

## Research Article

### A Fast Algorithm for Large-Scale MDP-Based Systems in Smart Grid

<sup>1</sup>Hua Xiao, <sup>1</sup>Huaizong Shao, <sup>1</sup>Fan Yang, <sup>1,2</sup>Yingjie Zhou and <sup>1</sup>Qicong Peng

<sup>1</sup>School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu, China

<sup>2</sup>Department of Electrical Engineering, Columbia University, New York, USA

**Abstract:** In this study, we investigate the fast algorithms for the Large-Scale Markov Decision Process (LSMDP) problem in smart grid. Markov decision process is one of the efficient mathematical tools to solve the control and optimization problems in wireless smart grid systems. However, the complexity and the memory requirements exponentially increase when the number of system state grows in. Moreover, the limited computational ability and small size of memory on board constraint the application of wireless smart grid systems. As a result, it is impractical to implement those LSMDP-based approaches in such systems. Therefore, we propose the fast algorithm with low computational overhead and good performance in this study. We first derive the factored MDP representation, which substitutes LSMDP in a compact way. Based on the factored MDP, we propose the fast algorithm, which considerably reduces the size of state space and remains reasonable performance compared to the optimal solution.

**Keywords:** Factored MDP, fast algorithm, large-scale MDP, smart grid, wireless communication

## INTRODUCTION

Smart grid (Li *et al.*, 2010; Moghe *et al.*, 2012; Moslehi *et al.*, 2010) is a promising technology to improve the efficiency and reliability of the generation, transmission and distribution of electricity services. Compared to the traditional power grid, it introduces a two-way communication network to assist control and monitor the electric energy movements. The information communication mechanism enhances reliably and efficiently control of suppliers and consumers in power grids. Due to the time-varying nature of wireless channel and the complicated actions in smart grid, Markov decision process becomes an ideal mathematical tool to model and solve related problems in such scenarios. In this study, we focus on the MDP-based control and optimization problems in wireless smart grid systems.

In what follows, we survey recent works on the MDP-based approaches in wireless communications. Fu and Schaar (2010) propose a multi-user Markov decision process by considering the video traffic characteristics and the time-varying network conditions. A decomposition method is proposed to enable each wireless user can individually solve its own local MDP problem. Niyato *et al.* (2009) study the QoS provisioning problem for multiple traffic types in MIMO communications. They model the minimization problem of the weighted packet dropping probability as a constrained Markov decision process, by which the

optimal antenna assignment and admission control is derived. The transmission power and modulation order adaptation strategies are investigated in Karmokar *et al.* (2008). The wireless channel is modeled as a first-order Markov chain. The optimal transmission power and rate policies are proposed base on the semi-Markov decision process. Phan *et al.* (2010) present an energy-efficient transmission policy for wireless sensor networks with a strict energy capacity constraint. The optimal transmission threshold is derived based on the Markov chain model. The condition to perform the vertical handoff in the 4<sup>th</sup> generation wireless communication is studied in Lin and Wong (2008). The maximization problem for the expected total rewards of connection is formulated as MDP.

However, the complexity of the above MDP-based approaches exponentially increases when the number of system states grows in, which causes that the problem cannot be efficiently solved in practice. In smart grid, the wireless systems not only undertake transmission tasks but also implement control actions. Furthermore, the systems normally have limited computation ability and small memory space. Therefore, it is difficult to obtain the control policy based on the large-scale MDP models in such systems. It is necessary to study the fast and efficient algorithms for such problems based on the practical wireless smart grid systems. To this end, our main contributions in this study are as follows. First, we propose the framework of the wireless smart grid system and propose the factored MDP, which

**Corresponding Author:** Hua Xiao, School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu, China

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

represents the large-scale MDP in a compact way. We derive the fast algorithm based on the factored model, which remains good performance compared to the optimal solution.

**System model:** We consider the wireless transmission system with energy harvesting devices in smart grid. Smart meter is one of the key components in smart grid, which is used to control, monitor grids and transmit data. In general, smart meters have a limited battery capacity and it is inconvenient and costly to replace the battery when the smart meters are deployed at substations. Therefore, we introduce the energy harvesting device into the smart meters, which can harvest environmental energy and sustainably provide energy support to smart meters. In addition, smart meters implement real-time monitoring, which results in an instant data delay requirements. The smart meters should have the ability to efficiently control the energy consumption and to meet the instant delay constraint. The basic system model is shown in Fig. 1.

We assume that the time is divided into slots and all system parameters are discretized. The system parameters keep stable during each time slot. Let  $Q_t$  denote the data queue state in time slot  $t$ , where  $0 < Q_t < Q_c$  and  $Q_t \in Q$ . Let  $Y_t$  denote the available energy level, where  $0 < Y_t < Y_c$  and  $Y_t \in Y$ . Let  $g_t$  denote the channel state and  $E_t$  denote the energy harvesting rate in time slot  $t$ , where  $g_t \in G$  and  $E_t \in E$ . Let  $S_t = \{Q_t, Y_t, g_t, E_t\}$  denote the system state in time slot  $t$ , where  $S_t \in S$ . Let  $R_c$  denote the required transmission rate and  $\sigma^2$  denote the noise power. Let  $\Pi$  denote the control policy. The energy control problem is formulated as follows:

$$\begin{aligned} \min_{\Pi} E \left\{ \sum_{t=1}^{\infty} \lambda^{t-1} r(S_t) \right\} \\ \text{s.t.} \quad 0 < Q_t < Q_c \\ \quad \quad 0 < Y_t < Y_c \end{aligned} \quad (1)$$

where the expectation is over the system state and the cost function is given by:

$$r(S_t) = a_t P_t \quad (2)$$

where,  $a_t$  denotes the action in time slot  $t$  and  $A$  denotes the action set, we have:

$$a_t = \{0, 1\} \quad (3)$$

And

$$P_t = (2^{R_c} - 1) \sigma^2 / g_t \quad (4)$$

The problem (1) is a discount infinite MDP problem, which can be solved by value iteration

(Puterman, 1994). However, the large-scale system states make it impossible to solve by the smart meters with only limited computational ability and small memory space. In the next sections, we will propose the factored MDP, which can represent the problem (1) in a compact way.

**Factored MDP:** Factored MDP is a compact representation for the original MDP problem. It is neither a new formulation nor a solution for the problem. However, it has some properties which are beneficial for solving the large-scale MDP problem in a compact way. It is based on the dynamic Bayesian network and can reduce the size of the original MDP model from two aspects. First, it uses state variables to substitute the system states with identical property. Therefore, the system state space is represented by state variables. Second, it adopts dynamic Bayesian network to characterize the transition probability of system states. The dynamic Bayesian network is a directed acyclic graphical network based on time sequence. The transition matrix is sparse in general. That is, the system states in the next time slot are dependent on part of system states in the current time slot. Dynamic Bayesian network can represent the connection between adjacent time slots in a compact way.

Let  $X'_1, X'_2, X'_3$  and  $X'_4$  denote the system variables corresponding to the system state  $Q_t, Y_t, g_t$  and  $E_t$ , respectively. Therefore, we have:

$$\begin{cases} X'_1 \in Q \\ X'_2 \in Y \\ X'_3 \in G \\ X'_4 \in E \end{cases} \quad (5)$$

Let  $\{\text{Parent}(X'_1), \text{Parent}(X'_2), \text{Parent}(X'_3), \text{Parent}(X'_4)\}$  denote the parents of  $\{X'_1, X'_2, X'_3, X'_4\}$ . The transition graph of the dynamic Bayesian network is a two-layer directed acyclic graph as shown in Fig. 2.

The system state  $S_t = \{Q_t, Y_t, g_t, E_t\}$  is substituted by four variables  $\{X'_1, X'_2, X'_3, X'_4\}$ . In Fig. 1, we have  $\text{Parent}(X'_1) = \{X_1, X_2\}$ . This is because the data queue dynamic depends on the previous queue state and the previous available energy level. Each node  $X'_i$  is associated with a conditional probability distribution  $P(X'_i | \text{Parent}(X'_i))$ . In Fig. 1,  $h_i$  denotes the basis of the value function. Let  $V(s)$  denote the value function of the system state  $s \in S$ . For the optimal policy  $\Pi^*$ , we have:

$$V^*(s) = r(s, \Pi^*(s)) + \sum_{s'} p(s' | s, \Pi^*(s)) V^*(s') \quad (6)$$

where,  $s' \in S$ . Let  $V_s$  denote  $V(s)$  in (6). The MDP problem can convert to the following linear programming:

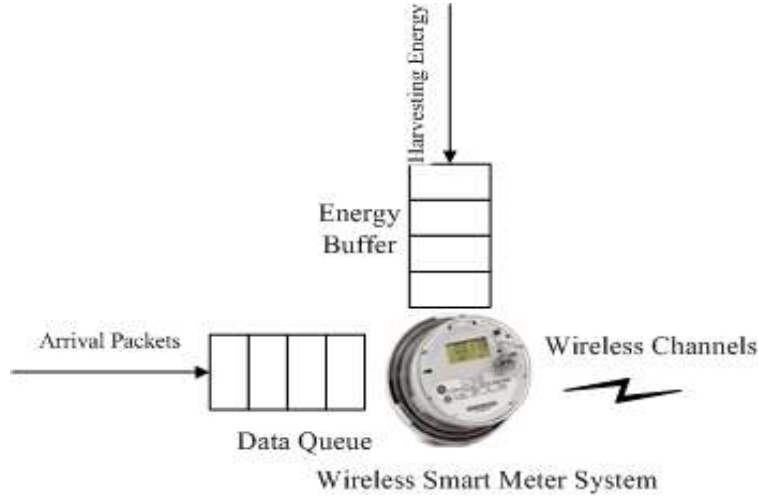


Fig. 1: The wireless smart meter system

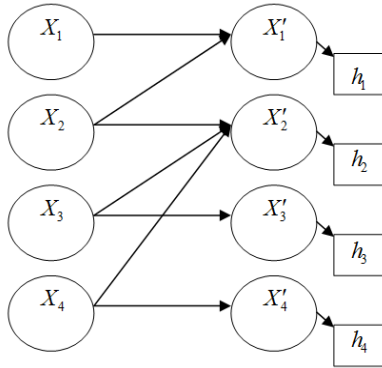


Fig. 2: The representation of the factored MDP

$$\begin{aligned}
 & \min_{V_{s_i}} \sum_{s_i} \alpha(s_i) V_{s_i} \\
 & s.t. V_{s_i} \geq r(s_i, a) + \gamma \sum_{s_j} p(s_j | s_i, a) V_{s_j} \\
 & \forall s_i, s_j \in S, a \in A
 \end{aligned} \tag{7}$$

where  $\alpha(s)$  denotes the state relevance weight, which should be positive. First, we factor the value function and simplify the objection function. Assume that the value function can be approximated by the subspace which is spanned by the basis function  $\{h_1(x), h_2(x), \dots, h_k(x)\}$ . We have:

$$V_{s_i} = \sum_k w_k h_k(s_i) \tag{8}$$

Plug the above equation into the problem (7). We have:

$$\begin{aligned}
 & \min \sum_{s_i} \alpha(s_i) \sum_k w_k h_k(s_i) \\
 & s.t. \sum_k w_k h_k(s_i) \geq r(s_i, a) + \gamma \sum_{s_j} p(s_j | s_i, a) \sum_k w_k h_k(s_j) \\
 & \forall s_i, s_j \in S, a \in A
 \end{aligned} \tag{9}$$

We can further obtain that:

$$\begin{aligned}
 & \min \sum_{s_i} w_k \sum_k \alpha(s_i) h_k(s_i) \\
 & s.t. \sum_k w_k h_k(s_i) \geq r(s_i, a) + \gamma \sum_k w_k \sum_{s_j} p(s_j | s_i, a) h_k(s_j) \\
 & \forall s_i, s_j \in S, a \in A
 \end{aligned} \tag{10}$$

Let:

$$\begin{cases} \alpha_k = \sum_{s_i} \alpha(s_i) h_k(s_i) \\ g_k(s_i) = \sum_{s_j} p(s_j | s_i, a) h_k(s_j) \end{cases} \tag{11}$$

We have:

$$\begin{aligned}
 & \min \sum_k w_k \alpha_k \\
 & s.t. \sum_k w_k h_k(x) \geq r(s_i, a) + \gamma \sum_i w_i g_i(x) \\
 & \forall s_i \in S, a \in A
 \end{aligned} \tag{12}$$

The problem (12) has only K variables compared to  $|Q| \times |Y| \times |G| \times |E|$  state in problem (7). However, the number of constraints in (12) is still very large. In the next section, we will propose the fast algorithm with much less complexity.

**The fast algorithm via state aggregation:** For the large-scale MDP problem, the system state space is very huge and the transition matrix is sparse in general. Factored MDP uses a compact representation to replace the original MDP problem. The traditional value iteration is then transformed into linear programming. Then we use finite linear basis to reduce the free variables in the linear programming. Next, we present the fast algorithm via system state aggregation to further reduce the size of the problem.

We divide the system states into several groups and aggregate the system states in each group into a macro-state and generate costs among macro states. The policy therefore becomes a hierarchical policy. The hierarchical policy is a mapping from one macro system states to another macro system state. The cost function and the transition matrix are based on the macro states. We then formulate a reduced MDP problem, which is equivalent to the original MDP problem (1).

The original MDP problem (1) is defined by a four-tuple set  $M = (S, A, P, r)$ , where  $S$  is system state space,  $A$  is the action space,  $P$  is the transition matrix,  $r$  is the cost functions. The reduced MDP model is defined by  $M' = (S', A, P', r')$ , which has the same action space with the original MDP model. And  $S' = \{B_1, B_2, \dots, B_L\}$ , where,

$$B_1 \cap B_2 \cap B_3 \cap \dots \cap B_L = S \quad (13)$$

The transition probability is given by:

$$P'(i, j, a_t) = \Pr(B_j | B_i, a_t) = \sum_{s_j \in B_j} \Pr(s_j | s_i, a_t), \quad \forall s_i \in B_i \quad (14)$$

And the cost function is defined by:

$$r'(B_i, a_t) = \sum_{s_i \in B_i} r(s_i, a_t) \quad (15)$$

Therefore, the stage aggregation method is given as follows:

**Input:** Initial partition  $b_1 = \{B_1, B_2, B_3, \dots, B_L\}$ , target partition size

**Output:** Target partition  $\{B_1, B_2, B_3, \dots, B_L\}'$

- **Initialize:**  $b_t = b_1$
- While (current size  $\leq$  target partition size) do
- For every block in partition  $b_t$  do
- Further partition the related block
- End
- End
- **Return:** The target partition  $\{B_1, B_2, B_3, \dots, B_L\}'$

After we divide the system state space, we implement the approximate linear programming to obtain the policy. In the next section, we will show the simulation results of the proposed algorithm.

## SIMULATION RESULTS

In our simulation studies, the length of the time slot is set as 5 ms. We consider the channel experience the Rayleigh fading. The Doppler shift is set as 10 Hz. The channel fading is characterized by 4 states. We assume that the required SNR is fixed. Therefore, there are 4 types of transmission powers. The data queue length is set as 3. The packet arrival process is set as a Poisson process with rate 0.1. We assume that one packet can be sent during one time slot. The energy buffer length is set as 80. The energy harvesting dynamic is characterized by 3 states. The energy harvesting trace is shown in Fig. 3. Therefore, the total number of system states is 2880. We run 1000 realization of the channel with 10 min.

We compare the proposed fast algorithm to the exact value iteration algorithm. The exact value iteration algorithm provides the optimal solution of the

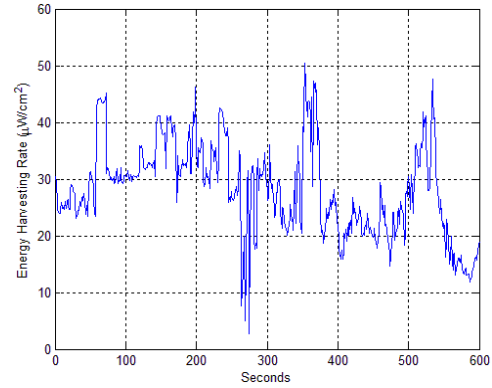


Fig. 3: The energy harvesting trace in 10 min

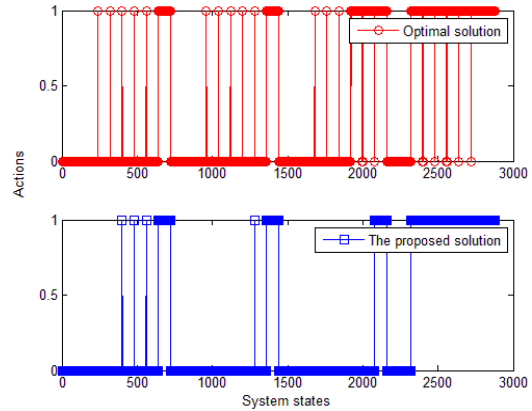


Fig. 4: Action comparison between the optimal solution and the proposed solution

Table 1: Computational performance comparison between the optimal solution and the proposed solution

	The optimal solution	The proposed solution
Number of system states	2880	68
CPU time (sec)	850.3594	0.3750
Accuracy	100%	93.8000%
Energy consumption (J)	0.0138	0.0144

problem (1). Figure 4 shows the actions of the optimal solution and the proposed solution for all system states.

It is shown that the proposed algorithm takes identical action as the optimal solution most of the time. The action accuracy reaches 93.8% compared to the optimal solution. Next, we compare the size of the two methods and the computational time as shown in Table 1. The proposed solution has much smaller number of system states and the computational time period is very short compared to the optimal solution. In addition, the energy consumption and the accuracy remain good performance similar to the optimal solution.

### CONCLUSION

In this study, we have proposed the fast algorithm for large-scale MDP in the wireless smart meter systems. We first addressed the compact representation problem for the large-scale MDP. Based on the factored MDP representation, we then proposed the state aggregation algorithm, which can largely reduce the size of the original problem with very low complexity. Extensive simulation results have verified the good performance of the proposed scheme.

### ACKNOWLEDGMENT

This study is supported by the Fundamental Research Funds for the Central Universities in China under Grant ZYGX2009J006.

### REFERENCES

Fu, F. and M. Van Der Schaar, 2010. A systematic framework for dynamically optimizing multi-user wireless video transmission. *IEEE J. Select. Areas Commun.*, 28(3): 308-320.

- Karmokar, A., D. Djonin and V. Bhargava, 2008. Cross-layer rate and power adaptation strategies for IR-HARQ systems over fading channels with memory: A SMDP-based approach. *IEEE Trans. Comm.*, 56(8): 1352-1365.
- Li, P.F., W. Qiao, H. Sun, H. Wan, J. Wang, *et al.*, 2010. Smart transmission grid: Vision and framework. *IEEE Trans. Smart Grid*, 1(2): 168-177.
- Lin, Y. and V. Wong, 2008. An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks. *IEEE Trans. Vehicular Technol.*, 57(2): 1243-1254.
- Moghe, R., F. Lambert and D. Divan, 2012. Smart "stick-on" sensors for the smart grid. *IEEE Trans. Smart Grid*, 3(1): 241-252.
- Moslehi, K. and R. Kumar, 2010. A reliability perspective of the smart grid. *IEEE Trans. Smart Grid*, 1(1): 57-64.
- Niyato, D., E. Hossain and D. Kim, 2009. Joint admission control and antenna assignment for multiclass QoS in spatial multiplexing MIMO wireless networks. *IEEE Trans. Wireless Comm.*, 8(9): 4855-4865.
- Phan, C., Y. Park, H. Choi, J. Cho and J. Kim, 2010. An energy-efficient transmission strategy for wireless sensor networks. *IEEE Trans. Consumer Electr.*, 56(2): 597-605.
- Puterman, M.L., 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley-Interscience, Hoboken, ISBN: 0471727822, pp: 649.