

# Maximizing Quality of Information From Multiple Sensor Devices: The Exploration vs Exploitation Tradeoff

Ertugrul Necdet Ciftcioglu, *Member, IEEE*, Aylin Yener, *Member, IEEE*, and Michael J. Neely, *Senior Member, IEEE*

**Abstract**—This paper investigates Quality of Information (QoI) aware adaptive sampling in a system where two sensor devices report information to an end user. The system carries out a sequence of tasks, where each task relates to a random event that must be observed. The accumulated information obtained from the sensor devices is reported once per task to a higher layer application at the end user. The utility of each report depends on the timeliness of the report and also on the quality of the observations. Quality can be improved by accumulating more observations for the same task, at the expense of delay. We assume new tasks arrive randomly, and the qualities of each new observation are also random. The goal is to maximize time average quality of information subject to cost constraints. We solve the problem by leveraging dynamic programming and Lyapunov optimization. Our algorithms involve solving a 2-dimensional optimal stopping problem, and result in a 2-dimensional threshold rule. When task arrivals are i.i.d., the optimal solution to the stopping problem can be closely approximated with a small number of simplified value iterations. When task arrivals are periodic, we derive a structured form approximately optimal stopping policy. We also introduce hybrid policies applied over the proposed adaptive sampling algorithms to further improve the performance. Numerical results demonstrate that our policies perform near optimal. Overall, this work provides new insights into network operation based on QoI attributes.

**Index Terms**—Approximate dynamic programming, network utility maximization, quality of information.

## I. INTRODUCTION

THE increasing diversity of wireless applications has led to the introduction of new performance attributes. Traditional performance metrics associated with Quality of Service

Manuscript received October 31, 2012; revised February 20, 2013; accepted April 16, 2013. Date of publication April 24, 2013; date of current version September 11, 2013. This work was supported by the U.S. Army Research Laboratory under the Network Science Collaborative Technology Alliance, Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Edwin Chong.

E. N. Ciftcioglu is with the Department of Computer Science and Engineering, The Pennsylvania State University, University Park, PA 16802 USA. (e-mail: enc118@psu.edu).

A. Yener is with the Department of Electrical Engineering, The Pennsylvania State University, University Park, PA 16802 USA (e-mail: yener@ee.psu.edu; yener@enr.psu.edu).

M. J. Neely is with the Department of Electrical Engineering—Systems Division, University of Southern California, Los Angeles, CA, 90049 USA (e-mail: mjneely@usc.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTSP.2013.2259798

(QoS) such as throughput, delay, and fairness, may not be accurate indicators of the suitability of information for a specific application. In order to address this issue, there has been interest in introducing new attributes which characterize the value of information relevant to a specific application [1], [2]. Attributes such as provenance, accuracy and precision, reliability, corroboration and credibility, age/freshness, and timeliness have been used to define the quality of information [1]–[3]. Event detection applications for QoI are studied in [1], [4]. Recently, there have also been studies which focus on transmission scheduling aspects in QoI-based networks [5], [6]. In [7], we have considered the trade-off between the attributes of accuracy and freshness for a single link.

In this work, we consider a network of multiple sensor devices that repeatedly process new tasks. The goal is to maximize a time average QoI-based utility minus (sensing and processing) costs. The QoI utility is general and can depend on a variety of QoI metrics such as credibility, accuracy, precision, and age. The main property of QoI is that it is a composite metric which deteriorates with age and increases with time due to additional information gathered. The amount of information that sensors collect varies randomly throughout time, which leads to uncertainty in the QoI utility evolution. Consequently, the essential decision to be made by the network via a centralized controller is whether to receive extra information, i.e., *exploration*, or report the information present in order to prevent the utility from degrading due to increased age, i.e., *exploitation*. This decision depends on the amount and quality of the information the sensors have already gathered at the decision instant. While the exploration-exploitation tradeoff has previously been studied for various wireless network scenarios, e.g., cognitive spectrum access [8], [9], autonomous resource management [10], and QoS routing [11], in this paper we specifically address the problem of adaptive sampling [12], [13] which focuses on the determination of the time to report sampled data. The proposed adaptive sampling algorithms attempt to maximize time average QoI-minus-cost. The optimal behavior for sum utilities is characterized by value iteration, leading to a multidimensional *continuing region* for the sensor utilities. We then consider two specific task arrival patterns: i.i.d. arrivals and periodic arrivals. For i.i.d. arrivals, we show that approximate adaptive sampling algorithms based on the value iteration algorithm [14] is sufficient to approach optimality. For periodic arrivals, we also utilize approximations on value functions [15]–[17], and provide algorithms which vary depending on both joint utility states and the amount of time left for the arrival of the next task. The algorithms provided are shown to avoid high computational complexity. For these two task arrival patterns, we then demonstrate that it is optimal to stop if the sum utility exceeds an upper threshold and continue if it is under a

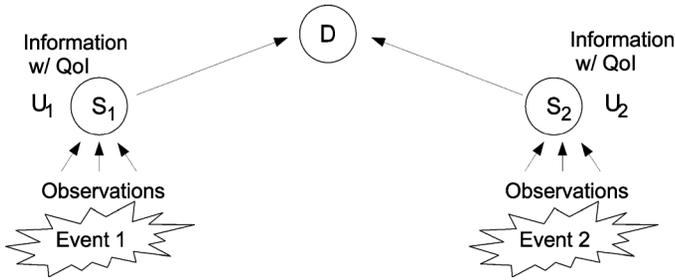


Fig. 1. QoI-based network with two sensor devices.

lower threshold. Finally, we address the alternative problem of maximizing total QoI utility subject to constraints on average cost expenditure. We apply methods from Lyapunov optimization [18], and present a dynamic adaptive sampling algorithm which satisfies the constraints. This work provides new insights on QoI-based networking, which can accommodate for many different QoI attributes and applications. While we have focused on the model of the two-sensor network, the results and methodology provided in this work can be readily generalized to more than two sensors.

## II. SYSTEM MODEL

We consider a network where two sensor devices ( $S_1$  and  $S_2$ ) are issued tasks from an end user. Sensors observe separate phenomena, and gather information related with the task, which are to be reported to the end user  $D$  (Fig. 1).

We consider a slotted system where a random amount of information is gathered by the sensors at each slot. Information gathered by the sensors are associated with a QoI-based utility value for the end user. In particular, the QoI utility is a composite attribute which deteriorates with age and increases with additional information gathered in time.

Tasks arrive sequentially as either a periodic or random process, and once a new task arrives, the sensors are obliged to send the information accumulated about the previous task. On the other hand, we assume that the end user possesses knowledge about utility states of the sensors and may request the sensors to respond to the task before a new task arrives<sup>1</sup>. In particular, this decision to **continue** for *exploration*, or to **stop** for *exploitation*, which is carried out in a centralized manner at the end user, will be the main focus of this paper. The transmission from  $S_1$  to  $D$  and  $S_2$  to  $D$  are performed simultaneously.

Let  $N, N = 0, 1, \dots$  denote the stage number, i.e., the number of time slots that have elapsed since the current task was issued. The QoI utility  $U_i(N)$  at sensor  $i, i = 1, 2$ , evolves as follows:

$$U_i(N+1) = \gamma_i U_i(N) + Q_i(N), \quad (1)$$

where  $0 < \gamma_i < 1, i = 1, 2$  are the discounting factors of the two sensors. The value  $Q_i(N)$  corresponds to the raw utility added due to information collected at stage  $N$  by sensor  $i$ , for  $i \in \{1, 2\}$ . We assume that both  $Q_1(j), j = 0, 1, \dots$ , and  $Q_2(j), j = 0, 1, \dots$  are non-negative i.i.d. random variables, that is, additional information gathered at any slot cannot have an adverse effect and utility degrades only due to age. This leads to the fol-

<sup>1</sup>Inter-node communication and coordination [19], [20] which are necessary to ensure these assumptions are satisfied are out of the scope of this paper.

lowing expression in terms of information arrivals at previous slots:

$$U_i(N) = \sum_{j=0}^{N-1} \gamma_i^{N-1-j} Q_i(j). \quad (2)$$

Sensors are heterogeneous in the sense that the  $\gamma_i$  values and  $Q_i$  processes can differ for each sensor.

In this setting, the end user will make a decision in order to maximize the total utility. The total utility function at the end user will be the summation of the individual utilities of the sensors ( $U_1(N), U_2(N)$ ).

$$U_s(N) = U_1(N) + U_2(N), \quad \forall N. \quad (3)$$

In addition, each slot where the sensors continue to gather information results in an operational cost  $\beta$ , which may occur due to sensing and processing.

## III. CONTROL OBJECTIVE

Let  $U_{i,m}, i = 1, 2$  denote the QoI that are delivered to the end user from sensor  $i$  for task  $m$  under any policy (where  $m \in \{1, 2, 3, \dots\}$ ). Similarly, let  $C(m)$  denote the total processing cost incurred on task  $m$ . This is equal to  $\beta$  times the number of slots in which the control action of **continue** was chosen for task  $m$ . Then, we consider the control objective of maximizing the time average QoI minus the average (processing and sensing) cost. Since the actions taken in a task do not effect the durations of other tasks, this is equivalent to maximizing the average QoI minus the total (processing and sensing) cost per task, which can be stated as:

$$\text{Maximize : } \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M \mathbb{E} \{U_{1,m} + U_{2,m} - C(m)\}. \quad (4)$$

Since the arrival times of tasks are independent of decisions, this maximization can be achieved by *separately* maximizing the expected QoI minus the total processing cost within each task arrival period. This problem falls into the class of *stochastic shortest path* (SSP) problems [21]. In particular, the problem always starts in the QoI state  $U_1 = U_2 = 0$  and terminates whenever the next task arrives. This SSP problem always terminates under any control policy due to the fact that the expected inter-arrival time of the tasks is finite. In the next section, we present the dynamic programming formulation for this problem.

## IV. GENERAL SOLUTION

At each stage  $N$ , the end user makes a decision to continue, or to stop. Stopping yields the current accumulated utility  $U_1(N) + U_2(N)$ . If a new task arrives on the current stage, the system is forced to stop. Else, the system can choose to continue, with the hopes of accumulating a larger utility before stopping. For clarity of exposition, we represent the state of the system as  $(U_1, U_2, N)$ , which implies that utilities  $U_1$  and  $U_2$  correspond to the values at stage  $N$ , i.e.,  $U_1(N)$  and  $U_2(N)$ , with also  $U_s$  corresponding to  $U_s(N)$ . Bellman's equation [21] for the system is:

$$V(U_1, U_2, N) = \max[U_1 + U_2, -\beta + p(N)(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - p(N))\mathbb{E}\{V(\gamma_1 U_1 + Q_1(N), \gamma_2 U_2 + Q_2(N), N+1)\}] \quad (5)$$

where  $V(U_1, U_2, N)$  is the optimal cost-to-go value function,  $\beta$  is the processing cost,  $p(N)$  is the probability that a new task arrives at the next time slot, and  $\overline{Q}_i$  is the expected value of  $Q_i$ . The term multiplying  $p(N)$  corresponds to the expected sum utility when the end user decides to continue but a task arrival occurs at the next stage, and the system is forced to stop. Note that the expectation in (5) is with respect to  $Q_1$  and  $Q_2$ , which are random variables. The Bellman equation shows that it is optimal to stop whenever the following inequality holds (and to continue otherwise):

$$U_1 + U_2 > -\beta + p(N)(\gamma_1 U_1 + \overline{Q}_1 + \gamma_2 U_2 + \overline{Q}_2) + (1 - p(N))\mathbb{E}\{V(\gamma_1 U_1 + Q_1(N), \gamma_2 U_2 + Q_2(N), N + 1)\}. \quad (6)$$

We note that the left-hand-side (LHS) corresponds to the instantaneous sum utility.

In [7], for the case with a single source, we were able to demonstrate that the optimal stopping rule is threshold-type for a class of task arrival patterns. Here, we also aim to characterize the decision by focusing on only the sum utility value, i.e.,  $U_s(N) = U_1(N) + U_2(N)$ .

To that end, we first introduce the following lemma:

*Lemma 1:* For any fixed stage  $N$ ,  $V(X, Y, N)$  is increasing in both  $X$  and  $Y$ .

*Proof:* Fix  $X, Y, N$ . Consider any  $(W, Z)$  such that  $W \geq X$  and  $Z \geq Y$ , where strict inequality holds for at least one. We want to show that  $V(W, Z, N) > V(X, Y, N)$ . Suppose we are in stage  $N$ , and let  $T$  be a random variable representing the number of slots we continue after stage  $N$  before stopping, assuming we use the optimal rule for achieving  $V(X, Y, N)$ . So  $T \in \{0, 1, 2, 3, \dots\}$ . Then we have:

$$V(X, Y, N) = \mathbb{E}\left\{\gamma_1^T X + \gamma_2^T Y + \sum_{j=1}^T [\gamma_1^{T-j} Q_1(N+j) + \gamma_2^{T-j} Q_2(N+j)]\right\}. \quad (7)$$

Now assume we are in stage  $N$ , but start with quantities  $(W, Z)$ . However, we use the *same* stopping rule  $T$ , associated with the lesser quantities  $(X, Y)$ . Then this is a particular algorithm, not necessarily optimal, and so:

$$V(W, Z, N) \geq \mathbb{E}\left\{\gamma_1^T W + \gamma_2^T Z + \sum_{j=1}^T [\gamma_1^{T-j} Q_1(N+j) + \gamma_2^{T-j} Q_2(N+j)]\right\} = \mathbb{E}\{\gamma_1^T\}(W - X) + \mathbb{E}\{\gamma_2^T\}(Z - Y) + V(X, Y, N), \quad (8)$$

where the last line follows from (7). We know that  $\mathbb{E}\{\gamma_1^T\} > 0$  and  $\mathbb{E}\{\gamma_2^T\} > 0$ , and so the sum of the first two terms on the right-hand-side above must be strictly positive. ■

Consider the one-dimensional version of the problem with only one sensor, and a single discount factor  $\gamma$ . Suppose  $V_H(U, N)$  and  $V_L(U, N)$  are defined as the optimal value functions for sensor state  $U$  at stage  $N$ , which solve the

one-dimensional problems associated with discount factors  $\gamma_H = \max(\gamma_1, \gamma_2)$  and  $\gamma_L = \min(\gamma_1, \gamma_2)$ , i.e., result from the following Bellman equations:

$$V_L(U, N) = \max[U, -\beta + p(N)(\gamma_L U + \overline{Q}_1 + \overline{Q}_2) + (1 - p(N))\mathbb{E}\{V_L(\gamma_L U + Q_1(N) + Q_2(N), N + 1)\}] \quad (9)$$

$$V_H(U, N) = \max[U, -\beta + p(N)(\gamma_H U + \overline{Q}_1 + \overline{Q}_2) + (1 - p(N))\mathbb{E}\{V_H(\gamma_H U + Q_1(N) + Q_2(N), N + 1)\}]. \quad (10)$$

*Lemma 2:* For all  $U_1 \geq 0$ ,  $U_2 \geq 0$ , and integers  $N \in \{1, 2, \dots\}$ , we have:

$$V_L(U_1 + U_2, N) \leq V(U_1, U_2, N) \leq V_H(U_1 + U_2, N). \quad (11)$$

*Proof:* We prove that  $V_L(U_1 + U_2, N) \leq V(U_1, U_2, N)$ . The proof of the other inequality is similar. Let  $\tau_L$  represent the optimal stopping time for the one-dimensional problem corresponding to  $\gamma_L$ , stage  $N$ , and utility  $U_1 + U_2$ . That is, given utility  $U_1 + U_2$  at stage  $N$ ,  $\tau_L$  is the random variable representing the optimal time to stop. That is:

$$V_L(U_1 + U_2, N) = \mathbb{E}\left\{\gamma_L^{\tau_L}(U_1 + U_2) + \sum_{j=1}^{\tau_L} \gamma_L^{\tau_L-j} [Q_1(N+j) + Q_2(N+j)]\right\}. \quad (12)$$

Define  $\tau$  as the optimal stopping time for the two-dimensional problem starting at stage  $N$  with utilities  $U_1, U_2$ , so that:

$$V(U_1, U_2, N) = \mathbb{E}\left\{U_1 \gamma_1^\tau + U_2 \gamma_2^\tau + \sum_{j=1}^{\tau} [\gamma_1^{\tau-j} Q_1(N+j) + \gamma_2^{\tau-j} Q_2(N+j)]\right\}. \quad (13)$$

Then, because  $\tau$  is optimal and maximizes the right-hand-side above, we have:

$$V(U_1, U_2, N) = \mathbb{E}\left\{U_1 \gamma_1^\tau + U_2 \gamma_2^\tau + \sum_{j=1}^{\tau} [\gamma_1^{\tau-j} Q_1(N+j) + \gamma_2^{\tau-j} Q_2(N+j)]\right\} \geq \mathbb{E}\left\{U_1 \gamma_1^{\tau_L} + U_2 \gamma_2^{\tau_L} + \sum_{j=1}^{\tau_L} [\gamma_1^{\tau_L-j} Q_1(N+j) + \gamma_2^{\tau_L-j} Q_2(N+j)]\right\} \quad (14)$$

$$\geq \mathbb{E}\left\{U_1 \gamma_1^{\tau_L} + U_2 \gamma_2^{\tau_L} + \sum_{j=1}^{\tau_L} [\gamma_1^{\tau_L-j} Q_1(N+j) + \gamma_2^{\tau_L-j} Q_2(N+j)]\right\} \quad (15)$$

$$\geq \mathbb{E}\left\{\gamma_L^{\tau_L}(U_1 + U_2) + \sum_{j=1}^{\tau_L} \gamma_L^{\tau_L-j} [Q_1(N+j) + Q_2(N+j)]\right\} \quad (16)$$

$$= V_L(U_1 + U_2, N), \quad (17)$$

which holds regardless of the relation between  $\tau$  and  $\tau_L$ . ■

While (6) provides a stopping rule, by the Lemma 2, we know that the right-hand-side of (6) satisfies:

*Lemma 3:* Suppose it is optimal to stop for the one-dimensional problem at stage  $N$ , with  $\gamma_H$  and  $U_1 + U_2$ . Then it is also optimal to stop for the two-dimensional problem. Similarly, if it is optimal to continue for the one-dimensional problem with  $\gamma_L$  and with  $U_1 + U_2$ , then it is also optimal to continue in the two-dimensional problem.

*Proof:* We just show the first statement, concerning comparison to the one-dimensional problem with  $\gamma_H$ , for brevity (the proof for  $\gamma_L$  is similar). If it is optimal to stop for the one-dimensional problem with  $\gamma_H$ , then we must have:

$$U_1 + U_2 > -\beta + p(N)(\gamma_H(U_1 + U_2) + \overline{Q_1} + \overline{Q_2}) + (1 - p(N))\mathbb{E}\{V(\gamma_H U_1 + Q_1(N), \gamma_H U_2 + Q_2(N), N + 1)\} \quad (18)$$

$$\geq -\beta + p(N)(\gamma_1 U_1 + \gamma_2 U_2 + \overline{Q_1} + \overline{Q_2}) + (1 - p(N))\mathbb{E}\{V(\gamma_H U_1 + Q_1(N), \gamma_H U_2 + Q_2(N), N + 1)\} \quad (19)$$

$$\geq -\beta + p(N)(\gamma_1 U_1 + \gamma_2 U_2 + \overline{Q_1} + \overline{Q_2}) + (1 - p(N))\mathbb{E}\{V(\gamma_1 U_1 + Q_1(N), \gamma_2 U_2 + Q_2(N), N + 1)\}, \quad (20)$$

where the first inequality is the stopping condition for the one-dimensional problem, the second follows from Lemma 2, and the third follows because  $V(X, Y, N)$  is non-decreasing in  $X$  and  $Y$ . Therefore,  $U_1 + U_2$  satisfies the condition (6). ■

In the following sections, we consider cases where the one-dimensional problems satisfy a threshold rule. Then, from Lemma 3, at each stage  $N$ , we can define thresholds  $T_H(N)$  as the threshold for the one-dimensional problem with  $\gamma_H$ , and  $T_L(N)$  as the threshold for the one-dimensional problem with  $\gamma_L$ , to claim that if  $U_1 + U_2 > T_H(N)$  it is definitely optimal to stop, and if  $U_1 + U_2 < T_L(N)$  it is definitely optimal to continue.

In general, the stopping rule depends on the vector  $(U_1(N), U_2(N))$  at each stage  $N$ , not only on the sum. We next characterize this region for two specific task arrival processes: (i) i.i.d. memoryless task arrivals with geometric interarrival times, and (ii) periodic task arrivals.

We first present simple monotonicity and convexity properties of  $V(X, Y, N)$ :

*Lemma 4:* For any fixed stage  $N$ ,  $V(X, Y, N)$  is convex in  $(X, Y)$ .

*Proof:* Consider any non-negative vectors  $(X, Y)$  and  $(W, Z)$  and any non-negative values  $p, q$  such that  $p + q = 1$ . We want to show that:

$$pV(X, Y, N) + qV(W, Z, N) \geq V(pX + qW, pY + qZ, N). \quad (21)$$

As in the proof of Lemma 1, we have:

$$\begin{aligned} V(X, Y, N) &= \mathbb{E} \left\{ \gamma_1^{T_{X,Y}} X + \gamma_2^{T_{X,Y}} Y \right. \\ &\quad \left. + \sum_{j=1}^{T_{X,Y}} [\gamma_1^{T_{X,Y}-j} Q_1(N+j) + \gamma_2^{T_{X,Y}-j} Q_2(N+j)] \right\} \quad (22) \end{aligned}$$

$$\begin{aligned} V(W, Z, N) &= \mathbb{E} \left\{ \gamma_1^{T_{W,Z}} W + \gamma_2^{T_{W,Z}} Z \right. \\ &\quad \left. + \sum_{j=1}^{T_{W,Z}} [\gamma_1^{T_{W,Z}-j} Q_1(N+j) + \gamma_2^{T_{W,Z}-j} Q_2(N+j)] \right\} \quad (23) \end{aligned}$$

where  $T_{X,Y}$  is the random variable associated with the optimal stopping rule, assuming we start in stage  $N$  with utilities  $(X, Y)$ , and  $T_{W,Z}$  is the corresponding optimal stopping time given we start with utilities  $(W, Z)$ . By definition of optimality, the values of  $V(X, Y, N)$  and  $V(W, Z, N)$  are greater than or equal to the corresponding values when a random variable  $\tilde{T}$  corresponding to any other stopping rule is used inside the expectations on the right-hand-side of (22) and (23), respectively:

$$\begin{aligned} V(X, Y, N) &\geq \mathbb{E} \left\{ \gamma_1^{\tilde{T}} X + \gamma_2^{\tilde{T}} Y \right. \\ &\quad \left. + \sum_{j=1}^{\tilde{T}} [\gamma_1^{\tilde{T}-j} Q_1(N+j) + \gamma_2^{\tilde{T}-j} Q_2(N+j)] \right\} \quad (24) \end{aligned}$$

$$\begin{aligned} V(W, Z, N) &\geq \mathbb{E} \left\{ \gamma_1^{\tilde{T}} W + \gamma_2^{\tilde{T}} Z \right. \\ &\quad \left. + \sum_{j=1}^{\tilde{T}} [\gamma_1^{\tilde{T}-j} Q_1(N+j) + \gamma_2^{\tilde{T}-j} Q_2(N+j)] \right\}. \quad (25) \end{aligned}$$

Multiplying (24) and (25) by  $p$  and  $q$ , respectively, and adding gives:

$$\begin{aligned} pV(X, Y, N) + qV(W, Z, N) &\geq \mathbb{E} \{ \gamma_1^{\tilde{T}} (pX + qW) + \gamma_2^{\tilde{T}} (pY + qZ) \} \\ &\quad + \mathbb{E} \left\{ \sum_{j=1}^{\tilde{T}} [\gamma_1^{\tilde{T}-j} Q_1(N+j) + \gamma_2^{\tilde{T}-j} Q_2(N+j)] \right\}. \quad (26) \end{aligned}$$

Now define  $\tilde{T}$  as the random variable associated with the optimal stopping rule in stage  $N$ , given utilities  $(pX + qW, pY + qZ)$ . Then the right-hand-side of (26) is the optimal value  $V(pX + qW, pY + qZ)$ , proving (21). ■

Convexity is a useful property, and implies (by Jensen's inequality) that for any non-negative random variables  $A$  and  $B$  we have:

$$\mathbb{E} \{ V(A, B, N) \} \geq V(\mathbb{E} \{ A \}, \mathbb{E} \{ B \}, N). \quad (27)$$

## V. GEOMETRIC ARRIVALS

We first consider the case where the task arrival is an i.i.d. Bernoulli process with rate  $0 < \lambda < 1$ . This means that  $p(N) = \lambda$  for all  $N$  and the interarrival times are geometrically distributed with mean  $1/\lambda$ . Due to the memoryless property of the

Geometric distribution, we can drop the stage index while representing the system state, and the Bellman equation becomes:

$$V(U_1, U_2) = \max[U_1 + U_2, -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - \lambda)\mathbb{E}\{V(\gamma_1 U_1 + Q_1, \gamma_2 U_2 + Q_2)\}]. \quad (28)$$

First, the optimal rule is to stop if the following is satisfied:

$$U_1 + U_2 > -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - \lambda)\mathbb{E}\{V(\gamma_1 U_1 + Q_1, \gamma_2 U_2 + Q_2)\}. \quad (29)$$

We first present the following structural result:

*Theorem 1:* For any instantaneous sum utility value  $U_s = U_1 + U_2$ , there exists thresholds  $T_H$  and  $T_L$  ( $T_L \leq T_H$ ) which depend on  $\gamma_i$ , distributions of  $Q_i$ ,  $i = 1, 2$ , and  $\beta$ , such that:

- If  $U_s > T_H$ , it is always optimal to stop.
- If  $U_s < T_L$ , it is always optimal to continue.

*Proof:* We refer to the full proof at [7] for the optimality of a threshold-type stopping rule for the one-dimensional case with geometric task arrivals. Accordingly, the two-threshold rule readily follows from Lemma 3. Note that these thresholds are independent from the stage number. ■

While this characterization throws light for cases when the sum utility is too large or too small, next we pursue solutions which accommodate for general  $(U_1, U_2)$  pairs.

#### A. Value Iteration Approximation

To characterize  $V(U_1, U_2)$ , we utilize the well known value iteration algorithm [21]. The general procedure is given as follows:

$$V_0(U_1, U_2) = 0 \quad \forall U_1, U_2 \quad (30)$$

$$V_{k+1}(U_1, U_2) = \max[U_1 + U_2, -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - \lambda)\mathbb{E}\{V_k(\gamma_1 U_1 + Q_1, \gamma_2 U_2 + Q_2)\}]. \quad (31)$$

The difficulty is that the  $V_k(\cdot)$  functions grow increasingly complex as functions of  $(U_1, U_2)$ . We can use convexity and Jensen's inequality via (27) to obtain a simple lower bound on the expectation of the value function in (31):

$$\mathbb{E}\{V_k(\gamma_1 U_1 + Q_1, \gamma_2 U_2 + Q_2)\} \geq V_k(\gamma_1 U_1 + \overline{Q_1}, \gamma_2 U_2 + \overline{Q_2}).$$

Therefore, by (31):

$$V_{k+1}(U_1, U_2) \geq \max[U_1 + U_2, -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - \lambda)V_k(\gamma_1 U_1 + \overline{Q_1}, \gamma_2 U_2 + \overline{Q_2})]. \quad (32)$$

This inspires the following iterations for *bounding functions*  $\tilde{V}_k(U_1, U_2)$ :

$$\tilde{V}_0(U_1, U_2) = 0, \quad \forall U_1 \geq 0, U_2 \geq 0 \quad (33)$$

$$\tilde{V}_{k+1}(U_1, U_2) = \max[U_1 + U_2, -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - \lambda)\tilde{V}_k(\gamma_1 U_1 + \overline{Q_1}, \gamma_2 U_2 + \overline{Q_2})]. \quad (34)$$

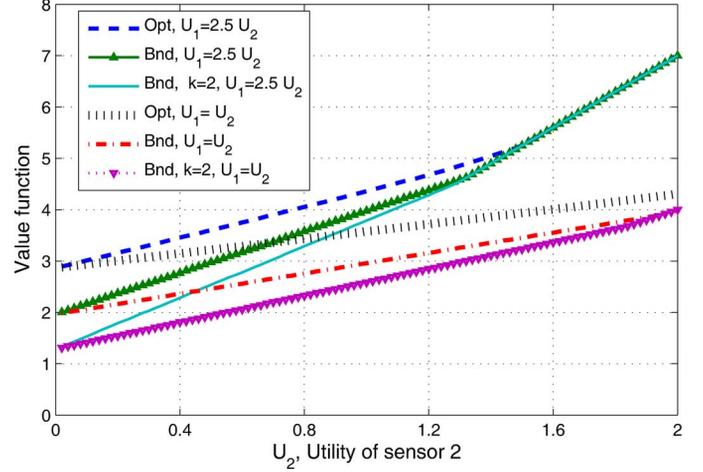


Fig. 2. Value functions for varying QoI utility state 2,  $\gamma_1 = 0.8$ ,  $\gamma_2 = 0.5$ .

The iterations (33)-(34) converge, because we observe they are the exact value iterations in a system where rewards  $(Q_1, Q_2)$  are deterministically equal to the constants  $(\overline{Q_1}, \overline{Q_2})$  on every stage.

*Lemma 5:* For each  $k \in \{0, 1, 2, \dots\}$  and every non-negative vector  $(U_1, U_2)$ , we have:

$$V_k(U_1, U_2) \geq \tilde{V}_k(U_1, U_2).$$

*Proof:* It holds for  $k = 0$ . Suppose it holds for a general  $k$ . We prove it also for  $k + 1$ . We have by (32):

$$V_{k+1}(U_1, U_2) \geq \max[U_1 + U_2, -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - \lambda)V_k(\gamma_1 U_1 + \overline{Q_1}, \gamma_2 U_2 + \overline{Q_2})] \quad (35)$$

$$\geq \max[U_1 + U_2, -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - \lambda)\tilde{V}_k(\gamma_1 U_1 + \overline{Q_1}, \gamma_2 U_2 + \overline{Q_2})], \quad (36)$$

where the final inequality holds by the induction assumption  $V_k(\cdot) \geq \tilde{V}_k(\cdot)$ . ■

We demonstrate the closeness of the bounding functions to the optimal value function in Fig. 2, for  $\beta = 0.01$ ,  $\gamma_1 = 0.8$ ,  $\gamma_2 = 0.5$ , with  $Q_1$  and  $Q_2$  independent Bernoulli with  $\overline{Q_1} = 0.9$ ,  $\overline{Q_2} = 0.4$ , with  $\lambda = 0.25$ . For illustrative purposes, we vary utility of sensor 2, and display the value functions for the special cases where  $U_1 = U_2$ , i.e., both sensor states are equal, and the case when  $U_1 = 2.5U_2$ . ‘‘Opt’’ corresponds to the optimal value function for each of these two cases, which was obtained by value iteration performed using the full distribution of  $Q_1$  and  $Q_2$ . On the other hand, ‘‘Bnd’’ corresponds to the value function resulting from the bounding algorithm with a sufficient number of iterations, while ‘‘Bnd,  $k = 2$ ’’ depicts the value function for the simpler case where only 2 iterations are used for the bounding function. Next, we utilize these bounding functions in a reduced-complexity adaptive sampling algorithm.

### B. Adaptive Sampling via Bounding Functions

Perfect value iterations require extensive computation requirements due to expectations over  $Q_1, Q_2$  values at each iteration step. In order to reduce computational requirements in adaptive sampling, we propose algorithms which are inspired by the approximate dynamic programming approach called Certainty Equivalent Control (CEC) [21]. CEC methods execute dynamic programming algorithms for cases where the uncertain quantities are fixed at some *typical* values. A natural choice for the typical values are mean values of random variables. Doing so results in the following Bellman equation:

$$\hat{V}(U_1, U_2) = \max[U_1 + U_2, -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - \lambda)\hat{V}(\gamma_1 U_1 + \overline{Q_1}, \gamma_2 U_2 + \overline{Q_2})]. \quad (37)$$

We immediately note that applying value iteration to get the value functions  $\hat{V}(U_1, U_2)$  is identical to the iterations for bounding functions in (33)-(34).

Accordingly, we propose an approximately optimal adaptive sampling algorithm which utilizes bounding functions instead of the actual value function, that is, for a given bounding function  $\tilde{V}(U_1, U_2)$ ,

(i) **Stop**, if

$$U_1 + U_2 > -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - \lambda) \times \tilde{V}(\gamma_1 U_1 + \overline{Q_1}, \gamma_2 U_2 + \overline{Q_2}). \quad (38)$$

(ii) **Continue**, if

$$U_1 + U_2 \leq -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - \lambda) \times \tilde{V}(\gamma_1 U_1 + \overline{Q_1}, \gamma_2 U_2 + \overline{Q_2}). \quad (39)$$

Indeed, from Lemma 4 and Lemma 5, we have

$$\mathbb{E}(V(\gamma_1 U_1 + Q_1, \gamma_2 U_2 + Q_2)) \geq V(\gamma_1 U_1 + \overline{Q_1}, \gamma_2 U_2 + \overline{Q_2}) \geq \tilde{V}(\gamma_1 U_1 + \overline{Q_1}, \gamma_2 U_2 + \overline{Q_2}).$$

This implies that if the continue condition (39) for approximate algorithm is satisfied, it will definitely be satisfied in the actual optimal algorithm(6).

Hence, we denote set of  $(U_1, U_2)$  points which lead to continue decisions in the approximate algorithm as the *Sufficient Region for Continuing*, in the sense that satisfying (39) is sufficient to guarantee that the optimal decision is to continue.

It is possible to further simplify the adaptive sampling algorithm by using a finite order-iteration for the bounding functions. For such simplified algorithms, it is possible to approximate the SRC with the union of finite number of halfspaces.

Consider  $V_1(U_1, U_2)$  in (31), where satisfying the following relation leads to the fact that continuing is optimal (Note that  $V_1(U_1, U_2) = \tilde{V}_1(U_1, U_2)$ ).

CR 1: (For  $V_1(U_1, U_2)$ )

$$U_1 + U_2 \leq -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}),$$

equivalent to the condition

$$(1 - \lambda\gamma_1)U_1 + (1 - \lambda\gamma_2)U_2 \leq -\beta + \lambda\overline{Q_1} + \lambda\overline{Q_2}. \quad (40)$$

From  $\tilde{V}_2(U_1, U_2)$  in (34), if  $(U_1, U_2)$  satisfies either of the

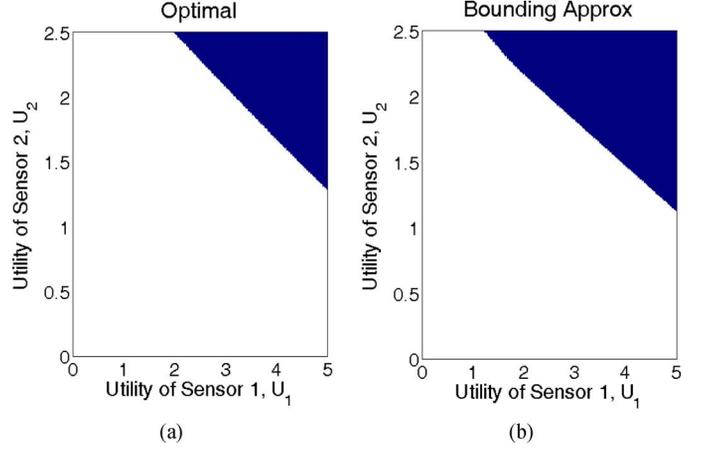


Fig. 3. Stopping regions for a) Optimal Algorithm b) Bounding Approximate Algorithm,  $\gamma_1 = 0.8, \gamma_2 = 0.6$ .

following 2 relations, continuing is the optimal decision.

CR 2.1: (For  $\tilde{V}_2(U_1, U_2)$ )

$$U_1 + U_2 \leq -\beta + \gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}, \quad (41)$$

equivalent to the relation

$$(1 - \gamma_1)U_1 + (1 - \gamma_2)U_2 \leq -\beta + \overline{Q_1} + \overline{Q_2}. \quad (42)$$

CR 2.2: (For  $\tilde{V}_2(U_1, U_2)$ )

$$U_1 + U_2 \leq -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - \lambda)(-\beta + \lambda(\gamma_1(\gamma_1 U_1 + \overline{Q_1} + \overline{Q_1}) + \gamma_2(\gamma_2 U_2 + \overline{Q_2}) + \overline{Q_2})), \quad (43)$$

equivalent to the condition

$$(1 - \lambda\gamma_1 - (1 - \lambda)\lambda\gamma_1^2)U_1 + (1 - \lambda\gamma_2 - (1 - \lambda)\lambda\gamma_2^2)U_2 \leq (-2 + \lambda)\beta + (1 + (1 - \lambda)\lambda\gamma_1)\overline{Q_1} + (1 + (1 - \lambda)\lambda\gamma_2)\overline{Q_2}. \quad (44)$$

For clarity, we do not present the approximate  $V_k$  forms or the relations explicitly for  $k > 2$ . The main observation is that after  $k = 2$ , some of conditions from the previous iterations are preserved, along with additional continuing relations.

The  $(U_1, U_2)$  pairs that satisfy either of the continuing conditions above lead to a subset of the continuing region for a given approximation order, i.e., the  $(U_1, U_2)$  pairs for which the continuing decision will definitely be made.

We illustrate stopping regions for the optimal algorithm and approximate algorithm with bounding functions in Figs. 3 and 4, specifically for  $\beta = 0.01, \gamma_1 = 0.8, Q_1$  and  $Q_2$  independent Bernoulli random variables with  $\overline{Q_1} = 0.9, \overline{Q_2} = 0.5$ , with  $\lambda = 0.2$ . Fig. 3 depicts the case with  $\gamma_2 = 0.6$  and Fig. 4 depicts the case with  $\gamma_2 = 0.333$ . Blue regions correspond to  $U_1, U_2$  where stopping is optimal and white regions correspond to regions where continuing is chosen by the algorithms.

As stated in Theorem 1, sum utility does not solely define the continuing condition for an intermediate region of  $U_s$  vales. In particular, the sensor with larger  $\gamma$  is encouraged to wait for a larger QoI before making the stopping decision. In other words, it can afford to wait more for exploration since its utility is not degraded as fast as the other sensor. Moreover, as expected, the

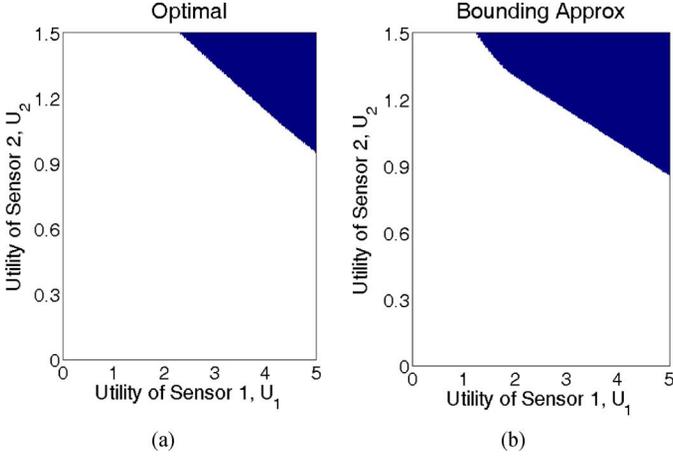


Fig. 4. Stopping Regions, a) Optimal Algorithm b) Bounding Approximate Algorithm,  $\gamma_1 = 0.8, \gamma_2 = 0.333$ .

set of points that are guaranteed to be in the continuing region shrinks as  $\gamma_2$  reduces. Consequently, the maximum  $U_s$  value which definitely leads to a continue decision reduces with lower  $\gamma_i, i = 1, 2$ .

### C. Hybrid Algorithms

We also propose a hybrid adaptive sampling algorithm which has intermediate complexity. While the algorithm uses the bounding functions to approximate the value functions, it uses the full distribution in the decision stage, i.e., chooses the following actions:

(i) **Stop**, if

$$U_1 + U_2 > -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - \lambda)\mathbb{E}\{\tilde{V}(\gamma_1 U_1 + Q_1, \gamma_2 U_2 + Q_2)\}. \quad (45)$$

(ii) **Continue**, if

$$U_1 + U_2 \leq -\beta + \lambda(\gamma_1 U_1 + \overline{Q_1} + \gamma_2 U_2 + \overline{Q_2}) + (1 - \lambda)\mathbb{E}\{\tilde{V}(\gamma_1 U_1 + Q_1, \gamma_2 U_2 + Q_2)\}. \quad (46)$$

We will demonstrate that the hybrid approach will further approach the performance of the optimal adaptive sampling algorithm.

## VI. PERIODIC ARRIVALS

We next consider the case where tasks arrive periodically, once every  $D$  slots. Thus, each task has a fixed deadline of  $D$  slots.  $p(N)$  in (6) becomes  $p(D - 1) = 1$  and  $p(N) = 0$  for all  $N \neq D - 1$ , and we have:

$$V(U_1, U_2, D) = U_1 + U_2, \quad (47)$$

$$V(U_1, U_2, D - 1) = \max \left[ U_1 + U_2, -\beta + \gamma_1 U_1 + \gamma_2 U_2 + \overline{Q_1} + \overline{Q_2} \right], \quad (48)$$

$$V(U_1, U_2, N) = \max \left[ U_1 + U_2, -\beta + \mathbb{E}\{V(\gamma_1 U_1 + Q_1(N), \gamma_2 U_2 + Q_2(N), N + 1)\} \right], \quad (49)$$

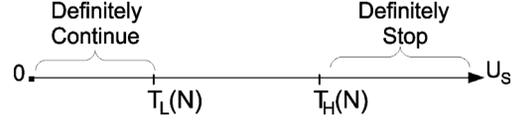


Fig. 5. Upper and lower thresholds for given  $U_s$ .

where  $0 \leq N \leq D - 2$  in (49). Thus, it is optimal to stop in any state  $(U_1, U_2, D - 1)$  if  $U_1 + U_2 > -\beta + \gamma_1 U_1 + \gamma_2 U_2 + \overline{Q_1} + \overline{Q_2}$ , i.e.,  $U_1(1 - \gamma_1) + U_2(1 - \gamma_2) > -\beta + \overline{Q_1} + \overline{Q_2}$ . Similarly, it is optimal to stop in any state  $(U_1, U_2, N), N < D - 1$ , if

$$U_1 + U_2 > -\beta + \mathbb{E}\{V(\gamma_1 U_1 + Q_1(N), \gamma_2 U_2 + Q_2(N), N + 1)\}. \quad (50)$$

Again, we first consider the option of making the stopping decision depending on only the sum utility, where we have the following result:

**Theorem 2:** For any instantaneous sum utility value  $U_s$ , at stage  $N$ , there exists thresholds  $T_H(N)$  and  $T_L(N)$  ( $T_L(N) \leq T_H(N)$ ) which depend on  $\gamma_i$ , distributions of  $Q_i, i = 1, 2$ , and  $\beta$ , such that (Fig. 5):

- If  $U_s > T_H(N)$ , it is always optimal to stop.
- If  $U_s < T_L(N)$ , it is always optimal to continue.

*Proof:* We again utilize Lemma 3 and the existence of a threshold-type rule for the one-dimensional case. We refer readers to [7] for the proof for the case with periodic task arrivals for a single sensor device. ■

In order to pursue more general two-dimensional stopping rules, let us re-visit (49). These conditions give a set of  $D$  equations to be satisfied simultaneously. Since the expressions are rather straightforward for  $N = D$  and  $N = D - 1$ , we pursue the solution via backward induction. As in the geometric arrival case, we use a sequence of bounding functions which use mean values of random variables. By adopting Lemma 4 and Lemma 5 for the expectation terms in the RHS of (49) we obtain lower bound approximations  $\tilde{V}(U_1, U_2, N) \leq V(U_1, U_2, N)$  to the value functions. For instance,

$$\tilde{V}(U_1, U_2, D - 2) = \begin{cases} U_1 + U_2, & \text{if } U_1 + U_2 > \max[-\beta + \gamma_1 U_1 + \gamma_2 U_2 + \overline{Q_1} + \overline{Q_2}, \\ & -2\beta + \gamma_1^2 U_1 + \gamma_2^2 U_2 + \overline{Q_1}(1 + \gamma_1) + \overline{Q_2}(1 + \gamma_2)], \\ -\beta + \gamma_1 U_1 + \gamma_2 U_2 + \overline{Q_1} + \overline{Q_2}, & \text{if } -\beta + \gamma_1 U_1 + \gamma_2 U_2 + \overline{Q_1} + \overline{Q_2} > \max[U_1 + U_2, \\ & -2\beta + \gamma_1^2 U_1 + \gamma_2^2 U_2 + \overline{Q_1}(1 + \gamma_1) + \overline{Q_2}(1 + \gamma_2)], \\ -2\beta + \gamma_1^2 U_1 + \gamma_2^2 U_2 + \overline{Q_1}(1 + \gamma_1) + \overline{Q_2}(1 + \gamma_2), & \text{else} \end{cases}$$

If  $U_1 + U_2$  is less than the resulting term in the RHS, it will definitely be less than the term that would have resulted using the original value iterates, since  $\tilde{V}_2(U_1, U_2, D - 2) \leq V_2(U_1, U_2, D - 2)$ . Consequently, the Sufficient Region for Continuing relations at stage  $D - 2$  are given by the union of the half-spaces defined by  $U_1(1 - \gamma_1) + U_2(1 - \gamma_2) < -\beta + \overline{Q_1} + \overline{Q_2}$ , and  $U_1(1 - \gamma_1^2) + U_2(1 - \gamma_2^2) < -2\beta + \overline{Q_1}(1 + \gamma_1) + \overline{Q_2}(1 + \gamma_2)$ . We next generalize this result for an arbitrary stage number.

**Theorem 3:** At any stage  $N$ , it can be shown that the sufficient region for continuing is given by the union of following  $D - N$  two-dimensional regions:

$$U_1(1 - \gamma_1) + U_2(1 - \gamma_2) < -\beta + \overline{Q}_1 + \overline{Q}_2 \quad (51)$$

$$U_1(1 - \gamma_1^2) + U_2(1 - \gamma_2^2) < -2\beta + \overline{Q}_1(1 + \gamma_1) + \overline{Q}_2(1 + \gamma_2) \quad (52)$$

$\vdots$

$$U_1(1 - \gamma_1^{D-N}) + U_2(1 - \gamma_2^{D-N}) < -(D - N)\beta + \overline{Q}_1 \frac{1 - \gamma_1^{D-N}}{1 - \gamma_1} + \overline{Q}_2 \frac{1 - \gamma_2^{D-N}}{1 - \gamma_2}. \quad (53)$$

*Proof:* Recall the recursive relation for the bounding functions:

$$\begin{aligned} \tilde{V}(U_1, U_2, N) \\ = \max \left[ U_1 + U_2, -\beta + \tilde{V}(\gamma_1 U_1 + \overline{Q}_1, \gamma_2 U_2 + \overline{Q}_2, N + 1) \right]. \end{aligned}$$

At each stage backwards, we argue that an additional relation for continuing is added to the rest satisfying the form (51)–(53). This relation is introduced due to an additional possible function form for the second term of the max operator in (54). At stage  $D$  and  $D - 1$  this is readily shown by (51)–(52).

Next, assume that the latest bounding function form added at stage  $L$  stage backwards, i.e.,  $N = D - L$  is

$$\gamma_1^{D-L} U_1 + \gamma_2^{D-L} U_2 - (D - L)\beta + \overline{Q}_1(1 + \gamma_1 + \gamma_1^2 + \dots + \gamma_1^{D-L-1})\overline{Q}_2(1 + \gamma_2 + \gamma_2^2 + \dots + \gamma_2^{D-L-1}). \quad (54)$$

The recursion will result  $\tilde{V}(U_1, U_2, N - 1)$  being the maximum of two terms, first being  $U_1 + U_2$ , or inserting the expression in (54) to the RHS of the recursion resulting in:

$$\begin{aligned} -\beta + \{ -(D - L)\beta + \gamma_1^{D-L}(\gamma_1 U_1 + \overline{Q}_1) + \gamma_2^{D-L}(\gamma_2 U_2 + \overline{Q}_2) \\ + \overline{Q}_1(1 + \gamma_1 + \gamma_1^2 + \dots + \gamma_1^{D-L-1}) + \overline{Q}_2(1 + \gamma_2 + \gamma_2^2 + \dots + \gamma_2^{D-L-1}) \} \end{aligned} \quad (55)$$

$$\begin{aligned} = -\beta(D - L + 1) + \gamma_1^{D-L+1} U_1 + \gamma_2^{D-L+1} U_2 \\ + \overline{Q}_1(1 + \gamma_1 + \dots + \gamma_1^{D-L}) + \overline{Q}_2(1 + \gamma_2 + \dots + \gamma_2^{D-L}) \end{aligned} \quad (56)$$

$$\begin{aligned} = -\beta(D - L + 1) + \gamma_1^{D-L+1} U_1 + \gamma_2^{D-L+1} U_2 \\ + \overline{Q}_1 \frac{1 - \gamma_1^{D-L+1}}{1 - \gamma_1} + \overline{Q}_2 \frac{1 - \gamma_2^{D-L+1}}{1 - \gamma_2}, \end{aligned} \quad (57)$$

i.e.,

$$\begin{aligned} \tilde{V}(U_1, U_2, N - 1) = \max \left[ U_1 + U_2, -\beta(D - L + 1) \right. \\ \left. + \gamma_1^{D-L+1} U_1 + \gamma_2^{D-L+1} U_2 + \overline{Q}_1 \frac{1 - \gamma_1^{D-L+1}}{1 - \gamma_1} \right. \\ \left. + \overline{Q}_2 \frac{1 - \gamma_2^{D-L+1}}{1 - \gamma_2} \right]. \end{aligned} \quad (58)$$

This implies a continuing condition in accordance with the structural form of (51)–(53) for  $L - 1$ , i.e., each stage backwards

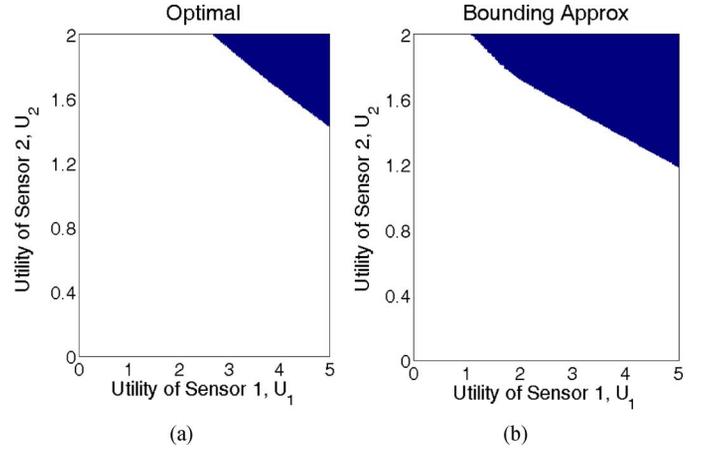


Fig. 6. Stopping Regions at stage 5, period  $D = 15$ ,  $\gamma_1 = 0.8$ ,  $\gamma_2 = 0.6$ ,  $\overline{Q}_1 = 0.9$  and  $\overline{Q}_2 = 0.4$  Bernoulli.

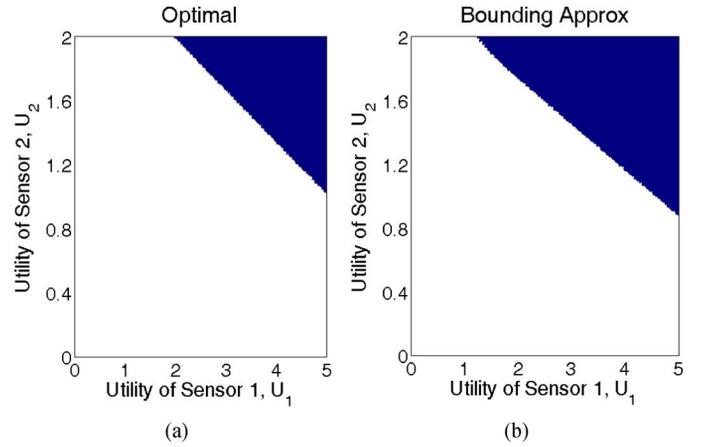


Fig. 7. Stopping Regions at stage 12, period  $D = 15$ ,  $\gamma_1 = 0.8$ ,  $\gamma_2 = 0.3$ ,  $\overline{Q}_1 = 0.9$  and  $\overline{Q}_2 = 0.4$  Bernoulli.

an additional expression with the structured form is introduced to be compared with  $U_1 + U_2$  as a continuing condition. ■

Next, we establish a relationship between the maximum values  $U_1 + U_2$  can take to be still in order to definitely make a continue decision for different stages.

**Lemma 6:** The maximum value of  $U_1 + U_2$  to guarantee continue decision at stage  $N$ ,  $T_N^p$ , is non-increasing with stage number, i.e.,  $T_N^p \geq T_{N+1}^p$ .

*Proof:* Recall that as the stage index increases, the number of possible continuing relations that can be satisfied reduces. Moreover, due to the structural form (51)–(53), at any stage  $N$ , the remaining  $D - N$  continuing conditions are identical to  $D - N$  of the conditions at previous stages. As a result, the sufficient region for continuing at a stage is a subset of the sufficient region for continuing of previous stages (Figs. 6–7). Accordingly, by graphical arguments, it can be readily shown that the maximum  $U_1 + U_2$  that continuing is guaranteed to be the optimal decision is non-increasing with stage number. Specifically, the  $U_1 + U_2 = c$  level curves for decreasing  $c$  intersect with the sufficient region for continuing either at a smaller  $c$  value, or at identical  $c$  compared with regions of previous stages. ■

**Remark 1:** This result is in line with the intuition that as the deadline approaches, the thresholds on sum utility to stop

decrease as there are fewer opportunities to increase the QoI utility.

Finally, for the periodic case, we propose the following approximate adaptive sampling algorithm using the bounding functions:

- Start from stage 1. Repeat until **stop** decision is taken or a new task arrives (Stage =  $D$ ). For stage  $N$ , observe  $U_1(N), U_2(N)$ . Start by checking (51). If it is satisfied, decide **continue** and increment stage number as  $N + 1$ .
- If (51) is not satisfied, check conditions up to (53) until any of the conditions are satisfied. If none of the  $D - N$  conditions from (51) to (53) are satisfied, decide **stop**.
- If **stop** decision is taken, stop the adaptive sampling algorithm and report information to the end node.
- If stage number is  $D$ , **stop**, report the information to the end node and start addressing the new task.

*Remark 2:* While it may seem that there are a potentially large number of continuing relations to check at initial stages (e.g.,  $D$  for the first stage), we argue that the actual operation is likely to result in a lower complexity. This is due to the fact that whenever one of the conditions are satisfied, the checking is aborted for that stage. Moreover, for initial stages it is likely that the utilities are low due to insufficient information accumulation and even the first continuing conditions are likely to succeed (Figs. 6–7), reducing computational time considerably. As for later stages, even though the utilities are expected to increase, the maximum number of continuing relations to be checked reduce as we approach the deadline. In this aspect, the adaptive sampling algorithm, resembles receding horizon approaches [22].

*Remark 3:* While for clarity purposes we have selected the most concise multiuser scenario with two sensors, the adaptive sampling algorithms proposed in this paper can be readily extended to the case of  $L$  sensors, with  $L > 2$ . Similar with the two-sensor case, Bellman equations will provide stopping rules comparing current sum utility of  $L$  sensors with expressions consisting of value functions as in (5). The upper and lower thresholds which guarantee correct decisions by only observing sum QoI utility over all sensors can equivalently be defined using the maximum and minimum  $\gamma$  parameters among all  $L$  sensors. The principles regarding convexity and relations via Jensen's inequality will provide bounding functions for  $L$ -dimensional value functions, which can be used for approximate algorithms. The main difference will be that the state dimensionality will increase. Accordingly, for both the case with geometric and periodic arrivals, the continuing regions will be defined by  $L$ -dimensional spaces. In order to reduce complexity of algorithms, one method we propose is to combine sensors with similar  $\gamma$  into classes, and treat them as a single sensor. This will reduce the dimensionality of the regions.

## VII. AVERAGE COST CONSTRAINED QOI-BASED ADAPTIVE SAMPLING

In this section, we consider the related problem to (4) where we are interested in maximizing the time average of the QoI reward given average cost constraints per each task. We approach

the problem by optimizing time averages in systems with independent and identical behavior over renewal frames.

The  $(U_1, U_2) = (0, 0)$  state at the beginning of each task can be considered as a *renewal* state. We are interested in developing a control algorithm which maximizes a time average of a reward process associated with the system.

We address the problem by Lyapunov optimization [18], [23]. However, unlike the classical setting where a one-shot greedy decision is performed each time slot, we are involved with a more complex problem, where renewal frames may have a different length and we are involved with a sequence of random events within each frame [24]. Hence, in contrast to one-shot decisions every slot, we must specify a *policy*, i.e., a contingency plan for making decisions over the course of the frame in reaction to the resulting system events.

More specifically, we choose a policy to minimize the drift plus penalty (or equivalently drift minus reward) every frame. We define the  $(0, 0)$  state as the recurrent state of the Markov chain defining the QoI reward evolution. Accordingly, re-visitations to the  $(0, 0)$  state lead to renewals, and the drift minus reward technique can be applied.

At the beginning of each task, i.e., renewal frame  $m$ , controller selects policy  $\pi(m)$  from the abstract policy space  $\mathcal{P}$ . We denote  $C(m)$  as the processing cost expended for frame  $m$ .  $C(m)$  is a random function of the policy  $\pi(m)$ , i.e.,  $C(m) = \hat{C}(\pi(m))$ , and is conditionally independent of events in previous frames given policy  $\pi = \pi(m)$ , identically distributed over all frames that use the same policy  $\pi$ .

Consider the frame average expressions, for  $M > 0$ :

$$\bar{C}(M) = \frac{1}{M} \sum_{m=0}^{M-1} \mathbb{E} \{C(m)\}. \quad (59)$$

We define the infinite horizon frame-average expectation  $\bar{C}$  by

$$\bar{C} = \lim_{M \rightarrow \infty} \bar{C}(M). \quad (60)$$

Long-term averages of the QoI utility reward can also be defined similarly, denoted as  $\bar{U}_1 + \bar{U}_2$ . The optimization problem we are interested is formally defined as:

$$\max \quad \bar{U}_1 + \bar{U}_2 \quad (61)$$

$$\text{s.t.} \quad \bar{C} \leq c_{av}, \quad (62)$$

where  $c_{av}$  is a bound on the long term average cost expenditure per task. Note that since task arrivals are not effected by actions taken by the sources, the above objectives and constraints are also equivalent to the time-average values. Specifically, the time average cost expenditure is equal to  $\bar{C}/\Delta$  and the time average cost constraint is  $c_{av}/\Delta$  with time-average objective  $(\bar{U}_1 + \bar{U}_2)/\Delta$ , where  $\Delta$  is the mean task inter-arrival time, e.g.,  $\Delta = 1/\lambda$  for the geometric task arrival case and  $\Delta = D$  for the periodic task arrival case.

To treat constraint (62), we define virtual queues [23]  $Z(m)$  with  $Z(0) = 0$  and update equations:

$$Z(m+1) = \max(Z(m) - c_{av}, 0) + C(m), \quad (63)$$

for task  $m = 1, 2, \dots$ . The intuition is that if we can stabilize the queue  $Z(m)$ , then the time average of the *service process*

$c_{av}$  is greater than or equal to the time average of the *arrival process*  $C(m)$ . It is well known that backpressure policies based on Lyapunov stability is the most prominent approach for stabilizing queues [18].

We define the Lyapunov function as

$$L(m) = L(Z(m)) = \frac{1}{2}Z(m)^2, \quad (64)$$

which is a measure of the size of the virtual queue backlog. Since

$$Z^2 \leq Y^2 + Z^2 + W^2 - 2Y(V - W) \quad (65)$$

for  $Z \leq \max(Y - V, 0) + W$ , the Lyapunov drift is written as [18]

$$\begin{aligned} \Delta(Z(m)) &= \mathbb{E} \{L(Z(m+1)) - L(Z(m)) | Z(m)\} \\ &= B - Z(m)\mathbb{E} \{c_{av} - C(m) | Z(m)\} \end{aligned} \quad (66)$$

where  $B$  is a bounded term, due to bounded second moments of  $C(m)$  and  $c_{av}$ .

If we add expected penalty terms (or equivalently subtract expected QoI reward terms) to the above expression, we have

$$\begin{aligned} \Delta(Z(m)) - \zeta \mathbb{E} \{U_{1,m} + U_{2,m} | Z(m)\} \\ = B - Z(m)\mathbb{E} \{c_{av} - C(m) | Z(m)\} \\ - \zeta \mathbb{E} \{U_{1,m} + U_{2,m} | Z(m)\} \end{aligned} \quad (67)$$

where  $\zeta$  is a non-negative parameter that weighs the extent to which we emphasize utility maximization. The algorithm aims to minimize Lyapunov drift thereby maximizing the following expression at each task  $m$ :

$$\mathbb{E} \{ \zeta U_{1,m} + \zeta U_{2,m} - Z(m)C(m) | Z(m) \}. \quad (68)$$

Maximization of (68) involves the dynamic program studied in previous sections for each task  $m$ . We refer to [24] for a detailed treatment of the class of Lyapunov optimization algorithms tailored for renewal systems.

In particular, the previous formulation which maximizes the difference between QoI utility minus cost corresponds to the case where the virtual queue size is unity for all tasks. The main effect of the virtual queues is to transform the unit processing cost  $\beta$  to  $Z(m)\beta$  for task  $m$ . As a result, a dynamic program is formulated for each task and solved with  $Z(m)\beta$  instead of  $\beta$ .

If the cost expended in previous tasks are high, the virtual queue grows and results in a high multiplier, i.e.,  $Z(m)$ . This, in turn increases the penalty for continuing, and results in a policy which tends to stop earlier. This can be also visualized by considering the fact that increasing  $\beta$  reduces expressions that are being compared with  $U_1 + U_2$  in (6). On the other hand, if cost consumptions at previous tasks are low, the virtual queue size is small, resulting in a reduced effect on cost for the current task, leading to an increased threshold for transmission.

Note that in addition to specifying the exact policy by solving the dynamic program for each task, we also perform the queue update by observing the resulting  $C(m)$  values, and update virtual queues  $Z(m)$  by (63).

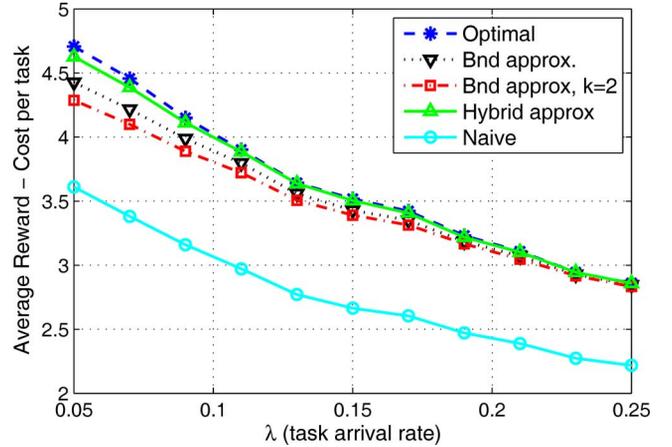


Fig. 8. QoI utility-minus-cost for geometric task arrivals,  $\gamma_1 = 0.8$ ,  $\gamma_2 = 0.5$ .

## VIII. NUMERICAL RESULTS

In this section, we present numerical results which demonstrate performance of the adaptive sampling algorithms discussed. Throughout majority of the simulations, we compare the derived algorithms with an algorithm which stops every slot with a fixed probability, which is equal to the reciprocal of the expected task period. In the special case when the new task arrives before any reporting decision, the adaptive sampling algorithm forces to send the information. We call this algorithm *naive*.

First, in Fig. 8, performance of algorithms for the case with Geometric task arrivals is investigated in terms of *average QoI utility(reward) minus cost per task* by varying  $\lambda$ . While the specific sensor network application and deployment scenario would affect parameters associated with utility, costs and amount of information gathered, the parameters used in this simulation are  $\gamma_1 = 0.8$ ,  $\gamma_2 = 0.5$ , and  $\beta = 0.01$ . Information arrivals every slot are Bernoulli with mean values  $\overline{Q_1} = 0.9$ , and  $\overline{Q_2} = 0.4$ . It is readily seen that the algorithm based on bounding functions performs closely to the optimal algorithm and significantly outperforms the naive algorithm, which does not use any explicit information about the current QoI state of the network. Moreover, it is observed that a low number of value iterates performs fairly well. Finally, we also observe that the hybrid algorithm reduces the gap with optimality with a moderate complexity. Specifically, to provide insight on the complexity of these different algorithms, the curves in Fig. 8 require computational time of 3.12 sec for naive, 4.62 sec for bounding function approximation with  $k = 2$ , 20.47 sec for bounding function, and 22.39 sec for the hybrid algorithm.

Next, in Fig. 9, we present performance of the case with periodic task arrivals. Specifically, the task period is varied for  $\gamma_1 = 0.8$ ,  $\gamma_2 = 0.5$ , Bernoulli  $Q_1, Q_2$  with  $\overline{Q_1} = 0.9$  and  $\overline{Q_2} = 0.4$ , and  $\beta = 0.01$ . We again observe that the approximate adaptive sampling algorithm performs closely to the optimal adaptive sampling algorithm and significantly outperforms the naive algorithm. Moreover, as task period increases, it is also observed that the maximum QoI utility per task also increases. This is attributed to the fact that there are more opportunities to increase the utility before the new task forces the sensors

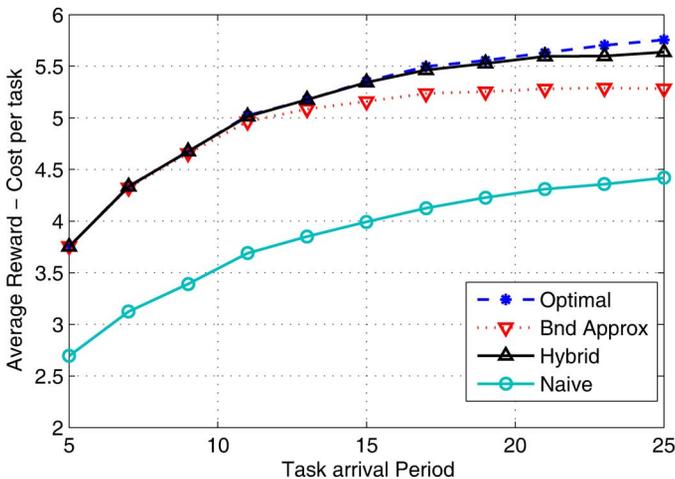


Fig. 9. QoI utility-minus-cost for periodic task arrivals,  $\gamma_1 = 0.8$ ,  $\gamma_2 = 0.5$ .

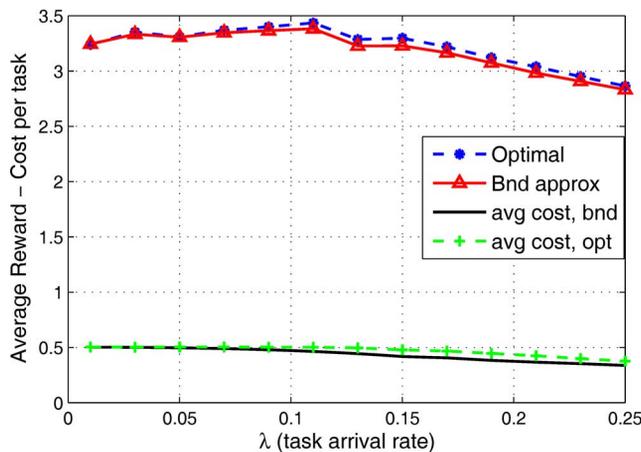


Fig. 10. Performance for the average cost constrained case,  $c_{av} = 0.5$ .

to respond. Another outcome is that as discussed in Remark 2, at the initial stages, the first condition in (51) is sufficient for the algorithm to make the continue decision for strong majority of the tasks. As stage number increases, additional continuing conditions might be needed more frequently, but the maximum number of conditions to be checked are also reduced.

Finally, we demonstrate results for the case with constrained average costs in Fig. 10. We limit the average cost constraint per cost as  $c_{av} = 0.5$ , with  $\beta = 0.1$ , and vary  $\lambda$ . It is observed that the modified adaptive sampling algorithm based on Lyapunov optimization indeed keeps the average cost per task below 0.5 for all cases. We also note that the constraint is not always tight, particularly as  $\lambda$  increases. Increased task arrival probability reduces the number of stages for each task, and consequently the cost expended per task.

## IX. CONCLUSIONS

In this paper, we have considered QoI-aware adaptive sampling for a network of multiple sensor devices. We have focused on the problem of maximizing QoI utility-minus-cost for the scenario when information arrives randomly to sensors.

We have provided various low complexity near-optimal adaptive sampling algorithms which decide whether the sensors should respond or continue accumulating information for different task arrival patterns. These adaptive sampling

decisions depend on the joint QoI state at the network, and decisions are made by checking whether the joint QoI states are in the continuing regions provided. The approximate adaptive sampling algorithms are demonstrated to approach optimal QoI-utility-minus cost for both geometric and periodic task arrival patterns. We have provided algorithms with different levels of complexity along with their closeness in performance to optimal algorithms. We have also proposed adaptive sampling algorithms which maximize QoI utility when the average operational cost per task is constrained. Leveraging tools from stochastic network optimization, we have developed algorithms which successfully balance exploration and exploitation for adaptive sampling with long-term cost constraints.

While the methodology provided in this paper can be readily extended to a multiple sensor network with more than two users, future work includes maximizing the sum QoI utility delivered in multi-sink environments, and other multi-user topologies. Consideration of arbitrary functions of the source utilities as the final QoI utility at the end user is also of future interest.

## REFERENCES

- [1] S. Zahedi and C. Bisdikian, "A Framework for QoI-inspired analysis for sensor network deployment planning," in *Proc. Int. Workshop Performance Control in Wireless Sensor Netw. (PWSN)*, Austin, TX, USA, Oct. 2007.
- [2] A. Bar-Noy, G. Cirincione, R. Govindan, S. Krishnamurthy, T. F. La-Porta, P. Mohapatra, M. Neely, and A. Yener, "Quality-of-information aware networking for tactical military networks," in *Proc. 3rd Int. Workshop Inf. Quality and Quality of Service for Pervasive Comput., in Conjunction with IEEE Percom '11*, Seattle, WA, USA, Mar. 2011, pp. 2–7.
- [3] B. Liu, P. Terlecky, A. Bar-Noy, R. Govindan, and M. J. Neely, "Optimizing information credibility in social swarming applications," in *Proc. IEEE INFOCOM '11 Mini-Conf.*, Shanghai, China, Apr. 2011, pp. 1147–1158.
- [4] Z. M. Charbiwala, S. Zahedi, Y. Kim, Y. H. Cho, and M. B. Srivastava, "Toward quality of information aware rate control for sensor networks," in *Proc. 4th Int. Workshop Feedback Control Implement. Design in Comput. Syst. Netw.*, San Francisco, CA, USA, Apr. 2009.
- [5] E. N. Ciftcioglu, A. Yener, R. Govindan, and K. Psounis, "Operational information content sum capacity: Formulation and examples," in *Proc. 14th Conf. Inf. Fusion*, Chicago, IL, USA, Jul. 2011.
- [6] E. N. Ciftcioglu and A. Yener, "Quality-of-information aware transmission policies with time-varying links," in *Proc. MILCOM '11*, Baltimore, MD, USA, Nov. 2011.
- [7] R. Uргаonkar, E. N. Ciftcioglu, A. Yener, and M. J. Neely, "Quality of information aware scheduling in task processing networks," in *Proc. 7th Int. Workshop Resource Alloc. Cooperat. Wireless Netw. (RAWNET), in Conjunction with IEEE WiOpt '11*, Princeton, NJ, USA, May 2011, pp. 401–406.
- [8] S. Gao, L. Qian, D. R. Vaman, and Z. Han, "Distributed cognitive sensing for time varying channels: Exploration and exploitation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC'10)*, Sydney, Australia, Apr. 2010, pp. 1–6.
- [9] L. Lai, H. El Gamal, H. Jiang, and V. V. Poor, "Cognitive medium access: Exploration, exploitation, and competition," *IEEE Trans. Mobile Comput.*, vol. 10, no. 2, pp. 239–253, Feb. 2011.
- [10] K. Shah and M. Kumar, "Distributed independent reinforcement learning (dir) approach to resource management in wireless sensor networks," in *Proc. IEEE Int. Conf. Mobile Adhoc and Sensor Syst. (MASS '07)*, Pisa, Italy, Oct. 2007, pp. 1–9.
- [11] Z. Liu and I. Elhanany, "RL-MAC: A QoS-aware reinforcement learning based MAC protocol for wireless sensor networks," in *Proc. IEEE Int. Conf. Network., Sens., Control (ICNSC'06)*, pp. 768–773.
- [12] J. Hardwick and Q. F. Stout, "Optimal few-stage designs," *J. Statist. Planning Inference*, vol. 104, pp. 121–145, 2002.
- [13] H. S. Chang, M. C. Fu, and R. H. Smith, "An adaptive sampling algorithm for solving Markov decision processes," *Operat. Res.*, pp. 126–139, Jan.-Feb. 2005.

- [14] D. P. Bertsekas, "A new value iteration method for the average cost dynamic programming problem," *SIAM J. Control Optimiz.*, vol. 36, no. 2, pp. 742–759, 1998.
- [15] M. Riedmiller, "Neural fitted Q iteration – First experiences with a data efficient neural reinforcement learning method," in *Proc. Mach. Learn.: ECML*, Porto, Portugal, Oct. 2005.
- [16] J. Schneider, W.-K. Wong, A. Moore, and M. Riedmiller, "Distributed value functions," in *Proc. Int. Conf. Mach. Learn.*, Bled, Slovenia, Jun. 1999.
- [17] E. D. Ferreira and P. K. Khosla, "Multi agent collaboration using distributed value functions," in *Proc. IEEE Intel. Veh. Symp., IV '00*, Dearborn, MI, USA, Oct. 2000, pp. 404–409.
- [18] L. Georgiadis, M. J. Neely, and L. Tassiulas, *Resource Allocation and Cross-layer Control in Wireless Networks*. Delft, The Netherlands: Now, 2006, Foundations and Trends in Networking.
- [19] J. A. Stankovic, T. E. Abdelzaher, C. Lu, L. Sha, and J. C. Hou, "Real-time communication and coordination in embedded sensor networks," *Proc. IEEE*, vol. 91, no. 7, pp. 1002–1022, Jul. 2003.
- [20] V. Lesser, C. C. L. Ortiz, and M. Tambe, *Distributed Sensor Networks: A Multiagent Perspective*. New York, NY, USA: Springer, 2003, vol. 9.
- [21] D. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont, MA, USA: Athena Scientific, 2001, vol. 1.
- [22] J. B. Rawlings and K. R. Muske, "The stability of constrained receding horizon control," *IEEE Trans. Autom. Control*, vol. 38, no. 10, pp. 1512–1516, Oct. 1993.
- [23] M. J. Neely, "Energy optimal control for time varying wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2915–2934, Jul. 2006.
- [24] M. J. Neely, "Dynamic optimization and learning for renewal systems," in *Proc. Asilomar Conf. Signals, Syst., Comput. (Invited Paper)*, Nov. 2010.



**Ertugrul Necdet Ciftcioglu** (S'06–M'13) received his B.S. degree in Electrical and Electronics Engineering from the Middle East Technical University (METU), Ankara, Turkey in 2004, M.S. degree in Electronics Engineering and Computer Science from Sabanci University, Istanbul, Turkey in 2006, and Ph.D. degree in Electrical Engineering from The Pennsylvania State University, PA in 2012. He is currently a Research Associate in the Department of Computer Science and Engineering, The Pennsylvania State University. He has been a visiting

scholar at Northwestern University, IL, University of Southern California, CA and BBN Technologies, MA. His research interests are cross-layer design and resource allocation for wireless communication networks, particularly stochastic network optimization for relaying, cooperative communications and multiuser networks, and recent emphasis on network science.



**Aylin Yener** (S'91–M'00) received the B.Sc. degree in electrical and electronics engineering, and the B.Sc. degree in physics, from Boğaziçi University, Istanbul, Turkey; and the M.S. and Ph.D. degrees in electrical and computer engineering from Wireless Information Network Laboratory (WINLAB), Rutgers University, New Brunswick, NJ. Commencing fall 2000, for three semesters, she was a P. C. Rossin Assistant Professor at the Electrical Engineering and Computer Science Department, Lehigh University, PA. In 2002, she joined the faculty of The Pennsylvania State University, University Park, PA, where she was an Assistant Professor, then Associate Professor, and is currently Professor of Electrical Engineering since 2010. During the academic year 2008–2009, she was a Visiting Associate Professor with the Department of Electrical Engineering, Stanford University, CA. Her research interests are in information theory, communication theory and network science, with recent emphasis on green communications and information security. She received the NSF CAREER award in 2003.

Dr. Yener previously served as a technical program chair or co-chair for various conferences for the IEEE Communications Society, as an associate editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, as an associate editor and an editorial advisory board member for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. She served as the student committee chair for the IEEE Information Theory Society 2007–2011, and was the co-founder of the Annual School of Information Theory in North America co-organizing the school in 2008, 2009 and 2010. Dr. Yener currently serves on the board of governors of the IEEE Information Theory Society as its treasurer.



**Michael J. Neely** (SM'09) received B.S. degrees in both Electrical Engineering and Mathematics from the University of Maryland, College Park, in 1997. He was then awarded a 3 year Department of Defense NDSEG Fellowship for graduate study at the Massachusetts Institute of Technology, where he received an M.S. degree in 1999 and a Ph.D. in 2003, both in Electrical Engineering. He joined the faculty of Electrical Engineering at the University of Southern California in 2004, where he is currently an Associate Professor. His research interests are in the areas

of stochastic network optimization and queueing theory, with applications to wireless networks, mobile ad-hoc networks, and switching systems. Michael received the NSF Career award in 2008, the Viterbi School of Engineering Junior Research Award in 2009, and the Okawa Foundation research award in 2012. He is a member of Tau Beta Pi and Phi Beta Kappa.