, defined over the finite discrete set of states Ω. Let $Z(t)$ be the Markov regenerative process defined by the kernel functions $\mathbf{E}(t)$ and $\mathbf{K}(t)$. In the case of semi-Markov process based decision processes, at each state transition instant an action is chosen. When the underlying structure state process is a Markov regenerative process, only a subset of the state transition instants are regenerative, thus we restricted our attention to processes where the decision points (i.e. where an action should be taken) are the regenerative time instants.

We shall now define the cost structure that is assigned to state transitions and sojourn times. If action a is taken in state i, it implies an impulse cost of $C(i,a)$. We assign a cost rate $r(i)$ to the states $(i \in \Omega)$ as well. Let us point out that by this cost structure the rate based cost accumulation is defined in each of the states, while impulse cost can be collected only in the decision points, i.e. at transitions to regenerative states. Figure 1 illustrates the cost accumulation of a Markov regenerative process. As it is demonstrated transitions to state i define regenerative instants, whence impulse cost is accumulated (instants T_1 and T_2). The rest of the state transitions do not define regenerative instants, while are allowed to happen.

If action a is taken in state i, the next state of the system and the distribution of the sojourn time is chosen accord-

ing to the following distributions:

$$p_{ij}(a) = \Pr\{Z(T_1) = j \,|\, Z(0) = i, a\},$$

$$H_{ij}(t \,|\, a) = \Pr\{T_1 \le t \,|\, Z(T_1) = j, Z(0) = i, a\}.$$

We shall define which action should be taken in the decision points. The answer is given in a set of decision rules, which is called a *policy*. Let the set of possible actions be the finite set A. Let the individual actions be denoted by a, $a \in A$. The policy chooses a decision a in each of the decision situations. We shall limit our scope to stationary processes, i.e. the action taken depends only on the state the process stays in.

Generally speaking our task is to find a policy (or policies) that is optimal in some sense. The optimality shall mean that the expected accumulated cost, or the cost accumulation rate of the system is minimal.

In the sequel we generalize the theory of semi markov process based decision processes to processes with Markov regenerative structure state process. The results are mostly built on two facts: first, the kernel $\mathbf{Q}(t)$ of semi-Markov processes and the global kernel $\mathbf{K}(t)$ of Markov regenerative processes play almost the same role in the corresponding results. The other major difference is that while all the states of semi-Markov processes must be regenerative, the states of Markov regenerative processes are not necessarily regenerative. This relationship is reflected in the usage of matrices $\mathbf{Q}(t)$ and $\mathbf{E}(t)$ defined for semi markov processes and Markov regenerative processes, respectively.

3. MARKOV REGENERATIVE DECISION PROCESSES

Infinite Time Horizon with Discounting

In this section we focus on processes where the time is not limited, and we apply discounting in the cost structure. The discounting scheme is the one usually applied in decision processes, i.e. the cost C that occurs at time t is equivalent to the cost $Ce^{\alpha t}$, incurring at time $t = 0$, where $\alpha > 0$ is the discount rate. Our task is to determine a policy that minimizes the expected cost accumulated during the whole life of the process. We introduce the notation $B_\alpha(t)$ for the accumulated discounted cost up to time t.

We shall define now the expected cost. Let Z_n be the nth state and a_m the mth decision of the process $Z(t)$, respectively. Furthermore let τ_n be the time between the $(n-1)$th and the nth transition, with $\tau_0 = 0$. The index of transitions and decisions are distinguished since decisions are not necessarily made only at regenerative state transitions.

Let $\Theta \subset \Omega \times \Omega$ be the set of state transitions, with $\Theta_R \subset \Theta$ being the set of transitions leading to regenerative states. With the above notation the expected discounted cost as a function of the initial state applying policy f and discount rate α, starting the process from state i

$$V_f(i) = \mathbf{E}_f\left[\sum_{n=0}^{\infty} e^{-\alpha(\tau_1+\cdots+\tau_{n-1})}\left(I_{\{(Z_{n-1},Z_n)\in\Theta_R\}}C(Z_n,a_n) + \int_0^{\tau_n} r(Z_n)e^{-\alpha t}\,dt\right) \,|\, Z_1 = i\right]. \tag{2}$$

The optimal cost function is defined as the cost function that is minimal in each state:

$$V(i) \hat{=} \min_f V_f(i), \quad \forall i \in \Omega.$$

The policy f is *α-optimal* if it yields the optimal cost function:

$$V_f(i) = V(i), \quad \forall i \in \Omega.$$

Now we shall show how to determine the expected cost function when policy f is applied in the process.

Theorem 3.1 *Applying policy f the expected discounted cost function V_f satisfies the following equation:*

$$V_f(i) = C(i, f(i)) + \sum_{k\in\Omega} r(k)E^*_{ik}(\alpha) + \sum_{j\in\Omega} V_f(j)K^{\sim}_{ij}(\alpha),$$

where $\mathbf{E}^*(\alpha)$ and $\mathbf{K}^{\sim}(\alpha)$ are the Laplace and Laplace-Stieltjes transforms of the local and global kernel functions of policy f, at $s = \alpha$.

Proof: Let us denote $\mathbf{E}[B_\alpha(T_1) \,|\, \tau, i, j] = \mathbf{E}[B_\alpha(T_1) \,|\, T_1 = \tau, Z(0) = i, Z(T_1) = j]$.

From the definition of the process we describe

$$V_f(i) = C(i, f(i)) + \sum_{j\in\Omega} p_{ij}(f(i)) \int_{\tau=0}^{\infty} \mathbf{E}[B_\alpha(T_1) \,|\, \tau, i, j] + V_f(j)e^{-\tau\alpha}\,dH_{ij}(\tau).$$

The above equation can be separated as

$$\begin{aligned} V_f(i) &= C(i, f(i)) + \sum_{j\in\Omega} p_{ij}(f(i)) \int_{\tau=0}^{\infty} \mathbf{E}[B_\alpha(T_1) \,|\, \tau, i, j]\,dH_{ij}(\tau) \\ &\quad + \sum_{j\in\Omega} p_{ij}(f(i))V_f(j)H^{\sim}_{ij}(\alpha), \end{aligned} \tag{3}$$

from which applying the definition of $H_{ij}(\tau)$

$$\begin{aligned} V_f(i) &= C(i, f(i)) + \sum_{j\in\Omega} p_{ij}(f(i))\mathbf{E}[B_\alpha(T_1) \,|\, Z(0) = i, Z(T_1) = j] \\ &\quad + \sum_{j\in\Omega} p_{ij}(f(i))V_f(j)H^{\sim}_{ij}(\alpha), \end{aligned}$$

then applying the definition of $p_{ij}(f(i))$

$$V_f(i) = C(i, f(i)) + \mathbf{E}[B_\alpha(T_1) \mid Z(0) = i] + \sum_{j \in \Omega} V_f(j) K_{ij}^{\sim}(\alpha).$$

We shall now determine the value of $\mathbf{E}[B_\alpha(T_1) \mid Z(0) = i]$. Since discounting is applied in the cost structure we shall write

$$\mathbf{E}[B_\alpha(T_1) \mid Z(0) = i] = \mathbf{E}\left[\sum_{k \in \Omega} r(k) \int_{\tau=0}^{\infty} I_{\{Z(\tau)=k, \tau<T_1\}} e^{-\alpha\tau}\, d\tau \mid Z(0) = i\right],$$

the expected value operator can be moved as

$$\mathbf{E}[B_\alpha(T_1) \mid Z(0) = i] = \sum_{k \in \Omega} r(k) \int_{\tau=0}^{\infty} \mathbf{E}\left[I_{\{Z(\tau)=k, \tau<T_1\}} \mid Z(0) = i\right] e^{-\alpha\tau}\, d\tau.$$

The value of the indicator function gives the corresponding probability, thus

$$\mathbf{E}[B_\alpha(T_1) \mid Z(0) = i] = \sum_{k \in \Omega} r(k) \int_{\tau=0}^{\infty} \Pr\{Z(\tau) = k, \tau < T_1 \mid Z(0) = i\} e^{-\alpha\tau}\, d\tau,$$

recognizing the definition of the $\mathbf{E}(t)$ local kernel function

$$\mathbf{E}[B_\alpha(T_1) \mid Z(0) = i] = \sum_{k \in \Omega} r(k) \int_{\tau=0}^{\infty} E_{ik}(\tau) e^{-\alpha\tau}\, d\tau = \sum_{k \in \Omega} r(k) E_{ik}^*(\alpha).$$

□

As a consequence of Theorem 3.1, if we enumerate all the possible policies, we can determine the corresponding cost function, thus the optimal policy can be determined (if it exists). However we should find more effective ways to get the optimum. In order to do this, we should now define an important property of the optimal cost function.

Theorem 3.2 *The optimal cost function V satisfies the following equation:*

$$V(i) = \min_a \left\{ C(i, a) + \sum_{k \in \Omega} r(k) E_{ik}^*(\alpha \mid a) + \sum_{j \in \Omega} V(j) K_{ij}^{\sim}(\alpha \mid a) \right\}. \quad (4)$$

Proof: The proof can follow the idea of the proof of Theorem 7.1 of [8]. □

Now the results of the semi-Markov decision theory corresponding to infinite time horizon and discounting can easily be generalized to Markov regenerative processes, applying the defined

$$\overline{C}_\alpha(i, a) = C(i, f(i)) + \sum_{k \in \Omega} r(k) E_{ik}^*(\alpha)$$

average transition cost and using the global kernel function $\mathbf{K}(t)$ instead of the kernel $\mathbf{Q}(t)$.

We will proceed the same way as it is done in [8]. First we define a mapping of bounded functions, and will show some properties of the defined mapping. Let $B(\Omega)$ be the bounded (real-valued) functions defined over the state space Ω. We define for each f policy the mapping $M_f : B(\Omega) \to B(\Omega)$ as

$$(M_f u)(i) = \overline{C}_\alpha(i, f(i)) + \sum_{j \in \Omega} p_{ij}(f(i)) K_{ij}^{\sim}(\alpha \mid f(i)) u(j),$$

where $\mathbf{K}(t)$ is the global kernel function when policy f is applied.

Lemma 3.3 *For all functions $u, v \in B(\Omega)$ and f policy:*

1. $u \leq v \Rightarrow M_f u \leq M_f v$,
2. $M_f V_f = V_f$,
3. $M_f^n u \to V_f$ *for all function* $u \in B(\Omega)$.

Proof: The proof can follow the idea of the proof of Theorem 7.2 of [8]. □

By point 3 of Lemma 3.3 another method can be identified to determine the expected cost function, since for any bounded function u the function series $M_f^n u$ will converge to the expected cost function V_f. In cases when approximate results for the cost function is satisfactory this method can be useful.

Now we determine the optimal policy knowing the optimal expected cost function. Let f_α be the policy that in state i makes the decision that minimizes the rhs of Equation (4):

$$f_\alpha(i) \hat{=} \arg\min_a \left\{ \overline{C}_\alpha(i, a) + \sum_{j \in \Omega} V(j) K_{ij}^{\sim}(\alpha \mid a) \right\}, \quad i \in \Omega.$$

Theorem 3.4 *Policy f_α is α-optimal, i.e.*

$$V_{f_\alpha}(i) = V(i), \quad i \in \Omega.$$

Proof: The proof can follow the idea of the proof of Theorem 6.3 of [8].

Applying the mapping M_{f_α} to the minimal cost function V:

$$(M_{f_\alpha}V)(i) = \overline{C}_\alpha(i, f_\alpha(i)) + \sum_{j\in\Omega} V(j)K_{ij}^{\sim}(\alpha \,|\, f_\alpha(i)) =$$

$$\min_a \left\{ \overline{C}_\alpha(i,a) + \sum_{j\in\Omega} V(j)K_{ij}^{\sim}(\alpha \,|\, a) \right\} = V(i).$$

The second transform was done making use of the definition of f_α, finally we applied Theorem 3.2. While $M_{f_\alpha}V = V$, by induction $M_{f_\alpha}^n V = V$, from which $V_{f_\alpha} = V$ applying point 3 of Lemma 3.3. □

As a consequence there exist an α-optimal policy, which is in addition stationary, since its decisions are only dependent on the actual state of the process. If we can determine the optimal expected cost function V, then applying Theorem 3.4 the optimal policy can also be determined.

Policy Iteration

The method of policy iteration simply tries to improve an existing policy by changing some of its decisions. A policy is called better than another one if the value of the expected cost function resulted by the policy is not higher starting the process in any of the initial states, and is less in case of at least one state. Let us now examine the relationship of policy f (with expected cost function V_f) and of policy f^* (with expected cost function V_{f^*}), derived from policy f the following way:

$$f^*(i) \hat{=} \arg\min_a \left\{ \overline{C}_\alpha(i,a) + \sum_{j\in\Omega} V_f(j)K_{ij}^{\sim}(\alpha \,|\, a) \right\}. \quad (5)$$

We claim that $V_{f^*}(i) \leq V_f(i)$ for all state $i \in \Omega$. The proof can be derived as it is done of Corollary 6.4 in [8]. Therefore we start from an arbitrary initial cost function, then determine the improved policy by (5), determine the corresponding expected cost function and continue the same process until we cannot improve the policy in any of the possible initial states. Applying Theorem 3.2 we can see that this way we have found the optimal expected cost function.

Successive Approximation

The method of successive approximations determines directly the optimal expected cost function V, from which the optimal policy can be determined by applying Theorem 3.4.

Let us define the mapping $M_\alpha : B(\Omega) \to B(\Omega)$ as

$$(M_\alpha u)(i) = \min_a \left\{ \overline{C}_\alpha(i,a) + \sum_{j\in\Omega} K_{ij}^{\sim}(\alpha \,|\, a)u(j) \right\}.$$

We claim that M_α is a contraction mapping. Then applying the definition of M_α and Theorem 3.2 the fix point of the mapping must be the optimal cost function. Furthermore since the fix point is unique, we proved that the optimal cost function is unique as well.

We shall finally draw two conclusions. First, the global and local kernels are shown to be sufficient for the analysis of MRDPs, e.g. the information they contain about the process is sufficient. Second, the use of the Laplace-Stieltjes transforms makes relatively simple the whole analysis, reducing the symbolic derivation to the expression of the kernel matrices.

Infinite Time Horizon without Discounting

We also have investigated the situation when no discounting is applied in the cost structure ($\alpha = 0$). Our task is again to determine the policy that minimizes the expected cost. We found out out, that most of the results corresponding to semi-Markov processes can be generalized to Markov Regenerative Processes as well.

Unlike with discounted costs the uniqueness of the optimal policy could not be shown, only the existence of it. Neither we have the method of policy iteration nor of successive approximation to find the optimal policy, the only possibility left is the exhaustive search of all the policies. We should notice however, that in case of unlimited time, and without discounting the expected cost will very rarely be bounded. In most of the cases the expected cost will grow infinitely, for which cases the aim of the analysis should be modified to find the policy, which ensures the slowest cost increase rate. It can be shown that the procedure described by Howard [4] to find the policy that is optimal in this sense can be generalized to Markov Regenerative Processes.

Let us again emphasize that the global and local kernels are sufficient for the analysis of MRDPs even with finite time horizon, when the decisions are not dependent on the time. The time dependent decision analysis is out of the scope of the current paper. The use of the Laplace-Stieltjes transforms makes again simpler the whole analysis, reducing the symbolic derivation to the expression of the kernel matrices.

Unfortunately the page limit does not allow us to present the precise definitions and proofs here, the interested reader is referred to [10].

Finite Time Horizon

For finite time horizon systems first we determine the cost accumulated by policy f during time t.

Theorem 3.5 *Using policy f, with discount rate α, having kernel functions $K(t)$ and $E(t)$, the cost accumulated in time t can be expressed by the following transform domain equation:*

$$\underline{V}_f^{\sim}(s) = (\mathbf{I} - \mathbf{K}^{\sim}(s+\alpha))^{-1}\Big(\underline{C}_f + \mathbf{E}^*(s+\alpha)\underline{r}\Big). \quad (6)$$

Proof: The process is handled as two separate, stochastically identical processes, differing only in the cost accumulation structure. One of the processes collects cost

only by the cost rates, while the other collects the impulse costs in the regeneration time instants.
For the first process the following considerations can be made:

$$\begin{aligned} V_f(i,t) &= \mathbf{E}[B(t) \mid Z(0)=i] \qquad (7)\\ &= \mathbf{E}\left[\sum_{k\in\Omega} r(k) \int_{\tau=0}^{t} I_{\{Z(\tau)=k\}} e^{-\alpha\tau}\, d\tau \mid Z(0)=i\right] \\ &= \sum_{k\in\Omega} r(k) \int_{\tau=0}^{t} \mathbf{E}\left[I_{\{Z(\tau)=k\}} \mid Z(0)=i\right] e^{-\alpha\tau}\, d\tau \\ &= \sum_{k\in\Omega} r(k) \int_{\tau=0}^{t} \Pr\{Z(\tau)=k \mid Z(0)=i\} e^{-\alpha\tau}\, d\tau \\ &= \sum_{k\in\Omega} r(k) \int_{\tau=0}^{t} U_{ik}(\tau) e^{-\alpha\tau}\, d\tau, \end{aligned}$$

where $\mathbf{U}(t)$ is the matrix of the transition probabilities. The Laplace-transform of the expression:

$$\begin{aligned} V_f^*(i,s) &= \sum_{k\in\Omega} r(k) \int_{t=0}^{\infty} e^{-st} \int_{\tau=0}^{t} U_{ik}(\tau) e^{-\alpha\tau}\, d\tau\, dt \\ &= \sum_{k\in\Omega} r(k) \int_{\tau=0}^{\infty} \frac{1}{s} e^{-\alpha\tau} e^{-st} U_{ik}(\tau)\, d\tau \\ &= \sum_{k\in\Omega} r(k) \frac{1}{s} U_{ik}^*(s+\alpha), \end{aligned}$$

which can be transformed making use of the defined vectors:

$$\underline{V_f^{\sim}}(s) = \mathbf{U}^*(s+\alpha)\underline{r} = \frac{1}{s+\alpha}(\mathbf{I} - \mathbf{K}^{\sim}(s+\alpha))^{-1}\mathbf{E}^{\sim}(s+\alpha)\underline{r}.$$

The second process is equivalent to a semi-Markov process with global kernel function $\mathbf{K}(t)$ from the viewpoint of the cost accumulation, thus the accumulated cost can be calculated applying the results of Howard ([4], Chapter 13.2), substituting zero cost rates. The sum of the above quantities gives the statement of the theorem. □
For the processes without discounting the above theorem yields the result by substituting $\alpha = 0$ in Equation (6). For not limited time horizon cases the result can be get without inverse transformation by applying the final value theorem, i.e. in Equation (6) we take the limit $s \to 0$.
To determine the optimal policy exhaustive search should be applied.
Let us again emphasize that the global and local kernels are sufficient for the analysis of MRDPs even with finite time horizon, when the decisions are not dependent on the time. The time dependent decision analysis is out of the scope of the current paper. The use of the Laplace-Stieltjes transforms makes again simpler the whole analysis, reducing the symbolic derivation to the expression of the kernel matrices.

4. CONCLUSION

The problem of optimal decision in MRGPs is studied in this paper. In the considered class of models decisions are allowed only at regeneration instances and an infinite time horizon is considered with and without discounting. It is shown that the widely used kernel representation of MRGPs ($\mathbf{K}(t)$ and $\mathbf{E}(t)$) is sufficient to obtain the optimal decision, and the solution is also provided.

5. REFERENCES

[1] Hoon Choi, V.G. Kulkarni, and K. Trivedi. Markov regenerative stochastic Petri nets. *Performance Evaluation*, 20:337–357, 1994.

[2] E. Cinlar. Markov renewal theory. *Advances in Applied Probability*, 1:123–187, 1969.

[3] R. Fricks, M. Telek, A. Puliafito, and K. Trivedi. Markov regenerative theory applied to performability evaluation. In K.K. Bagchi and G. Zobrist, editors, *Modeling and Simulation of Advanced Computer Systems: Applications and Systems*, State-of-the Art in Performance Modeling and Simulation, chapter 11, pages 193–236. Gordon and Breach Publishers, 1996.

[4] R.A. Howard. *Dynamic Probabilistic Systems, Volume II: Semi-Markov and Decision Processes.* John Wiley and Sons, New York, 1971.

[5] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems.* Chapman Hall, 1995.

[6] D. Logothetis, K. S. Trivedi, and A. Puliafito. Markov regenerative models. In *Int. Computer Performance and Dependability Symposium - IPDS95.* IEEE CS Press, 1995.

[7] H. De Meer, K. S. Trivedi, and M. D. Cin. Guarded repair of dependable systems. *Theoretical Computer Science*, (128):179–210, 1994.

[8] S. M. Ross. *Applied Probability Models with Optimization Applications.* Dover Publications, Inc., New York, 1992.

[9] M. Telek and A. Pfening. Performance analysis of markov regenerative reward models. *Performance Evaluation*, 27&28:1–18, 1996.

[10] András Pfening. Markov Decision Processes with Applications in Telecommunications (in Hungarian). Doctor's thesis, Technical University of Budapest, 1996.