# Solving large structured Markov Decision Problems for perishable inventory management and traffic control

Haijema, R.

**Publication date**
2008

# Summary in English

## Introduction

Quite a number of optimization problems arising in practice involve sequential decision making and can be formulated as a Markov decision problem (MDP). However, the size of an MDP for many problems that arise in practice is often too large to determine an optimal policy. Fortunately, the problem's state space has some additional structure.

In this thesis we illustrate how to exploit the problem structure for two different applications in order to find good approximate solutions. The first problem that we treat is the production-inventory management of blood platelet pools (BPPs), a typical example of a perishable product with a shelf life of only 4-7 periods. The second problem is the dynamic control of traffic lights. The motivation to study these two problems stems from the fact that both problems are high-dimensional MDPs. The structure of the state space is different, consequently we apply different approaches to the two problems.

## Part I: Chapters 2 to 4 – Inventory management of BPPs

Blood platelets are of life saving importance: as small particles in the blood stream they help repair damaged blood vessels and prevent excessive bleeding. Some patients need to be transfused with platelets to keep the concentration of platelets in their blood stream at a safe level. To meet the uncertain demand at hospitals blood banks produce BPPs, also called thrombocytes concentrates, by pooling platelets from 5 different voluntary donors into a storage bag: the platelet pool. A platelet pool is the most perishable and the most expensive blood product processed at blood banks. BPPs are processed from a side product that is in The Netherlands fortunately plenty available after processing red blood cell concentrates (RBCs). Dutch blood banks usually process BPPs only of the blood groups O and A, since for platelet transfusion the blood groups of most patients are to some degree compatible with other blood groups.

Despite the compatibility of blood groups Veihola et al. [155] report in 2006 European figures for 17 blood centers (spread over 10 European countries) that the waste of BPPs due to outdating is considerable: 7-25% of the BPPs is not transfused because of outdating. In the USA [163] the fraction of outdated BPPs was 18% over 2004 . Outdating is the result of the short shelf life in combination with the uncertain demand and 'sub-optimal' replenishment strategies for inventories held at hospitals and/or blood banks. In The Netherlands the inventory of BPPs is mostly centralized at four blood banks. Shortages due to out-of-stock are to be kept at a low level, but can be resolved by meeting demand with BPPs from another blood bank. We have investigated (nearly) optimal production orders for one of the Dutch blood banks, where production happens 5 days a week (Monday to Friday), and production and laboratory tests take 1 day.

When modeling the problem as an MDP we firstly aggregate all blood groups into a single, say universal blood group. Later this modeling assumption is checked by a simulation study in which we distinguish patients and donors of eight different blood groups according to the ABO-system and the Rhesus-D factor (negative or positive). Depending on regulations by law and clinical studies BPPs have a maximal shelf life of 4 to 7 days, consequently one distinguishes BPPs of 4, 5, 6 or 7 age categories. The state vector of the MDP contains next to the day of the week also the number of BPPs in each age category.

On the one hand the high-dimensionality of the state vector does not allow an analytical approach, due to the complexity of modeling the transitions. On the other hand a numerical approach seems to be doomed to fail due to the large number of states that may occur. To allow numerical approximation of an optimal strategy by stochastic dynamic programming (SDP) we scale the problem by aggregating demand as if it happens in multiples of $k$ pools. When restricting production orders to be also in multiples of $k$ pools, the state space is at a more coarse grid. Consequently, we have to repair the transition law by transforming the demand distributions.

After truncating the state space to reduce the potential number of states even more, an optimal ordering strategy is computed for a given issuing policy. The optimal ordering strategy may be too complicated for practical use by blood bank managers. Therefore we check by simulation whether the optimal strategy resembles a simple rule that performs nearly optimal and that is applicable to the unscaled problem.

For data obtained for one of the Dutch blood banks it appears that a simple ordering rule with fixed order-up-to levels for each working day performs very well: outdating can be reduced to only a few percent or even less than 1 percent depending on the issuing discipline. Shortages occur even less often and the age of the pools upon transfusion may be improved as well. These conclusions are supported by an extensive sensitivity study using different data sets. Outdating is expected to be higher when the problem plays at a much smaller scale, when demand is much more uncertain or when all donor blood is needed to produce BPPs. The combination of SDP and Simulation may still be helpful in finding an approximate solution, as we have shown how more advanced order-up-to rules can be read from frequency tables obtained by Simulation.

Next to a description of the methodology, Chapters 3 and 4 of the thesis offer a detailed validation through a sensitivity analysis using data from a Dutch blood bank. Since production breaks have a great impact on the outdating of BPPs, we also present an SDP-Simulation approach to solve the problem around short holiday breaks, during which production is hampered not only on Saturday and Sunday. Also the ordering of BPPs at a smaller blood bank or a hospital is considered, where set-up costs for ordering platelets may apply. The suggested SDP-Simulation approach can also be adopted to determine ordering quantities at other blood banks and to solve similar problems in other industries that deal with perishable products that have a short limiting shelf life. The software developed during this thesis project is made more user-friendly and named TIMO (for Thrombocytes Inventory Management Optimizer). TIMO currently runs at Sanquin's blood bank division South East.

## Part II: Chapters 5 to 8 – Dynamic control of traffic lights

Traffic lights are introduced to make road traffic safer and more efficient by controlling the waiting time of car drivers. The problem of minimizing the long-run average waiting time over all $F$ traffic flows at a single intersection can be formulated as a Markov decision problem with state description, next to the state of the lights, the number of cars present at each queue: $\mathbf{q} = (q_1, q_2, \ldots, q_F)$.

All cases that we consider are under-saturated cases: in the long-run all traffic can be handled by the intersection. The potential number of states grows exponentially in the number of queues ($F$). Solving the MDP requires approximate solutions when the intersection consists of a larger number of queues. When the problem of a single intersection requires an approximate solution, then one certainly has to rely on approximations for the control of a network of intersections.

Whereas in the previous application we were looking for problem structure under an optimal policy as computed after scaling, we now add structure to the problem by imposing a well-structured policy: the fixed cycle control (FC). Under FC queues get green in a fixed order and the effective green times are fixed. The state of the traffic lights is thus simply the position in the cycle or the time slot. By adjusting the position in the cycle the green periods of the flows are extended or ended dynamically based on the number of cars queued. Therefore we need to know the relative appreciation or relative value of each time slot and for each possible $(q_1, q_2, \ldots, q_F)$. The state space under FC can be decomposed, such that one can evaluate each traffic flow in isolation of the others. The relative values of the states occurring under FC can be computed by summing the related relative values of the states for each flow. By a one-step policy improvement algorithm we obtain an improved strategy that hopefully performs nearly optimal. The improved strategy is called the RV rule (with RV for relative value).

The RV rule is put to the test and compared against basic control policies such as FC and exhaustive control. Therefore a simulation study is executed with different intersections under varying circumstances. The results are promising: in all (fully) symmetric cases considered the long-run average waiting time over all flows is reduced by at least 20 percent and in some cases even more than 70 percent compared to FC and (pure) exhaustive control (see Chapter 6).

Modern technology allows not only to track the number of cars at each queue, but it also detects the position of the cars approaching the intersection. Estimates of the expected arrival times of these cars can be included in the state description. The state space of each flow consists now of the state of the traffic light, the number of cars waiting at it's queue and a vector containing the arrival information. Assuming in the decision model that cars proceed at a constant and identical speed and that they do not change lanes, the RV rule can be easily extended to include arrival information. Some test by simulation show that including arrival information over only a few slots, say the next 5 slots, shows already an improvement of the control rule (see Chapter 7). Even when car speeds differ in the simulation model, the RV rule gives good results.

To solve the problem for a network of intersections (see Chapter 8), we decompose the control of the network by controlling the intersections in isolation. Starting with a good FC for each intersection in isolation, decisions are taken locally based on the RV rule. The synchronization of the local decisions at the intersections may be partly accomplished by using a few slots of arrival information. A more rigid synchronization can be achieved under FC by setting so-called offsets: in some to-FC-ideal cases green waves may be set in multiple directions (to maximize progression). An optimal synchronized FC is in general hard to find, but may result in low long-run average waiting times over all flows. For the RV rule we need just a locally-good FC, not necessarily an optimal synchronized one.

In a detailed simulation model in which cars may drive at constant or different speeds and queued cars take a length of 7 meters each, we put the resulting RV policies with arrival information to the test for several network cases. Although, the average waiting times under synchronized FC can be very low when green waves apply, the RV rule with arrival information yields lower overall long-run average waiting times in all cases that we have studied. In the network cases with nine intersections that we have considered, the waiting time under the synchronized FC is still more than 40% above that of the RV rule with arrival information.

Adding structure to the problem by starting with a good FC is thus very beneficial in order to get to an improved approximate solution.

## Conclusion

In the thesis we have studied two applications that may look so different, but that have in common that the underlying MDP is too large to solve. For both applications an approximate decision model is created and tested by (discrete) Simulation.

For the platelet production-inventory problem we have exploited the problem structure as present in the Dutch case: first blood groups are aggregated, next the pools are aggregated into batches to down-scale the problem and finally it appears that the age-categories of the pools can be aggregated to derive a simple rule for practical use. After re-scaling, this simple rule can be used in practice. For the dynamic control of traffic lights we suggest a one-step policy improvement algorithm. This algorithm can be applied only to a well-structured policy for which the relative values of the states can be computed. Under FC control the state space is well-structured and can be decomposed such that relative values can be computed for each flow in isolation. The control of a network of intersection also relies on decomposition.

Both decision models arose by exploiting the problem structure and rely on numerical techniques. For the platelet production-inventory problem we have looked for structure in the problem and data under consideration. For the dynamic control of traffic lights we have structured the problem by imposing a well-structured strategy for further improvement. For both applications simulation models are developed that contain more realistic features than can be incorporated in the decision models. Results show that the decision models perform very well, at least for the cases that are studied. Outdating of platelet pools can be reduced significantly and the waiting time at signalized intersections can be much lower than under the selected benchmark heuristics.

MDPs can thus be approximated successfully when exploiting the structure of the problem to come to approximate solutions. A uniform way of tackling high-dimensional MDPs may not exist, hence each problem requires a problem specific approach. This thesis helps in gaining insight in how to do so.