Optimal Energy Allocation for Linear Control over a Packet-Dropping Link with Energy Harvesting Constraints

Steffi Knorn *, **Subhrakanti Dey** *

* *Uppsala University, Sweden; email: {steffi.knorn; subhra.dey}@signal.uu.se*

Abstract
A sensor computes a state estimate of a closed loop linear control system. The state estimate is packetized and sent to the controller in the receiver block over a randomly time-varying (fading) packet dropping link. The receiver sends an ACK/NACK packet to the transmitter over a perfect feedback channel. The energy used in packet transmission depletes a battery of limited capacity at the sensor. The battery is replenished by an energy harvester, which has access to a source of everlasting but random harvested energy. Further, the energy harvesting and the fading channel gain processes are described as finite-state Markov chain models. The objective is to design an optimal energy allocation policy at the transmitter and an optimal control policy at the receiver so that an average infinite horizon linear quadratic Gaussian (LQG) control cost is minimised. It is shown that a separation principle holds, the optimal controller is linear, the Kalman filter at the sensor is optimal, and the optimal energy allocation policy at the transmitter can be obtained via solving the Bellman dynamic programming equation to a Markov decision process based stochastic control problem. A Q-learning algorithm is used to approximate the optimal energy allocation policy. Numerical simulations illustrate that the dynamic programming based policies outperform the simple heuristic policies.

Keywords: Kalman filtering (KF), energy harvesting, optimal control, energy allocation

1 Introduction

Wireless sensor systems have become much more powerful than their early predecessors, affordable and compact. They are increasingly being used in many areas such as environmental data gathering Akyildiz et al. [2002], industrial process monitoring Gungor and Hancke [2009], mobile robots Chong and Kumar [2003], and for monitoring of smart electricity grids Gungor et al. [2010]. Since sensors are often located in remote places, they cannot be connected to reliable power sources and are instead powered by batteries. In some cases, this independence from the power grid may also be beneficial to simplify the installation process or system changes.

When relying on limited energy sources and hence using only limited energy for wireless communication, transmitted information might be lost randomly due to noise, interference and fading in the wireless communication channel. It is important to study the effects of such unreliable communication channels on filtering and control. An important line of research in this area started with Sinopoli et al. [2004a], where a Kalman filter relying on measurements received via a packet dropping channel, was considered. It was shown that the resulting Kalman filter and its error covariance matrix are time-varying and stochastic. The mean state estimation error covariance is guaranteed to be bounded if the probability of receiving a packet is above a lower bound. These results were later extended to derive conditions on the packet arrival rate to guarantee the stability of the Kalman filter under various generalisations of the underlying model, transmission scheme, multiple sensors, delayed systems, and with more complex transmitters capable of transmission power control etc. in Liu and Goldsmith [2004], Xu and Hespanha [2005], Huang and Dey [2007], Epstein et al. [2008], Schenato [2008], Mo and Sinopoli [2008], Quevedo et al. [2012]. An overview of some of the earlier results can be found in Schenato et al. [2007]. Other researchers have studied how to minimise the expected estimation error, that is, the *performance* of the Kalman filter, see for instance Quevedo et al. [2010], Shi et al. [2011].

The impact of packet dropping links in closed loop control systems was also investigated. For example, Sinopoli et al. [2004b] studied a closed loop control system with a linear Gaussian quadratic optimal controller. If the sensor receives perfect feedback about the packet loss process (TCP-like case), the separation principle holds. Further, there exists a critical arrival probability below which the resulting optimal controller fails to stabilize the system. Gatsis et al. [2014] studied a similar system but assumed that the transmission energy can be chosen in order to influence the packet arrival probability, and derived an optimal transmission policy, which minimises the infinite horizon cost combining transmission power costs and a quadratic control cost. Sinopoli et al. [2005a] assumed that the control signal is also transmitted via an unreliable communication channel. If perfect channel feedback is available at the actuator, the separation principle holds and the optimal LQG control is linear. However, without perfect channel feedback, the separation principle does not hold and the resulting optimal controller is in general nonlinear - see Sinopoli et al. [2005c] (apart from some special cases, Sinopoli et al. [2005b]).

Wireless sensor systems/networks are often placed in an environment where energy can be harvested using solar panels, wind mills or other devices capable of harvesting vibrational/mechanical energy. The harvested energy can then be used to recharge the battery or immediately transmit data. Since most renewable energy sources are unreliable and hard to predict for longer time horizons, finding an optimal energy allocation policy in achieving a long-term performance measure is a challenging task. Recent literature focusing on wireless communications with energy harvesting transmitters have investigated optimal energy allocation policies in order to optimize various metrics related to information transmission such as throughput or delay etc. For example, in Sharma

⋆ Corresponding author: Steffi Knorn. Tel. +46-18-471-7389.

et al. [2010], the authors studied energy allocation policies in a single sensor node for throughput maximization and mean delay minimization. Ho and Zhang [2012] studied the optimal energy allocation policy to maximize the mutual information of a wireless link. The derivation of an optimal packet scheduling problem for a single-user energy harvesting wireless communication system (minimizing the delivery time for all packets) can be found in Yang et al. [2012]. Optimal off-line transmission policies assuming limited battery capacities are investigated in Tutuncuoglu and Yener [2012]. These results are further generalized in Ozel et al. [2011] considering fading channels and optimal online policies.

In the context of estimation and control of dynamical systems, estimation of a dynamical system with a packet dropping link under energy harvesting constraints was first studied in Nourian et al. [2014a], where a sensor equipped with an energy harvester and a rechargeable battery sends its measurements over a packet dropping link to the receiver. Optimal transmission energy allocation policies to minimise the expected error covariance in the presence of perfect or imperfect channel feedback were derived. In Nourian et al. [2014b], the authors assumed a smart sensor, which at every time step can send either a quantized version of its local state estimate or its local innovation via a packet dropping link with a possibly imperfect packet acknowledgement link.

We extend the results of Nourian et al. [2014a] to a closed loop control system with a packet dropping link between the sensor and the controller at the receiver. We study the optimal energy allocation policy at the transmitter and the optimal control design at the receiver such that an infinite-time horizon average LQG control cost is minimised. The "smart" sensor performs state estimation of the observed linear dynamical system and transmits the current state estimate (as opposed to the measurements as in Nourian et al. [2014a]) to the receiver via a packet dropping link. This transmission strategy is chosen based on the results in Gupta et al. [2007], where it was shown that it is optimal to send estimates over packet dropping links. The receiver sends an acknowledgement whether it has received the state estimate to the transmitter. In contrast to Sinopoli et al. [2004b, 2005a,c,b], the transmitter at the sensor is equipped with a rechargeable battery of finite capacity and an energy harvester, and can choose how much energy should be used for transmission. The transmission energy is limited by the available energy at the battery, which fluctuates randomly due to the stochastic nature of harvested energy. It is assumed that the time varying fading channel gain and the harvested energy processes are described by independent finite-state Markov chains. Hence, the probability of dropping the packet depends on the used transmission energy and the current channel gain and is, therefore, time varying (in contrast to the fixed probabilities considered in Sinopoli et al. [2004b, 2005a,c,b]). This dependence of the packet loss probability on the random channel gain forces the transmitter to find a tradeoff amongst spending energy to transmit the current state estimate, keeping energy in reserve for future transmissions in good channel conditions, as well as reducing energy overflow due to a finite battery capacity. It is shown that the separation principle holds and the optimal controller is linear when the receiver acknowledgement is received without error. The optimal energy allocation policy to determine the energy used to transmit the current state estimate to the receiver is obtained by solving an average cost optimal Bellman equation. In the case where environmental parameters such as the transition probabilities of the underlying Markov chains describing the energy harvesting and fading channel gains are unknown, a Q-learning based suboptimal algorithm is also provided. The optimal energy allocation algorithm and the Q-learning based algorithm are compared to various other strategies, such as sending the current measurement instead of the state estimate, and two suboptimal heuristic transmission energy allocation policies.

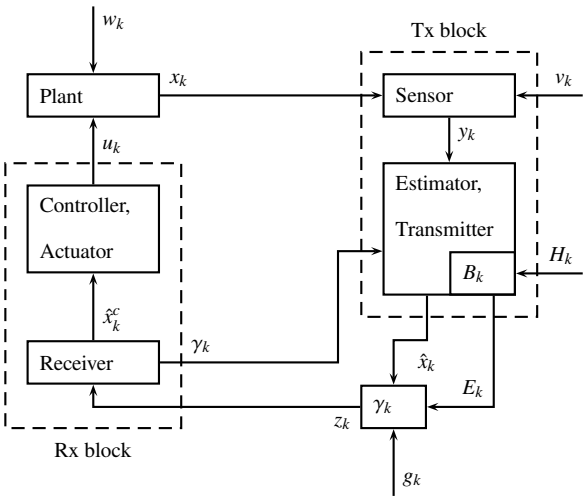

Figure 1. Scheme of system model

Section 2 describes the system model, and the optimal control and energy allocation policies are described in Section 3. Section 4 describes the Q-learning algorithm and Section 5 describes two suboptimal heuristic energy allocation policies and all policies are compared via numerical studies in Section 6, followed by concluding remarks in Section 7.

2 System Model

A scheme of the system model can be found in Figure 1. A detailed description of the components is given below.

2.1 Plant Model

The plant is modeled as a simple linear system with state $x_k \in \mathbb{R}^n$, process noise $w_k \in \mathbb{R}^n$, and a control input $u_k \in \mathbb{R}^p$: $x_{k+1} = Ax_k + Bu_k + w_k$, where it is assumed that (A,B) is stabilizable. The process noise is assumed to be i.i.d. Gaussian noise with zero mean and covariance matrix $M = \mathbb{E}\{w_k w_k^{\mathrm{T}}\} \geq 0$. The initial state x_0 is also Gaussian with mean $\bar{x}_0$, and covariance $\bar{P}_0$, and A, B are matrices of appropriate dimensions. Similar to Schenato et al. [2007], we also assume that (A, B) and $(A, M^{\frac{1}{2}})$ are controllable.

2.2 Sensor

The sensor produces a noisy measurement of the state given by $y_k = Cx_k + v_k$ where (A,C) is assumed to be observable, $y_k \in \mathbb{R}^q$, and $v_k \in \mathbb{R}^q$ is assumed to be i.i.d. Gaussian noise (independent of x_0 and w_k) with zero mean and covariance matrix $N = \mathbb{E}\{v_k v_k^{\mathrm{T}}\} > 0$.

2.3 State Estimator at the Transmitter

We assume a smart sensor with computational capability, and that the sensor transmitter forwards a state estimate to the remote estimator/controller. The sensor measurements are used at the transmitter to estimate the current state x_k based on the information set $\mathcal{I}_k = \{\hat{x}_0, y_l, \gamma_{l-1} : 1 \leq l \leq k\}$, where γ_l denotes the packet loss process in the sensor-receiver communication link as well as the channel feedback acknowledgment (due to perfect channel feedback assumed in this work), which will be discussed in detail in Section 2.5. Since the transmitter knows the exact packet loss sequence, it can reconstruct the time-varying Kalman filter at the receiver, and the exact control input applied to the plant, which is calculated at the receiver based on the receiver state estimate. The estimate at the transmitter is

$$\hat{x}_k := \hat{x}_{k|k} = \mathbb{E}\{x_k|\mathcal{I}_k\} = \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}), \quad (1)$$

$$\hat{x}_{k+1|k} = \mathbb{E}\{x_{k+1}|\mathcal{I}_k\} = A\hat{x}_{k|k} + Bu_k. \quad (2)$$

The matrix K_k should be chosen such that it minimises the error covariance matrix of the state estimation error. The error covariance matrices at the transmitter are

$$P_{k|k} = \mathbb{E}\left\{(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^{\mathrm{T}}|\mathcal{I}_k\right\}, \quad (3)$$

$$P_{k+1} := P_{k+1|k} = \mathbb{E}\left\{(x_{k+1} - \hat{x}_{k+1|k})(x_{k+1} - \hat{x}_{k+1|k})^{\mathrm{T}}|\mathcal{I}_k\right\}. \quad (4)$$

With $e_{k|k} = x_k - \hat{x}_{k|k}$ and $e_{k+1|k} = x_{k+1} - \hat{x}_{k+1|k} = Ae_{k|k} + w_k$, this yields

$$P_{k+1} = \mathbb{E}\left\{(Ae_{k|k} + w_k)(Ae_{k|k} + w_k)^{\mathrm{T}}\right\} = AP_{k|k}A^{\mathrm{T}} + M. \quad (5)$$

Further, choosing $K_k = P_{k|k-1}C^{\mathrm{T}}\left(CP_{k|k-1}C^{\mathrm{T}} + N\right)^{-1}$ such that $\hat{x}_k = \hat{x}_{k|k-1} + P_{k|k-1}C^{\mathrm{T}}\left(CP_{k|k-1}C^{\mathrm{T}} + N\right)^{-1}(y_k - C\hat{x}_{k|k-1})$ leads to the minimal error covariance matrix after updating the estimate $P_{k|k}$ in the standard form

$$P_{k|k} = P_{k-1} - P_{k-1}C^{\mathrm{T}}\left(CP_{k-1}C^{\mathrm{T}} + N\right)^{-1}CP_{k-1}. \quad (6)$$

The initial covariance matrix is given by $P_0 = \bar{P}_0$. Due to the standard controllability and observability conditions, we assume that the Kalman filter at the transmitter has been running long enough to reach a steady state such that the error covariance matrix at the transmitter is given by P_∞, which is the steady state estimation error covariance $P_{k|k}$ as $k \to \infty$.

2.4 Energy Harvester and Battery Dynamics

The transmitter has a rechargeable battery equipped with an energy harvester, that can gather energy from the environment. The amount of energy available to be harvested at time slot k, denoted by H_k, is unpredictable and is described as a stationary first-order homogeneous finite-state Markov process, Ho et al. [2010]. The energy harvested at time slot k is stored in the battery and can be used for data transmission. In this paper, we assume that the energy used for sensing and computational purposes at the transmitter are negligible compared to the amount of energy required for transmission. This is particularly true if data is transmitted over a wireless channel to a receiver that is a long distance away. The amount of stored energy in the battery at time k, B_k, evolves according to

$$B_{k+1} = \min\{B_k - E_k + H_{k+1}; \quad \bar{B}\} \quad (7)$$

with $0 \le B_0 \le \bar{B}$ and where $\bar{B}$ is the battery capacity, and E_k is the energy used for packet transmission during the k-th slot.

2.5 Communication Channel

A wireless communication channel is used to transmit the state estimate $\hat{x}_k$ to the controller/actuator unit, referred to as Rx block. The channel is a packet dropping link such that the estimate is either exactly received (that is for $\gamma_k = 1$) or completely lost due to corrupted data or substantial delay (that is for $\gamma_k = 0$), where γ_k is the Bernoulli random variable modelling the packet loss process. The received signal is $z_k = \gamma_k\hat{x}_k$. The probability of successfully transmitting the packet is

$$\mathbb{P}(\gamma_k = 1|g_k, E_k) := h(g_kE_k) \quad (8)$$

where g_k is the time-varying wireless fading channel gain and E_k is the transmission energy for transmitting the packet at k over the channel. The function $h : [0,\infty] \to [0,1]$ is monotonically increasing and continuous.

We assume that the fading channel gain $\{g_k\}$ is a first-order stationary homogeneous finite-state Markov fading process where the channel gains remain constant over each fading block and are independent of the energy harvesting process H_k, and known to the transmitter. This can be achieved by the receiver sending a pilot signal at the beginning of each slot for the transmitter to estimate the channel from the receiver to the transmitter. Under a time-division-duplex transmission scheme (essentially where the transmitter uses the remainder of the same time slot in which the channel estimation is performed), the channel from the sensor transmitter to the receiver is the same due to channel reciprocity.

Based on the channel gain g_k, and the current battery level B_k, the transmitter finds an optimal energy allocation policy $\{E_k\}$ in order to minimise a suitable finite horizon control cost. The details of this optimal energy allocation scheme will be provided in the next section.

After receiving z_k over the lossy communication channel, the receiver sends an ACK/NACK packet to the transmitter, over a perfect feedback channel, and is therefore equivalent to γ_k.

2.6 Estimator/Controller and Actuator in the Receiver block

The controller in the receiver block has access to the information set $\mathcal{I}_k^{\mathrm{c}} := \{\hat{x}_0^{\mathrm{c}}, z_l, \gamma_l : 1 \le l \le k\}$. As the estimates from the transmitter Kalman filter are dropped with probability $1 - h(g_kE_k)$, the current state estimate is not always available at the Rx block. Hence, the state estimate at the Rx block, $\hat{x}_{k|k}^{\mathrm{c}} = E[x_k|\mathcal{I}_k^{\mathrm{c}}]$, is given by

$$\hat{x}_k^{\mathrm{c}} := \hat{x}_{k|k}^{\mathrm{c}} = \gamma_k\hat{x}_k + (1 - \gamma_k)\left(A\hat{x}_{k-1}^{\mathrm{c}} + Bu_{k-1}\right). \quad (9)$$

Thus, the estimation error covariance matrix at the Rx block is

$$\begin{aligned} P_{k|k}^{\mathrm{c}} :=& \mathbb{E}\left\{(x_k - \hat{x}_k^{\mathrm{c}})(x_k - \hat{x}_k^{\mathrm{c}})^{\mathrm{T}}|\mathcal{I}_k^{\mathrm{c}}\right\} \\ =& \gamma_kP_\infty + (1 - \gamma_k)\left(AP_{k-1|k-1}^{\mathrm{c}}A^{\mathrm{T}} + M\right). \end{aligned} \quad (10)$$

For simplicity, we assume that the Rx block uses the same initial state distribution, such that $P_{0|0}^{\mathrm{c}} := \bar{P}_0$. Since it is assumed that the Tx block receives the ACK/NACK packet without fault, a copy of $P_{k|k}^{\mathrm{c}}$ can be kept at the Tx block.

The task of the controller is to design an optimal control sequence $\{u_k\}$ based on the information pattern $\mathcal{I}_k^{\mathrm{c}}$ such that a suitable average control cost is minimised. It is assumed that the link between the Rx block and the plant is lossless, such that the correct control signal u_k is applied to the plant. This is a reasonable assumption in case the actuator is directly connected or located very close to the plant. The optimization problem for finding the optimal transmission energy allocation and optimal control policy is described below.

3 Optimal Energy Allocation and Control Policy Design

In this section, our aim is to find the stationary optimal transmission energy allocation policy $\{E\}^*$ (if it exists) and the optimal control policy $\{u\}^*$, that jointly minimise the following infinite-horizon average LQG control cost (for a given mean and covariance of the initial state)

$$\begin{aligned} J(\{u\},\{E\},\bar{x}_0,\bar{P}_0) &= \lim_{T\to\infty}\frac{1}{T}\sum_{k=1}^{T}\mathbb{E}\left\{x_k^{\mathrm{T}}Wx_k + u_k^{\mathrm{T}}Uu_k\right\} \\ &= \lim_{T\to\infty}\frac{1}{T}\sum_{k=1}^{T}\mathbb{E}\left\{\mathbb{E}\left\{x_k^{\mathrm{T}}Wx_k + u_k^{\mathrm{T}}Uu_k\middle|\mathcal{I}_k^{\mathrm{c}}\right\}\right\}. \end{aligned} \quad (11)$$

We also assume that in addition to (A,C) being observable, $(A, W^{\frac{1}{2}})$ is also observable. The results of this section are largely based on the proof of a separation principle as shown in Schenato et al. [2007], which leads to a linear optimal control law, and the optimal energy allocation policy that minimises an infinite-horizon average receiver estimation error covariance cost as considered in Nourian et al. [2014a]. Due to space constraints, we only provide a sketch of the arguments necessary for the proof. The detailed proofs will be provided in a longer version currently under preparation.

Since we consider the perfect channel feedback case, this is similar to the TCP-like protocol considered in Schenato et al. [2007]. In order to show that a separation principle holds, one can consider a finite horizon version of the above problem as considered in Schenato et al. [2007]. Following an almost identical analysis as in Sec. V of this paper, but with the receiver information set $\mathcal{I}_k^{\mathrm{c}}$, one can use a value function approach and a corresponding induction based proof to show that the value function is a quadratic function of the control input u_k. This implies, that the control cost (11) can be minimised by solely optimizing over u_k while keeping E_k fixed, the optimal controller is linear and is of the form

$$u_k^* = L\hat{x}_k^{\mathrm{c}}. \tag{12}$$

It also follows that the tasks of obtaining the optimal Kalman filtered state estimate $\hat{x}_k, \hat{x}_k^{\mathrm{c}}$, calculating the optimal control input u_k^* at the controller, and computing the optimal energy allocation E_k^* at the transmitter can be carried out separately. Further, since the link between the controller and the actuator is perfect, the optimal controller gain L has the form

$$L = L_\infty = -\left(B^{\mathrm{T}} S_\infty B + U\right)^{-1} B^{\mathrm{T}} S_\infty A \tag{13}$$

where S_∞ is the solution of the standard ARE

$$S_\infty = A^{\mathrm{T}} S_\infty A + W - A^{\mathrm{T}} S_\infty B \left(B^{\mathrm{T}} S_\infty B + U\right)^{-1} B^{\mathrm{T}} S_\infty A. \tag{14}$$

Details can be found for instance in Sinopoli et al. [2004b].

It can be further shown that with this optimal controller, minimising the cost given by (11) with respect to the energy allocation policy is equivalent to solving the following stochastic control problem

$$\min_{E_k : k\geq 1} \limsup_{T\to\infty} \frac{1}{T} \sum_{k=1}^{T} \mathbb{E}\left\{tr\left(P_{k|k}^{\mathrm{c}}\right)\right\}. \tag{15}$$

This can be regarded as a Markov Decision Process (MDP) formulation with state space $\mathcal{S} = \{P_{k|k}^{\mathrm{c}}, g_k, H_k, B_k\}$ and action space $\mathcal{A} = \{E_k\}$. More details on MDPs can be found in-Bertsekas [1995], Altman [1999]. Clearly, just like the optimal controller, the optimal energy allocation policy can be obtained at the receiver and fed back to the transmitter provided the receiver knows the transmitter's current battery state and harvested energy patterns. This is obviously impractical. However, due to perfect channel feedback, the receiver error covariance is known at the Tx block, along with its battery state and harvested energy patterns. Hence, the optimization problem is solved at the Tx block.

Despite using the optimal time-varying Kalman filter at the receiver and the optimal LQG controller, boundedness of the cost function (15) with an unstable open loop system cannot be guaranteed if the packet dropping probability of the forward channel is too high. The following theorem shows that the average error covariance matrix at the receiver is bounded under suitable conditions.

Theorem 1. Assume the error covariance matrix at the controller $P_{k|k}^{\mathrm{c}}$ in (10). If there exists a $\xi \in [0,1)$ such that

$$\sup_{g,H} \int_{g_k} \int_{H_k} \left(1 - h\left(g_k \min\left\{H_k, \bar{B}\right\}\right)\right) \times \mathbb{P}\left(g_k | g_{k-1} = g\right) \times \mathbb{P}\left(H_k | H_{k-1} = H\right) dg_k dH_k \leq \frac{\xi}{\|A\|^2} \tag{16}$$

for all $k \geq 0$, then there exists an energy allocation policy $\{E_k\}$ such that the norm of P_k^{c} in (10) is exponentially bounded by

$$\mathbb{E}\left\{\left\|P_{k|k}^{\mathrm{c}}\right\|\right\} \leq \alpha \xi^k + \beta \tag{17}$$

for $k \geq 0$ and some non-negative scalars α and β.

Assume the error covariance matrix, channel gain, harvested energy, battery level and energy consumption at time k are denoted $P^{\mathrm{c}} = P_{k|k}^{\mathrm{c}}$, $g = g_k$, $H = H_k$, $B = B_k$, and $E = E_k$, respectively, and the corresponding error covariance matrix, channel gain, harvested energy and battery level at time $k+1$ are $\tilde{P}^{\mathrm{c}} = P_{k+1|k+1}^{\mathrm{c}}$, $\tilde{g} = g_{k+1}$, $\tilde{H} = H_{k+1}$ and $\tilde{B} = B_{k+1}$, respectively. Then the following theorem illustrates that an average cost Bellman optimality equation can be proved to hold.

Theorem 2. If conditions (W) and (B) of Schäl [1993] (requiring for example, the state space to be a compact set, the state to action mapping to be upper semicontinuous, the transition probabilities being weakly continuous, and the cost function being upper semicontinuous), then the infinite-time horizon stochastic control problem (15) is given by ρ, which is the unique solution of the average-cost optimality Bellman equation

$$\rho + V(P^{\mathrm{c}}, g, H, B) = \min_{E\in[0,B]} \mathbb{E}\left\{tr\left(P^{\mathrm{c}}\right) + V\left(\tilde{P}^{\mathrm{c}}, \tilde{g}, \tilde{H}, \tilde{B} \middle| P^{\mathrm{c}}, g, H, B, E\right)\right\} \tag{18}$$

where V is the relative value function. The optimal average cost ρ is independent of the initial conditions P_0, g_0, H_0 and B_0.

The stationary optimal solution to the stochastic control problem (15) is then given by

$$E^o(P^{\mathrm{c}}, g, H, B) = \underset{0\leq E\leq B}{\operatorname{argmin}} \mathbb{E}\left\{tr\left(P^{\mathrm{c}}\right) + V\left(\tilde{P}^{\mathrm{c}}, \tilde{g}, \tilde{H}, \tilde{B} \middle| P^{\mathrm{c}}, g, H, E\right)\right\}. \tag{19}$$

4 Q-Learning

Note that the Bellman equation to determine the optimal energy allocation policy cannot be used if some of the underlying system parameters are not completely known such as, for instance, the transition probabilities of the Markov process generating the channel gains or the energy harvesting process. Hence, finding algorithms, which do not rely on the complete knowledge of the underlying system, is an important task. Assume that the state space is discrete or can be approximately discretised. In the current work, this corresponds to a discretised approximation to the space of error covariance matrices, whereas the channel gain, the harvested energy, the battery level and the allocated energy usage are assumed to take values in a finite-discrete state space. Since the fading channel gains and harvested energy are independent Markov processes, the average-cost optimality Bellman equation (18) can be simplified to the Q-Bellman equation

$$Q^*(P^{\mathrm{c}}, g, H, B, E) = \mathbb{E}\left\{tr\left(P^{\mathrm{c}}\right)\right\} + \sum_{\tilde{g},\tilde{H},\tilde{B}} \mathbb{P}(\tilde{g}|g)\mathbb{P}(\tilde{H}|H)\mathbb{P}(\tilde{B}|B,H,E) \min_{\tilde{E}\in\mathcal{A}(\tilde{B})} Q^*(\tilde{P}^{\mathrm{c}}, \tilde{g}, \tilde{H}, \tilde{B}, \tilde{E}) \tag{20}$$

where $\mathcal{A}(\tilde{B})$ is the set of all feasible choices of $\tilde{E}$ given $\tilde{B}$. Note that $\mathbb{P}(x|y)$ is the probability of x given y. As discussed in Sutton and Barto [1998] and Prabuchandran et al. [2013], equation (20) can be solved using the stochastic approximation based Q-learning algorithm. Assuming that the probabilities $\mathbb{P}(\tilde{g}|g)$, $\mathbb{P}(\tilde{H}|H)$ and $\mathbb{P}(\tilde{B}|B,H,E)$ are unknown, but that the states can be observed it is given by:

$$Q_0(P^{\mathrm{c}}, g, H, B, E) = 0 \quad \forall P^{\mathrm{c}}, g, H.B \text{ and } E \in \mathcal{A}(B) \tag{21}$$

and for all $k \geq 0$

$$Q_{k+1}(P^{\mathrm{c}}, g, H, B, E) = Q_k(P^{\mathrm{c}}, g, H, B, E) + \gamma(k) \cdot \left(\mathbb{E}\left\{tr\left(P^{\mathrm{c}}\right)\right\} + \min_{\tilde{E}\in\mathcal{A}(\tilde{B})} Q_k(\tilde{P}^{\mathrm{c}}, \tilde{g}, \tilde{H}, \tilde{B}, \tilde{E}) - Q_k(P^{\mathrm{c}}, g, H, B, E)\right) \tag{22}$$

where now $\{\tilde{P}^{\mathrm{c}}, \tilde{g}, \tilde{H}, \tilde{B}, \tilde{E}\}$ is the next state, whereas $\{P^{\mathrm{c}}, g, H, B, E\}$ is the previous state at which $E \in \mathcal{A}(B)$ is selected according to the ϵ-greedy method:

$$E = \begin{cases} \operatorname{argmin}_{E\in A(B)} Q_k(P^{\mathrm{c}}, g, H, B, E) & \text{w/ prob. } 1-\epsilon \\ \text{chosen randomly } \in \mathcal{A}(B) & \text{w/ prob. } \epsilon \end{cases} \tag{23}$$

The algorithm in (22) converges to the optimal Q values if the step sizes $\gamma(k)$ for all $k \geq 0$ satisfies $\gamma(k) > 0$, $\sum_k \gamma(k) = \infty$ and $\sum_k \gamma^2(k) < \infty$, and the convergence is guaranteed for all $\epsilon > 0$, Sutton and Barto [1998], Prabuchandran et al. [2013]. The algorithm spends more computational effort in exploring the effect of possible choices of E if ϵ is large. However, a small value of ϵ is usually preferred as it allows to exploit the knowledge of which choice of E leads to the minimal expected cost based on Q_k.

5 Heuristic Policies for Energy Allocation

It is well known that solving the backward dynamic programming equation or the Q-learning algorithm to determine the optimal energy allocation policy requires a large computational overhead. Hence, it is often desirable to find suboptimal policies, that require much less computational effort.

We consider two suboptimal policies in this work for comparison purposes. The first simple suboptimal policy is a "greedy policy" which sets $E_k = B_k, \forall k$. Hence, at every time step all available energy is used to transmit data regardless of the channel gain.

A second simple heuristic policy is the "inverted channel policy". Assume the required transmission energy such that the expected drop-out probability of the communication channel with channel gain g_k is equal to a desired probability $\bar{\gamma}$, is denoted by $E_{\bar{\gamma}}(\bar{\gamma}, g_k)$. Then, the inverted channel energy allocation policy follows the simple rule $E_k = \min\{B_k, E_{\bar{\gamma}}(\bar{\gamma}, g_k)\}$.

6 Numerical Examples

In this section, we evaluate the performance of various optimal and suboptimal energy allocation policies. A scalar system with parameters $A = 1.1$, $B = 1$, $C = 1$, $M = 1$, $N = 1$ and $P_{x_0} = 1$ is considered. It is assumed that the sensor uses a binary phase shift keying (BPSK) transmission scheme, Proakis [2001], with $b = 4$ bits per packet. Therefore, (8) has the form

$$\mathbb{P}(\gamma_k = 1 | g_k, E_k) = h(g_k E_k) = \left(\int_{-\infty}^{\sqrt{g_k E_k}} \frac{1}{\sqrt{2\pi}} e^{-t^2/d} dt \right)^b . \quad (24)$$

The battery capacity is varied between 1mW and 5mW. The fading channel gain and harvested energy are given by independent 3-level discrete Markov chains with values {0,0.5,1} and {0,1,2}, respectively, and the transition probability matrix for both processes is taken to be the same (for simplicity)

$$\mathbb{T} = \begin{bmatrix} 0.2 & 0.3 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.7 \end{bmatrix}. \quad (25)$$

Seven different scenarios have been simulated. In the first two scenarios, the optimal solution is obtained using dynamic programming to solve the average-cost Bellman optimality equation approximately using suitable discretisation of the relevant state spaces. While in the first scenario (red dashed line) the measurement is sent via the noisy communication channel, the state estimate is sent in the second scenario (red solid line).

In the third, fourth and fifth scenarios the Q-learning algorithm is used and the current state estimate is send via the communication channel. The learning horizon is increased from 10^4 (third scenario, QL1, green dashed dotted line) to 10^6 (fourth scenario, QL2, green dashed line) and to 10^8 (fifth scenario, QL3, green solid line).

The sixth and seventh scenarios consider the two heuristic policies described in Section 5. In scenario 6 (blue) the greedy policy is used whereas in the seventh scenario (black) the inverted channel policy with $\bar{\gamma} = 0.7$ is used. This average packet loss probability is calculated based on the previous simulations by averaging over the time-varying packet loss probabilities

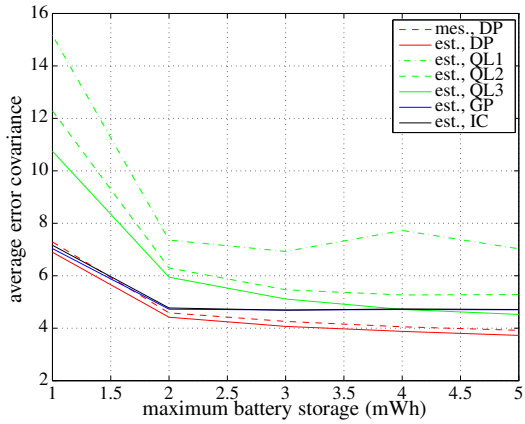

Figure 2. Average error covariance P^c vs. battery capacity, 'mes' = send measurements, 'est' = send estimates, 'DP' = dynamic programming, 'QL' = Q-learning horizon, 'GP' = greedy policy, 'IC' = inverted channel policy

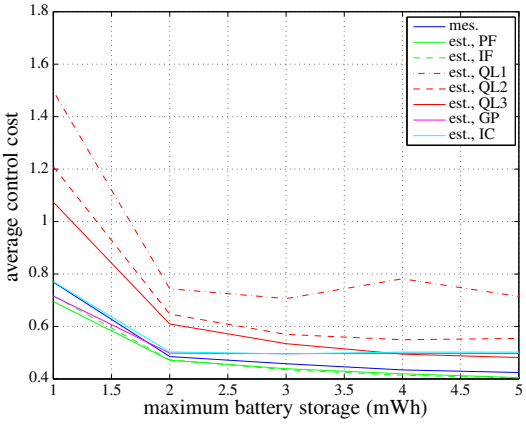

Figure 3. Average control cost J vs. battery capacity (for a detailed legend refer to the caption of Fig. 2)

$h(g_k E_k)$ for a given time-slot k, where E_k corresponds to the optimal energy allocation policy obtained by dynamic programming.

The average error covariance and average control cost for all scenarios (averaging over 10^6 time steps) are shown in Fig. 2 and Fig. 3, respectively. It can be observed that in the first five scenarios (dynamic programming and Q-learning) the average control cost and error covariance matrix decrease if the battery level increases. It can also be seen that sending measurements leads to a slightly worse performance compared to sending the state estimate, as expected. When the learning horizon of the Q-learning algorithm increases, the performance also improves but is worse than those obtained by using dynamic programming and sending measurements or state estimates. The performance of the two heuristic policies are much better than the Q-learning algorithm for low battery capacities but do not improve when increasing the battery capacity. Hence, it is only beneficial to invest in computation time to run the dynamic programming or the Q-learning algorithms for larger battery capacities.

7 Conclusions

This paper studied a linear control system with a packet dropping link between the smart sensor (calculating the current state estimate) and the controller. Since the link is a time-varying fading channel, the probability of receiving the current state

estimate is time-varying. The receiver at the controller sends an ACK/NACK packet to the transmitter to acknowledge the arrival of the packet. It is assumed that this feedback is received without faults. The transmitter at the sensor is equipped with a finite battery and an energy harvester to gather energy from its environment. The objective is to design a jointly optimal sensor transmission energy allocation and optimal control policy to minimise an infinite-horizon average LQG control cost.

Because of perfect channel feedback, the separation principle holds. Hence, the Kalman filters at the sensor and a linear controller are optimal, and the transmission energy allocation policy minimizing the expected average estimation error covariance can be obtained by standard dynamic programming techniques for solving the average-cost optimality equation. In case certain underlying system parameters such as the transition probabilities of the associated Markov processes are unknown, we employ Q-learning based suboptimal energy allocation algorithms. We also proposed two simple heuristic energy allocation algorithms. All these algorithms are compared by numerical studies illustrating that the optimal energy allocation policy obtained by dynamic programming outperforms several suboptimal policies especially for higher battery capacities.

References

I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *IEEE Commun. Mag.*, 40(8): 102–114, 2002.

E. Altman. *Constrained Markov Decision Processes*, volume 7. CRC Press, 1999.

D.P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 1. Athena Scientific, 1995.

C.-Y. Chong and S.P. Kumar. Sensor networks: Evolution, opportunities and challenges. *Proc. IEEE*, 91(8):1247–1256, 2003.

M. Epstein, L. Shi, A. Tiwari, and R.M. Murray. Probabilistic performance of state estimation across a lossy network. *Automatica*, 44(12):3046–3053, 2008.

K. Gatsis, A. Ribeiro, and G.J. Pappas. Optimal power management in wireless control systems. *IEEE Transactions on Automatic Control*, 59(6):1495–1510, June 2014.

V.C. Gungor and G.P. Hancke. Industrial wireless sensor networks: Challengesm design, principles and technical approaches. *IEEE Trans. Ind. Electron.*, 56(10):4258–4265, 2009.

V.C. Gungor, B. Lu, and G.P. Hancke. Opportunities and challenges of wireless sensor networks in smart grid. *IEEE Trans. Ind. Electron.*, 57(10):3557–3564, 2010.

V. Gupta, B. Hassibi, and R.M. Murray. Optimal LQG control across packet-dropping links. *System and Control Letters*, 56 (6):439–446, 2007.

C.K. Ho and R. Zhang. Optimal energy allocation for wireless communications with energy harvesting constraints. *IEEE Trans. Signal Process.*, 60(9):4808–4818, 2012.

C.K. Ho, P.D. Khoa, and P.C. Ming. Markovian models for harvested energy in wireless communications. In *IEEE Intern. Conf. Comm. Systems*, pages 311–315, 2010.

M. Huang and S. Dey. Stability of Kalman filtering with Markovian packet losses. *Automatica*, 43(4):698–607, 2007.

X. Liu and A.J. Goldsmith. Kalman filtering with partial observation losses. In *43rd IEEE Conference on Decision and Control*, pages 4180–4186, 2004.

Y. Mo and B. Sinopoli. A characterization of the critical value for Kalman filtering with intermittent observations. In *47th IEEE Conf. Decision and Control*, pages 2692–2216, 2008.

M. Nourian, A.S. Leong, and S. Dey. Optimal energy allocation for Kalman filtering over packet dropping links with imperfect acknowledgments and energy harvesting constraints. *IEEE Trans. Autom. Control*, 59(8):2128–2143, 2014a.

M. Nourian, A.S. Leong, S. Dey, and D.E. Quevedo. An optimal transmission strategy for kalman filtering over packet dropping links with imperfect acknowledgements. *IEEE Trans. Control Netw. Syst.*, 1(3):259–271, 2014b.

O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener. Transmission with energy harvesting nodes in fading wireless channels: Optimal policies. *IEEE J IEEE J. Sel. Areas Commun.*, 29(8):1732–1743, 2011.

K. J. Prabuchandran, S.K. Meena, and S. Bhatnagar. Q-learning based energy management policies for a single sensor node with finite buffer. *IEEE Wireless Commun. Let.*, 2(1):82–85, February 2013.

J.G. Proakis. *Digital Communications*. New York: McGraw-Hill, 4th edition, 2001.

D.E. Quevedo, A. Ahlén, and J. Østergaard. Energy efficient state estimation with wireless sensors through the use of predictive power control and coding. *IEEE Transactions on Signal Processing*, 58(9):4811–4823, 2010.

D.E. Quevedo, A. Ahlén, A.S. Leong, and S. Dey. On Kalman filtering over fading wireless channels with controlled transmission powers. *Automatica*, 48(7):1306–1316, 2012.

D.E. Quevedo, A. Ahlén, and K.H. Johansson. State estimation over sensor networks with correlated wireless fading channels. *IEEE Trans. Autom. Control*, 58(3):581–593, 2013.

M. Schäl. Average optimality in dynamic programming with general state space. *Mathematics of Operational Research*, 18(1):163–172, 1993.

L. Schenato. Optimal estimation in networked control systems subject to random delay and packet drop. *IEEE Trans. Autom. Control*, 53(5):1311–1317, 2008.

L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S.S. Sastry. Foundations of control and estimation over lossy networks. *Proc. IEEE*, 95(1):163–187, 2007.

V. Sharma, U. Mukherji, V. Joseph, and S. Gupta. Optimal energy managment policies for energy harvesting sensor nodes. *IEEE Trans. Wireless Commun.*, 9(4), 2010.

L. Shi, P. Cheng, and J. Chen. Sensor data scheduling for optimal state estimation with communication energy constraints. *Automatica*, 47(8):1693–1698, 2011.

B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M.I. Jordan, and S.S. Sastry. Kalman filtering with intermittent observations. *IEEE Trans. Autom. Control*, 49(9):1453–1464, 2004a.

B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, and S.S. Sastry. Time varying optimal control with packet losses. In *43rd IEEE Conference on Decision and Control*, pages 1938–1943, 2004b.

B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, and S.S. Sastry. Optimal control with unreliable communication: the TCP case. In *American Control Conference*, pages 3354–3359, 2005a.

B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, and S.S. Sastry. An LQG optimal linear controller for control systems with packet losses. In *44th IEEE Conf. on Decision and Control*, pages 458–463, 2005b.

B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, and S.S. Sastry. LQG control with missing observation and control packets. In *16th IFAC World Congress*, 2005c.

R.S. Sutton and A.G. Barto. *Reinforcement learning: An introduction*. Cambridge Univ Press, 1998.

K. Tutuncuoglu and A. Yener. Optimum transmission policies for battery limited energy harvesting nodes. *IEEE Trans. Wireless Commun.*, 11(3):1180–1189, 2012.

Y. Xu and J.P. Hespanha. Estimation under uncontrolled and controlled communications in networked control systems. In *44th IEEE Conference on Decision and Control*, pages 842–847, 2005.

J. Yang, O. Ozel, and S. Ulukus. Broadcasting with an energy harvesting rechargeable transmitter. *IEEE Trans. Wireless Commun.*, 11(2):571–583, 2012.