

# DECENTRALIZED VS. CENTRALIZED SCHEDULING IN WIRELESS SENSOR NETWORKS FOR DATA FUSION

Mihaela Mitici\* Jasper Goseling\*<sup>†</sup> Maurits de Graaf\*<sup>‡</sup> Richard J. Boucherie\*

\* Stochastic Operations Research, University of Twente, The Netherlands

<sup>†</sup>Department of Intelligent Systems, Delft University of Technology, The Netherlands

<sup>‡</sup>Thales Nederland B.V., The Netherlands

{m.a.mitici, j.goseling, m.degraaf, r.j.boucherie}@utwente.nl

## ABSTRACT

We consider the problem of data estimation in a sensor wireless network where sensors transmit their observations according to decentralized and centralized transmission schedules. A data collector is interested in achieving a data estimation using several sensor observations such that the variance of the estimation is below a targeted threshold. We analyze the waiting time for a collector to receive sufficient sensor observations. We show that, for sufficiently large sensor sets, the decentralized schedule results in a waiting time that is a constant factor approximation of the waiting time under the optimal centralized scheme.

**Index Terms**— Wireless sensor network, data fusion, scheduling, waiting time

## 1. INTRODUCTION

We consider a network of wireless sensors that have i.i.d. observations of an attribute, e.g. temperature. A collector is interested in estimating this attribute by combining a subset of available observations such that the accuracy of the estimate is below a targeted threshold. Since the sensor observations are independent and identically distributed, any sufficiently large subset of observations can achieve the required accuracy. In this work we analyze scheduling mechanisms for sensors that allow the collector to obtain these sensor observations.

The problem of sensor data fusion has been extensively studied in [1, 2, 3, 4]. Data fusion techniques combine data from several sensors to improve data accuracy, which is difficult to achieve interrogating a single sensor alone. Estimation of a variable using a set of sensor nodes and a fusion center has been studied in [5, 6, 7]. Generally, the sensor observations are transmitted to a fusion center where a final estimate is determined by performing a linear combination of the sensor observations, a technique referred to as the centralized BLUE [8]. In [5, 7], the authors study the problem of energy minimization while keeping the mean square estimation error of the sensor observations below a targeted threshold. In

[6], the estimation of a noise-corrupted parameter given bandwidth constraints is considered. Complementary to the work mentioned above, our focus is on scheduling mechanisms to support data fusion, when the sensors are contending for the medium.

Scheduling transmission techniques for wireless sensor networks have been studied extensively [9, 10]. Most of the work considered scheduling independently from the data fusion task. In contrast, in [11] a scheduling method to maximize the lifespan of a wireless sensor network is investigated. The authors provide an algorithm to partition the sensors in adjacent sets and schedule each sets for transmission such that the observations collected by the active sensors provide an accurate estimate of an attribute. In [12], the problem of data fusion is considered, where sensors transmit their observations according to a slotted ALOHA protocol (see [13] for an introduction to the ALOHA protocol). The authors show that the fusion rule is a weighted sum of the received messages and their collisions.

In this paper, we consider a decentralized scheduling, i.e. similar to slotted ALOHA studied in [12]. Compared to [12], in our case collisions do not provide any useful information to the receiver. Also, the estimation problem itself is trivial due to the i.i.d. observations. The main benefit of a completely decentralized scheduling is that the sensors do not require any knowledge about the state of the network. However, a potential drawback is low performance. The main focus of this work is to compare the expected time to obtain sufficient observations under the decentralized scheduling with the expected time of an optimal centralized scheduling.

Our main contribution is to show that despite the contrasting settings of the two transmission schedules, the decentralized scheduling provides a constant factor approximation to the optimal centralized schedule for the expected waiting time. This demonstrates that scheduling schemes for wireless sensor networks should be designed jointly with the intended data fusion task. Indeed, whereas it is well known that for traditional communication networks, slotted ALOHA is limited in performance, in the current work it is shown that it is

at most a constant factor away from an optimal centralized scheduling scheme.

The remainder of this paper is organized as follows. In Section 2 we formulate the problem statement. In Section 3 we compute the expected time for the decentralized and centralized transmission schedules. In Section 4 we discuss the results and provide conclusions.

## 2. MODEL AND PROBLEM STATEMENT

We consider a wireless sensor network consisting of  $N$  sensor nodes. Each sensor makes an observation  $X_i$  on a attribute  $\theta$ . The observations are subject to independent and identically distributed additive Gaussian noise with variance  $\sigma^2$ , i.e.  $X_i \sim \mathcal{N}(\theta, \sigma^2)$ .

A data collector is interested in estimating  $\theta$  based on the sensor observations. Any subset of  $s$  observations is sufficient. The variance of the estimate  $\bar{X}$  at the collector needs to be below a targeted threshold  $T$ . Since

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{s} \sum_{i=1}^s X_i\right) = \frac{1}{s^2} \sum_{i=1}^s \text{Var}(X_i) = \frac{\sigma^2}{s},$$

it follows that  $s = \lceil \frac{\sigma^2}{T} \rceil$ .

The sensors transmit their observations to a collector. However, simultaneous transmissions lead to a destructive collision and the collector does not obtain any information.

The sensor nodes are awake with probability  $p$ ,  $0 < p < 1$  and asleep with probability  $1 - p$ . Being awake or asleep reflects, for instance, the availability of energy in case of energy harvesting sensor nodes. Sensors can only transmit if they are awake.

In the decentralized  $R$  transmission scheme, an awake sensor transmits the observation with probability  $q$ ,  $0 < q < 1$  and remains silent with probability  $1 - q$ . In the analysis of this transmission scheme we will optimize over  $q$ .

The optimal centralized  $C$  scheme assumes that the sensor nodes are centrally scheduled for transmission based on their on/off status and whether the collector has already received their observation. More precisely, a sensor that has not successfully transmitted its observation previously is considered eligible. If two or more sensor nodes are awake and eligible for transmission during the same time slot, then one of the sensors is randomly selected for transmission.

We are interested in the expected waiting time  $E[W]$  under the decentralized and centralized transmission scheduling such that the collector retrieves data of sufficient accuracy.

We will make use of the digamma function, defined as  $\psi(n) = H_{n-1} - \gamma$ ,  $n \in \mathbb{N}$ , where  $H_n = \sum_{k=1}^n \frac{1}{k}$  is the  $n^{\text{th}}$  harmonic number, and  $\gamma$  is the Euler-Mascheroni constant.

## 3. ANALYSIS

We analyze the expected waiting time for a collector to retrieve sufficient sensor observations under the decentralized and centralized transmission scheduling.

**Theorem 1.** *For the optimal choice of  $q$ , the expected waiting time for a collector to get  $s$  distinct observations under the decentralized  $R$  scheduling is:*

$$\mathbb{E}[W^R] = \begin{cases} f(N, s) \frac{1}{p(1-p)^{N-1}}, & \text{if } p \in (0, \frac{1}{N}) \\ N f(N, s) (\frac{N}{N-1})^{N-1}, & \text{if } p \in [\frac{1}{N}, 1) \end{cases}$$

where  $f(N, s) = [\psi(N+1) - \psi(N-s+1)]$ .

*Proof.* Since the scheduling mechanism is independent and identical over time, the time it takes for the collector to obtain the  $i$ -th distinct observation after having received the  $(i-1)$ -th observation is the time until the first success in a Bernoulli trail. The  $i$ -th observation is successfully received if one of the  $N-i+1$  sensors have not yet been collected, is the single sensor transmitting, i.e. the probability of success is

$$S_i = \binom{N}{1} pq (1-pq)^{N-1} \frac{N-i+1}{N}. \quad (1)$$

Now,

$$\begin{aligned} \mathbb{E}[W^R] &= \sum_{i=1}^s \frac{1}{S_i} \\ &= \sum_{i=1}^s \frac{1}{\left[ pq (1-pq)^{N-1} \right] (N-i+1)} \\ &= \frac{\psi(N+1) - \psi(N-s+1)}{pq (1-pq)^{N-1}}. \end{aligned} \quad (2)$$

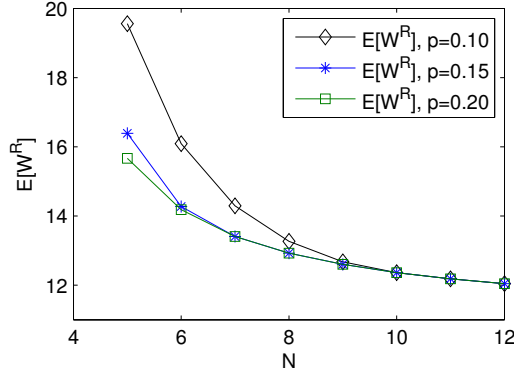
It remains to optimize over  $q \in [0, 1]$ . The gradient of  $\mathbb{E}[W^R]$  is

$$\frac{d}{dq} \mathbb{E}[W^R] = \frac{\psi(N+1) - \psi(N-s+1)}{p} \frac{Npq - 1}{q^2(1-pq)^N}.$$

This shows that if  $p \geq \frac{1}{N}$ , then  $E[W^R]$  is minimized for  $q = \frac{1}{Np}$ . In the case that  $p < \frac{1}{N}$ , the value of  $q$  for which  $E[W^R]$  is minimized is  $q = 1$ .  $\square$

Figure 1 shows that for a sufficiently large size of the sampling sensor set  $N$ ,  $\mathbb{E}[W^R]$  is independent of the transmission probability  $p$  and becomes a function of  $N$ . This is explained by the fact that the probabilities of one sensor transmitting and being awake compensate each other in order to minimize the waiting time. For example, a low probability of being awake is compensated by a high probability of transmitting when awake such that the waiting time is minimized.

Next, we analyze the centralized transmission scheme.



**Fig. 1.**  $E[W^R]$  for different transmitting probabilities,  $s = 4$ .

**Theorem 2.** *The expected waiting time for the data collector to retrieve  $s$  distinct observations under the centralized scheduling is:*

$$\mathbb{E}[W^C] = \sum_{i=1}^s \frac{1}{1 - (1-p)^{N-i+1}}.$$

*Proof.* Again, the scheduling mechanism is independent and identical over time. Therefore, the time until the collector retrieves the  $i$ -th sensor observation, given it already received  $i-1$  distinct observations, can be viewed as the time until a first success in a Bernoulli trial, where the success probability  $T_i$  is the probability that one sensor of those that have not previously transmitted is awake. Hence, the probability of success  $T_i$  is:

$$T_i = 1 - (1-p)^{N-i+1}$$

Now,

$$\mathbb{E}[W^C] = \sum_{i=1}^s \frac{1}{T_i} = \sum_{i=1}^s \frac{1}{1 - (1-p)^{N-i+1}}.$$

**Theorem 3.** *Let  $s \in \mathbb{N}$ . Then*

$$\lim_{N \rightarrow \infty} \frac{E[W^R]}{E[W^C]} = e.$$

*Proof.* For  $N$  sufficiently large,

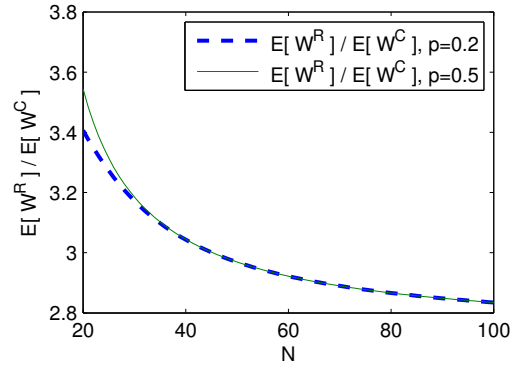
$$\begin{aligned} \lim_{N \rightarrow \infty} E[W^R] &= \lim_{N \rightarrow \infty} \sum_{i=1}^s \frac{N}{N-i+1} \left(1 + \frac{1}{N-1}\right)^{N-1} \\ &= se. \end{aligned} \quad (3)$$

Also, considering  $p$  to be fixed,

$$\begin{aligned} \mathbb{E}[W^C] &= \sum_{i=1}^s \frac{1}{1 - (1-p)^{N-i+1}} = \sum_{i=1}^s \sum_{j=0}^{\infty} ((1-p)^{N-i+1})^j \\ &\geq \sum_{i=1}^s 1 + (1-p)^{N-i+1} \\ &= s + \frac{1 - (1-p)^s}{1 - (1-p)} (1-p)^{N+1-s}. \end{aligned} \quad (4)$$

Using (3) and (4),  $\lim_{N \rightarrow \infty} \frac{E[W^R]}{E[W^C]} = e$ .  $\square$

Figure 2 shows that for large  $N$ , the ratio of the expected waiting time under the decentralized and centralized schemes approaches a constant  $e$  from above. Also, for sufficiently large  $N$ , the ratio is bounded from above by a constant.



**Fig. 2.** The ratio  $E[W^R]/E[W^C]$  for large  $N$ ,  $s = 10$ .

We now consider  $s$  to be a fraction smaller than the sampling sensor size  $N$ , with  $N/s = \beta$ ,  $\beta > 1$ .

**Theorem 4.** *For sufficiently large  $N$  and  $s$  and a constant ratio  $\beta = N/s$ , where  $\beta > 1$ ,*

$$\frac{\mathbb{E}[W^R]}{\mathbb{E}[W^C]} \leq e[\beta] \log\left(\frac{\beta}{\beta-1}\right).$$

$\square$

*Proof.* We now consider both  $N$  and  $s$  sufficiently large and keep the ratio  $\beta = N/s$  constant, where  $\beta > 1$ .

$$\begin{aligned} \mathbb{E}[W^R] &= \lceil \beta s \rceil \sum_{x=\lceil \beta s \rceil - s + 1}^{\lceil \beta s \rceil} \frac{1}{x} \cdot \left( \frac{\lceil \beta s \rceil}{\lceil \beta s \rceil - 1} \right)^{\lceil \beta s \rceil - 1} \\ &\leq \lceil \beta s \rceil \sum_{x=\beta s - s + 1}^{\beta s} \frac{1}{x} \cdot \left( \frac{\lceil \beta s \rceil}{\lceil \beta s \rceil - 1} \right)^{\lceil \beta s \rceil - 1} \\ &\leq \lceil \beta s \rceil \int_{\beta s - s + 1}^{\beta s} \frac{1}{x-1} dx \cdot \left( \frac{\lceil \beta s \rceil}{\lceil \beta s \rceil - 1} \right)^{\lceil \beta s \rceil - 1} \\ &\leq \lceil \beta \rceil s \log\left(\frac{\beta s}{\beta s - s}\right) \left( \frac{\lceil \beta s \rceil}{\lceil \beta s \rceil - 1} \right)^{\lceil \beta s \rceil - 1} \end{aligned} \quad (5)$$

Using (5) and the fact that  $\mathbb{E}[W^C] \geq s$ , see (4),

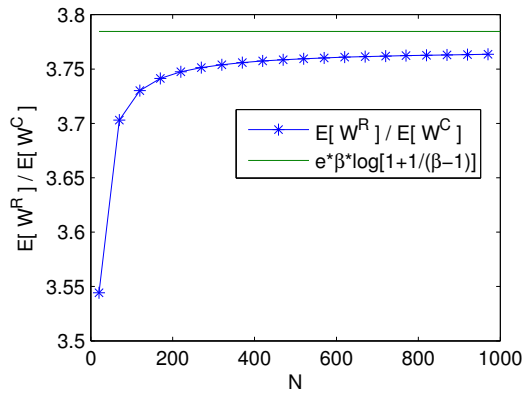
$$\frac{\mathbb{E}[W^R]}{\mathbb{E}[W^C]} \leq \lceil \beta \rceil \log \left( \frac{\beta}{\beta - 1} \right) \cdot \left( 1 + \frac{1}{\lceil \beta s \rceil - 1} \right)^{\lceil \beta s \rceil - 1}$$

Note that  $\left( 1 + \frac{1}{\lceil \beta s \rceil - 1} \right)^{\lceil \beta s \rceil - 1}$  approaches  $e$  from below. Therefore,

$$\frac{\mathbb{E}[W^R]}{\mathbb{E}[W^C]} \leq e \lceil \beta \rceil \log \left( \frac{\beta}{\beta - 1} \right)$$

Notice that  $\lim_{N \rightarrow \infty} \lceil \beta \rceil \log \left( \frac{\beta}{\beta - 1} \right) = 1$ , in accordance with the result of Theorem 3.  $\square$

Figure 3 shows that for a fixed ratio  $\beta$  of the size of the sampling sensor set  $N$  and the  $s$  retrieved observations, where  $N$  and  $s$  are large, the waiting time under the decentralized schedule is no higher than a constant  $e \lceil \beta \rceil \log \left( \frac{\beta}{\beta - 1} \right)$  than the centralized schedule.



**Fig. 3.** The limiting ratio  $E[W^R]/E[W^C]$  for large  $N$ , when  $\beta = 2$ .

#### 4. DISCUSSION AND CONCLUSIONS

We investigated the expected waiting time for a collector to retrieve sufficient sensor observations on an attribute such that the estimate of the attribute has a variance below a targeted threshold. We analyzed the expected waiting time for both decentralized and centralized sensor transmission schemes. We showed that the optimal centralized schedule has a lower expected waiting time than the decentralized scheme. However, the centralized schedule assumes information on the awake/asleep status of the sensors and the redundancy of the observations transmitted, which is difficult to achieve in reality. Nonetheless, we showed that the decentralized schedule, which requires no coordination between the sensors is a constant larger than the optimal centralized scheme when

the sampling set of sensors  $N$  is sufficiently large. Additionally, we showed that for large  $N$ , the waiting time to retrieve sufficient observations does not depend on the probability of transmitting or being awake.

Future work includes investigating the waiting time for multiple collectors to retrieve sufficient data from sensors randomly placed in the plane. In addition, we plan to investigate the waiting time for data estimation when sensors' observations are correlated.

**Acknowledgements:** This work was performed within the project RRR (**R**ealisation of **R**eliable and **S**ecure **R**esidential Sensor Platforms) of the Dutch program *IOP Generieke Communicatie*, number IGC1020, supported by the *Subsideregeling Sterktes in Innovatie*, and is partly supported by the *Netherlands Organisation for Scientific Research* (NWO), grant 612.001.107.

#### 5. REFERENCES

- [1] J. Manyika and H. Durrant-Whyte, *Data fusion and sensor management: a decentralized information-theoretic approach*. Prentice Hall PTR, 1995.
- [2] P. Varshney, *Distributed detection and data fusion*. Springer-Verlag New York, Inc., 1996.
- [3] Z. Luo and J. Tsitsiklis, "Data fusion with minimal communication," *IEEE Transactions on Information Theory*, vol. 40, no. 5, pp. 1551–1563, 1994.
- [4] P. Varshney, "Distributed bayesian hypothesis testing with distributed data fusion," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 18, no. 5, pp. 695–699, 1988.
- [5] J. Xiao, S. Cui, Z. Luo, and A. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 413–422, 2006.
- [6] J. Xiao and Z. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3564–3575, 2005.
- [7] J. Wu, Q. Huang, and T. Lee, "Minimal energy decentralized estimation via exploiting the statistical knowledge of sensor noise variance," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 2171–2176, 2008.
- [8] S. Sengijpta, "Fundamentals of statistical signal processing: Estimation theory," *Technometrics*, vol. 37, no. 4, pp. 465–466, 1995.
- [9] A. Bachir, M. Dohler, T. Watteyne, and K. Leung, "MAC essentials for wireless sensor networks," *IEEE*

*Communications Surveys & Tutorials*, vol. 12, no. 2, pp. 222–248, 2010.

- [10] I. Demirkol, C. Ersoy, and F. Alagoz, “MAC protocols for wireless sensor networks: a survey,” *IEEE Communications Magazine*, vol. 44, no. 4, pp. 115–121, 2006.
- [11] J. Zhang and Y. Hu, “Data centric multi-shift sensor scheduling for wireless sensor networks,” in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Vancouver*, pp. 4594–4597, 2013.
- [12] G. Whipps, E. Ertin, and R. Moses, “Distributed detection with collisions in a random, single-hop wireless sensor network,” in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Vancouver*, pp. 4603–4607, 2013.
- [13] D. Bertsekas, R. Gallager, and P. Humblet, *Data networks*. Prentice-Hall International, 1992.