MAXIMUM LIFETIME TRANSMISSION IN WIRELESS SENSOR NETWORKS FOR A COMMON SOURCE OBSERVATION

Xiaolu Zhang¹, Meixia Tao² and Chun Sum Ng¹

¹Dept. of Electrical & Computer Engineering, National University of Singapore, Singapore 117576

²Dept. of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China 200240

Emails: zhangxiaolu@nus.edu.sg; mxtao@sjtu.edu.cn; elengcs@nus.edu.sg

ABSTRACT

Energy efficient transmission is of vital importance in the design of wireless sensor networks (WSN). Minimizing the sum of energy consumption of all sensor nodes in the network may lead to some nodes with non-rechargeable batteries running out of energy quickly. In this paper, we consider to maximize the time until the first node fails in the network, namely, the network lifetime. Transmission schemes in a WSN where multiple sensors observe a common physical phenomenon are investigated. We show that the optimal joint time and power control (JTPC) takes advantage of “lazy scheduling” and “opportunistic quantization”. We then demonstrate that the JTJC can be implemented in a partially distributed manner. Numerical results show that significant lifetime gain can be achieved when compared with existing scheduling algorithms.

1. INTRODUCTION

Wireless sensor networks (WSN) have enormous potential for situation awareness applications. A WSN usually consists of a set of sensor nodes deployed in a certain region and performs distributed detection or estimation of application-specific information. In the context of distributed data collection, each node monitors its surrounding area, collects the data, and transmits it to a fusion center (FC). The FC then makes the final estimation using the collected data. A typical low cost sensor node has only limited processing capacity and often powered by small batteries. In many situations, battery recharging is impossible. If one or more nodes fail due to lack of energy, the sensor network may not sustain normal functionality. The lifetime of the network thus often refers to the time it takes for the first node in the network to die. Prolonging the network lifetime while maintaining a reasonably low computational cost has become the major challenge in designing communication protocols for WSN.

Energy efficient transmission strategies for single-hop data collection have recently attracted attention. In fundamental communication theory, the transmission rate is an increasing and concave function of the transmission power, and the total amount of energy required for transmitting a given amount of

data bits can be reduced by lowering the transmission power. This energy saving is achieved at the expense of increased transmission delay. Making use of this delay-energy tradeoff, Yao *et al.* proposed an energy-efficient transmission scheduling in [1]. Given that each sensor node has a fixed number of bits to transmit, the total energy consumption is minimized by varying the transmission times assigned to different sensors. Due to limited channel bandwidth, quantization needs to be done for the observed data by each sensor before transmission. An optimal power control policy for such decentralized estimation is proposed by Xiao *et al.* in [2]. This policy minimizes the energy consumption by varying the transmission power and quantization level. It is suggested that the sensors with better channel condition and/or good observation quality should increase their quantization resolution. It is however important to point out that maximizing energy efficiency does not necessarily lead to network lifetime maximization. In both [1] and [2], minimizing the total energy consumption or some variants of it may result in some nodes running out of energy quickly.

In this paper we propose to maximize directly the network lifetime by varying both transmission time and power for a single-hop sensor network. Several fundamental features of sensor networks are taken into account in our formulation. First, for a sensor network deployed for decentralized estimation, the estimation accuracy achieved at the FC is usually more important than the total or individual transmission rate. Secondly, since the energy limitation of each node may differ from each other, our lifetime maximization strategy also considers the available energy left on each sensor. Lastly, since FC usually cannot afford high computational cost and high control signaling overhead, our algorithm is designed to be of partially distributed nature.

2. SYSTEM MODEL

Consider a WSN with an FC and a set of K distributed sensor nodes, denoted as $\mathcal{K} = \{1, 2, \dots, K\}$. All sensors are located densely and observe only one common physical phenomenon θ . The observation by sensor k corrupted by additive noise is

denoted by

$$x_k = \theta + \omega_k, \quad k \in \mathcal{K}, \quad (1)$$

where ω_k is a zero-mean random variable with variance σ_k^2 and is independent for different k 's. Each observation x_k is quantized locally by each sensor using L_k bits. The resulting discrete message $m_k(x_k, L_k)$ is then transmitted to the FC over a common channel shared by all the sensors in the cluster. After collecting the data from all sensors, the FC generates the estimate $\hat{\theta}$. This decentralized estimation procedure with a common source observation is illustrated in Fig. 1.

We define the interval between two consecutive data reports as a time frame. Without loss of generality, let the duration of each time frame be one unit. Each sensor node takes turn to forward messages and occupies ρ_k fraction of time in each time frame. The channel gain between each sensor node and FC is assumed to be Gaussian. The channel gains are denoted as g_k , for $k \in \mathcal{K}$, and change slowly so as to remain constant during the network lifetime. Assume capacity-achieving coding is applied to each link. Thus, the number of quantization bits on each sensor node is given by:

$$L_k = \left\lceil \rho_k \log_2 \left(1 + \frac{g_k p_k}{N_0} \right) \right\rceil, \quad \forall k \in \mathcal{K}, \quad (2)$$

where $\lfloor a \rfloor$ represents the largest integer that is not greater than a . When $a \gg 1$, $\lfloor a \rfloor \approx a$. N_0 represents the noise power, and $\mathbf{p} = \{p_k, \forall k \in \mathcal{K}\}$ is the transmission power vector. Thus the energy consumed by sensor k in a time frame is $\rho_k p_k$. Without loss of generality, we let $N_0 = 1$. It is assumed that g_k is known at sensor node k . It is also assumed that each sensor is aware of its own available energy E_k and observation quality measured by σ_k^2 . Our resource allocation strategy aims to find the optimal ρ and $\mathbf{p}$ under the knowledge of σ_k^2 , g_k and E_k such that the network lifetime is maximized.

Some efficient local quantization schemes and distributed estimators have been discussed in [2, 3]. If the probability density function (PDF) of observation noise is unknown, one can use uniform quantizer. According to the result in [2], the estimate of θ has a mean square error (MSE)

$$D \leq \left(\sum_{k=1}^K \frac{1}{\sigma_k^2 + \delta_k^2} \right)^{-1}, \quad (3)$$

where δ_k^2 is the quantization noise variance given by

$$\delta_k^2 = \frac{W^2}{(2^{L_k} - 1)^2}. \quad (4)$$

In (4), $[-W, W]$ is the signal range that sensors can observe.

3. JOINT TIME SHARING AND POWER CONTROL (JTPC)

In this section, we will discuss joint time sharing and power control. The goal is to prolong the minimum lifetime among

Fig. 1. Decentralized estimation procedure

all the nodes under an estimation accuracy constraint. When MSE is chosen as the performance metric of estimation accuracy, the problem of network lifetime maximization can be formulated as a constrained optimization problem, **P1**, denoted by

$$\mathbf{P1} : \max_{\{\rho, \mathbf{p}\}} \min_k T_k \quad (5)$$

$$\text{s.t. } D(\mathbf{L}) \leq D_0 \quad (6)$$

$$\begin{aligned} T_k \rho_k p_k &= E_k, \quad k \in \mathcal{K} \\ \sum_{k=1}^K \rho_k &= 1 \end{aligned} \quad (7)$$

where T_k denotes the lifetime of sensor node k , as defined in (7), with E_k being the total available energy and $\rho_k p_k$ being the energy consumed in each time frame. $\mathbf{L}$ is defined as $\mathbf{L} := \{L_k, \forall k \in \mathcal{K}\}$ (see (2)). D_0 is the target average MSE constraint. If we define

$$U_k(L_k) := \frac{1}{W^2 (2^{L_k} - 1)^{-2} + \sigma_k^2}, \quad (8)$$

then, we have $D(\mathbf{L}) \leq 1/(\sum_{k=1}^K U_k(L_k))$ from (3). By taking the first and second derivatives of $U_k(L_k)$, it can be shown that U_k is an increasing and concave function of L_k when the observation noise variance σ_k^2 is greater than quantization noise variance δ_k^2 . The reason that we exclude the case of high observation quality is that when the sensor node has good observation quality (small σ_k^2), multiple copies of quantized observation cannot bring significant benefit to the system MSE. Meanwhile, the energy wastage brought by the transmission of multiple copies cannot be neglected. This observation motivates the focus on sensor networks with relatively poor observation quality. According to the property of convexity preservation [4], the function D is convex in L_k since D is a decreasing and convex function of U_k .

Definition: An average MSE target D_0 is said to be a *feasible MSE requirement* if there exists $\mathbf{p}$ and ρ such that D_0 can be achieved. Any feasible MSE requirement D_0 must satisfy

$$D_0 > D_{\min} = 1 / \left(\sum_{k=1}^K \frac{1}{\sigma_k^2} \right), \quad (9)$$

where $D_{\min}$ is the smallest MSE in the absence of quantization error. When $\sigma_k^2 = \sigma^2$ ($\forall k \in \mathcal{K}$), we have $D_0 > \sigma^2/K$. This describes how the number of sensor nodes enhances the estimation accuracy.

3.1. Some Properties of The Optimal Solution

For notation brevity, we define a new variable vector $\mathbf{s}$ to denote energy consumption in a time frame, the k^{th} element of which is $s_k := p_k \rho_k$. Due to the convexity of the optimization problem **P1**, it follows easily the first property of the optimal solution.

Property 1. *The optimal solution $s_k^* = p_k^* \rho_k^*$ to the problem **P1** satisfies that $s_k^*/E_k = x^*$ ($\forall k \in \mathcal{K}$), where $x^* = \max_k s_k^*/E_k$. In other words, the normalized energy consumptions in one time frame for all users are identical.*

Since the normalized energy consumption in a time frame is equivalent to the inverse of lifetime, *Property 1* indicates that all sensors in the network should fail simultaneously.

According to *Property 1*, we have $x = 1/T_k$ ($\forall k \in \mathcal{K}$). After substituting $p_k = E_k x / \rho_k$ into (7), **P1** can be transformed into a new problem, **P2**, with optimization variables ρ and x .

$$\mathbf{P2}: \min_{\{\rho, x\}} x \quad (10)$$

$$\text{s.t.} \quad \sum_{k=1}^K U_k(\rho_k, x) \geq \frac{1}{D_0} \quad (11)$$

$$\sum_{k=1}^K \rho_k = 1 \quad (12)$$

Here, $U_k(L_k)$ in (8) is re-expressed as $U_k(\rho_k, x)$ since L_k is a function of ρ_k and x . The new problem **P2** is also convex. Using the Karush-Kuhn-Tucker (KKT) optimality conditions we obtain the following two properties, which describe the relation between the optimal resource allocation and channel condition, initial power and observation quality.

Property 2. *When all the nodes have an identical observation quality, i.e., $\sigma_k^2 = \sigma_0^2$ ($\forall k \in \mathcal{K}$), the optimal time sharing fraction only relates to the product of initial energy and channel gain. More time sharing fraction should be given to the user with smaller product of initial energy and channel gain.*

Proof. When $\sigma_k^2 = \sigma^2$ ($\forall k \in \mathcal{K}$), substituting $p_k = E_k x / \rho_k$ into L_k in (2), and substituting the resulting L_k into (8), we have

$$U_k(L_k) = \frac{1}{W^2 \left[\left(1 + \frac{g_k E_k x}{\rho_k N_0} \right)^{\rho_k} - 1 \right]^2 + \sigma^2}.$$

Thus, $U_k(L_k)$ can be treated as a function of ρ_k , $E_k g_k$ and x . Here, the product of E_k and g_k is viewed as a single

variable. Then, we can define a function $U(\rho, E g, x)$ with $U(\rho_k, E_k g_k, x) = U_k(L_k)$. Let the derivative of $U(\rho, E g, x)$ with respect to ρ be denoted by $f(\rho, E g, x) = \partial U(\rho, E g, x) / \partial \rho$. It can be shown easily that the function $f(\rho, E g, x)$ is monotonically decreasing in both ρ and $E g$. According to KKT conditions, we have $f(\rho_i^*, E_i g_i, x^*) = f(\rho_k^*, E_k g_k, x^*)$, for all i, k . Therefore, to maintain this equality, one must have $\rho_i > \rho_k$ ($\rho_i < \rho_k$) if $E_i g_i < E_k g_k$ ($E_i g_i > E_k g_k$). $\square$

Property 2 suggests that the sensors with bad channel and/or energy condition will take longer time for transmission. This is in agreement with the idea of “lazy scheduling” [5], which achieves energy saving by lowering the transmission power and hence taking longer transmission time.

Property 3. *When all the nodes have an identical product of E_k and g_k , the optimal time sharing fraction only relates to the observation noise variation σ_k^2 . More time sharing fraction should be given to the sensor node with better observation quality. Correspondingly, the sensor node with better observation quality should use higher quantization resolution.*

Proof. A similar approach used in the proof of *Property 2* applies here. The details are omitted. $\square$

Property 3 indicates that for prolonging the network lifetime, the optimal solution takes advantage of “opportunistic quantization”, that is, the sensor nodes with high observation qualities should increase their quantization resolution.

3.2. Partially Distributed Adaptation

It is observed from the problem **P2** that if the optimal x^* is given, the optimal time sharing policy ρ^* obviously satisfies

$$\rho^* = \arg \max_{\substack{\sum_k \rho_k = 1, \\ 0 \leq \rho_k \leq 1}} \sum_{k=1}^K U_k(\rho_k, x^*). \quad (13)$$

Therefore, we propose to decompose **P2** into two subproblems: (a) binary search for x^* ; (b) partially distributed implementation of time-sharing adaptation using gradient projection method. Subproblem (b) is addressed in detail as follows.

For a given x^* , define the Lagrangian

$$\begin{aligned} L(\rho, \lambda) &= \sum_{k=1}^K U_k(\rho_k, x^*) - \lambda \left(\sum_{k=1}^K \rho_k - 1 \right) \\ &= \sum_{k=1}^K [U_k(\rho_k, x^*) - \lambda \rho_k] + \lambda. \end{aligned}$$

The dual function of (13) is

$$F(\lambda) = \max_{\rho} L(\rho, \lambda) = \sum_{k=1}^K B_k(\lambda) + \lambda, \quad (14)$$

where

$$B_k(\lambda) = \max_{0 \leq \rho_k \leq 1} U_k(\rho_k, x^*) - \lambda \rho_k. \quad (15)$$

If we interpret λ as the price of time sharing fraction, $B_k(\lambda)$ represents the maximum benefit of node k at the given price λ . The K independent subproblems in (15) can be solved at each node individually, and the resulting optimal $\rho_k(\lambda)$ is obtained as

$$\rho_k(\lambda) = [(U'_k)^{-1}(\lambda, x^*)]_0^1, \quad (16)$$

where $[\cdot]_a^b = \min\{\max\{\cdot, a\}, b\}$, and $(U'_k)^{-1}(\cdot, x^*)$ is the inverse of $\partial U_k / \partial \rho_k$ with respect to ρ_k which exists since U_k is twice differentiable and strictly concave.

Minimizing the dual function (14) can now be solved by adjusting the price λ using gradient projection. That is,

$$\lambda(t+1) = \left[\lambda(t) - \gamma \left(\sum_{k=1}^K \rho_k - 1 \right) \right]^+. \quad (17)$$

Here, $\gamma > 0$ is a step size, and $[\cdot]^+ := \max\{\cdot, 0\}$.

The time sharing adaptation and price adjustment have the decentralized nature. To implement the update, the FC needs to know the sum of ρ_k 's, whereas each node k only needs to know its own function U_k and the price λ .

Theorem 1. *There exists a constant γ_0 such that for any step-size γ satisfying $0 < \gamma < \gamma_0$, and any initial time sharing fractions $0 \leq \rho(0) \leq 1$ and price $\lambda(0)$, the sequence $\{\rho(t), \lambda(t)\}$ generated by (17) and $\rho_k(t+1) = \rho_k(\lambda(t))$ (see (16)) converges to the unique optimal point (ρ^*, λ^*) .*

Theorem 1 holds since the dual objective function F is convex, lower bounded and Lipschitz continuous (see [6, Prop. 3.4]).

The implementation of binary search can also be realized at the node side only using 1-bit control signal from fusion center since each node only needs to know whether $\sum_k U_k(\rho_k, x) > 1/D_0$ or not.

4. NUMERICAL RESULTS

The simulation is carried out under various network and channel configurations by varying path loss, available energy, and observation noise variance. The channel gain is modelled as $g = \alpha d^{-n}$, where d^{-n} is the effect of path loss with n being the path loss exponent, and α is an attenuation constant. The distance between sensor node k and the FC, d_k , is randomly generated following a uniform distribution in the region $[1, 10]$. Similar to [7], observation noise variances σ_k^2 ($\forall k \in \mathcal{K}$) are generated according to the distribution $0.05(1 + aY^2)$, where Y^2 follows the exponential distribution. In addition, we set available energies $E_k = E^{\min}(1 + bY^2)$, where $E^{\min}$ is the minimum energy of all nodes. Through adjusting the parameters n , a , and b , we generate different sets of $\{g_k, E_k, \sigma_k^2\}$

Fig. 2. The network Lifetimes of JTPC, UTP and PS vs. the number of sensor nodes

to model different sensing environments. We use the normalized deviation $\sqrt{\mathbb{V}(t)}/\mathbb{E}(t)$ defined in [7] to measure the heterogeneity of a variable t , where $\mathbb{V}(t)$ denotes the variance of t and $\mathbb{E}(t)$ the mean of t .

To illustrate the performance of JTPC, two benchmark schemes are considered:

1. *Uniform TDMA with power control (UTP)*: It is a sub-optimal solution of the proposed JTPC by letting $\rho_k = 1/K, \forall k \in \mathcal{K}$. The best power allocation $\mathbf{p}^*$ is obtained from *Property 1* such that all nodes die simultaneously.
2. *Power scheduling (PS)* [2]: all nodes also have a same time sharing fraction, but the power is adjusted to minimize the L^2 -norm of transmission power vector.

Fig. 2 compares the network lifetimes in the number of time frames achieved by JTPC, UTP and PS at different number of sensor nodes. We let $n = 3$, $a = 1$, and $b = 0$. The MSE requirement D_0 is fixed to be 0.01. It is observed that the network lifetimes of JTPC increases exponentially with the number of sensor nodes. Since the transmission time of each user, $\rho = 1/K$, in UTP and PS, decreases linearly with the number of sensor nodes, increasing the number of sensor nodes cannot prolong the lifetime of the network. The performance of PS is worse than that of UTP since PS suggests that the nodes with good channels and/or observation qualities should increase their quantization resolutions which leads them to run out of energy quickly.

In the remaining simulations, we study the lifetime gains (defined as the ratio of lifetime) of JTPC over UTP and PS under different sensing environments. The MSE requirement D_0 is fixed to be $2D_{\min}$, where $D_{\min}$ is the MSE of estimation without quantization noise given in (9).

Fig. 3 plots the lifetime of JTPC over UTP and PS at different heterogeneity of channel path loss. Here we fix σ_k^2 and E_k by choosing $a = b = 0$. The normalized deviation of

channel path losses is varied by adjusting n ($2 \leq n \leq 4$). Fig. 4 shows the results by fixing $d_k = 5.5, \forall k \in \mathcal{K}, b = 0$ and varying a . In Fig. 4, the lifetime gain of JTPC over UTP and PS increases when observation noise variances become more heterogeneous. In addition, for fixed normalized deviation noise variances, the lifetime gain increases as the number of nodes increases. The relation between the lifetime gain and the heterogeneity of initial energies is shown in Fig. 5.

5. CONCLUSION

We considered transmission scheduling and power control to prolong the lifetime of WSNs under the estimation accuracy constraint. We focused on the scenario where all sensors observe only one common source. A partially distributed algorithm is developed where the transmission time, power and quantization levels can be computed at each node side using its own local information (channel condition, available energy and observation quality) and little common information broadcast from fusion center. Simulation results show that significant lifetime gain can be achieved when compared with the existing energy efficient schemes. In specific, the gain increases as the network becomes more heterogenous. Note that the definition of network lifetime considered in this work is a simple one. In a real WSN, the death of some nodes may not affect the functioning of the whole network. More useful definitions of network lifetime, for example, the time when the L^{th} ($L < K$) node dies, are needed.

6. REFERENCES

- [1] Y. Yao and G. B. Giannakis, "Energy-efficient scheduling for wireless sensor networks," *IEEE Trans. on Commun.*, vol. 53, no. 8, pp. 1333–1342, Aug. 2005.
- [2] J. Xiao, S. Cui, Z. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 413–422, Feb. 2006.
- [3] D. Blatt and A. Hero, "Distributed maximum likelihood for sensor networks," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, 2004.
- [4] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, United Kingdom: Cambridge Univ. Press, 2004.
- [5] A. E. Gamal, C. Nair, B. Prabhakar, E. Uysal-Biyikoglu, and S. Zahedi, "Energy-efficient scheduling of packet transmissions over wireless networks," in *IEEE INFOCOM '02*, 2002.
- [6] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.
- [7] J. Xiao, A. Ribeiro, G. B. Giannakis, and Z. Luo, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 27–41, July 2006.

Fig. 3. Lifetime gains of JTPC over UTP and PS vs. normalized deviation of channel path losses

Fig. 4. Lifetime gains of JTPC over UTP and PS vs. normalized deviation of observation noise variances

Fig. 5. Lifetime gains of JTPC over UTP and PS vs. normalized deviation of initial energies