A Note on the Complexity of Some Quality of Information Optimisation Problems in Sensor Networks

Research Report 441
Ellen Munthe-Kaas
3rd October, 2014

ISBN 978-82-7368-406-6
ISSN 0806-3036
A Note on the Complexity of Some Quality of Information Optimisation Problems in Sensor Networks

Ellen Munthe-Kaas

Abstract

We prove that a selection of quality of information optimisation problems related to event detection in sensor networks are NP-hard.

1 Introduction

In event detection systems, where the aim is to utilize sensors in a network to detect events of interest, there is a need to balance the sensors’ energy usage against the frequency of data transfer in order to prolong the lifetime of the sensors. This leads to a number of optimisation problems, all of which tend to be NP-hard. In this note we review most of the optimisation problems in [1], [2], [3], [4] and prove that the corresponding decision problems are NP-complete.

2 Cost and Quality of Information Functions

Consider an event detection system with a set of sensors \( S \). The aim of the system is to use data from the sensors for detecting an event characterised by an expression \( E \). Each sensor \( i \) has a hop count \( h_i \in N \) (the number of communication links from the sensor to where the sensed data are collected and processed) and a sending rate \( r_i \), where \( N \) is the set of natural numbers (positive integers).

A sensor’s energy usage is proportional to the number of transmitted data tuples. The cost of employing a subset of the sensors \( \theta \subseteq S \) is therefore represented by \( \gamma \sum_{i \in \theta} h_i r_i \), where \( \gamma \) is a scaling constant. The sending rates can be changed dynamically, for instance to lower the cost. For each sensor \( i \), \( r_i \) reflects a percentage of its maximum sending rate. In practice the sensor might accomplish the requested sending rate by repeatedly sampling
and sending $r_i$ data tuples out of a 100 possible, thus we assume that $r_i$ is a non-negative integer in the range $[0..100]$. Since $\gamma$ does not affect the computational complexity of any of the optimisation problems, we for simplicity assume that $\gamma = 1$ and consequently use the cost function

$$\text{Cost}_i(\theta) = \sum_{i \in \theta} h_i r_i \tag{1}$$

where $\bar{r}$ are the sending rates. We write $\text{Cost}(\theta)$ when $\bar{r}$ is obvious from the context.

Each sensor contributes to the system’s information extraction. The quality of this information is influenced by the sensor’s sending rate; the more data tuples provided, the more accurate the information gained from the sensed values. The contribution to the quality of information from a sensor that is actively producing and sending data, is represented by a function $q(\alpha, \beta, r)$, where $r > 0$ is the sending rate and $\alpha > 0$ and $\beta > 0$ are used for tuning the value of $q$ to each individual sensor. The actual shape of $q$ can be chosen in a number of different ways as long as $q(\alpha, \beta, r)$ is monotonically increasing in $r$ and obeys $1/2 \leq q(\alpha, \beta, r) \leq 1$ (assuming reasonable values for $\alpha$ and $\beta$). The rationale behind $q(\alpha, \beta, r) \geq 1/2$ is related to Equation (4), which is introduced further down. When $r_i = 0$, i.e., sensor $i$ does not provide any data, its contribution to the quality of information is 0. Let $\alpha_i$ and $\beta_i$ be the values of $\alpha$ and $\beta$ respectively for sensor $i$. We write $q_i$ for $q(\alpha_i, \beta_i, r_i)$ when the value of $r_i$ is obvious from the context.

We consider the following two possible definitions of $q$, both of which are monotonically increasing in each of $\alpha$, $\beta$ and $r$ (slight variations of these definitions appear in the papers [1], [2], [3], [4]):

$$q(\alpha, \beta, r) = \alpha b^{-\frac{r}{\beta}} \tag{2}$$

$$q(\alpha, \beta, r) = 1 - \frac{1}{\alpha} b^{-\beta r} \tag{3}$$

where $b$ is the base of the exponent (with $b = e$ or $b = 2$ as the most prominent candidates). The quality of information function for a set of mutually independent sensors $\theta$ and an event expression $E$, $QoI[E]_r(\theta)$, is defined recursively wrt. the structure of $E$ [3]:

- $E$ is an atomic event: For atomic events the quality of information function is a fusion of those for the individual sensors, and follows a Bayesian formulation.

$$QoI[E]_r(\theta) = \frac{\prod_{i \in \theta} q_i^{\alpha_i}}{\prod_{i \in \theta} q_i + \prod_{i \in \theta} (1-q_i)} \tag{4}$$

where $\prod_{i \in \theta} q_i$ should be monotonically increasing in the size of $\theta$; the more sensors employed, the more accurately an atomic event can be detected. In order to ensure this, we require that $q_i \geq 1/2$.

- $E$ is a conjunction of events $E_1 \land \cdots \land E_n$, consecutive events $E_1 \rightarrow \cdots \rightarrow E_n$, or concurrent events $E_1 \parallel \cdots \parallel E_n$ (assuming that for $i \neq j$ the atomic events occurring in $E_i$ and $E_j$ are independent):

$$QoI[E]_r(\theta) = \prod_{j=1}^n QoI[E_j]_r(\theta) \tag{5}$$
• \( E \) is a disjunction of events \( E_1 \lor \cdots \lor E_n \) (assuming that for \( i \neq j \) the atomic events occurring in \( E_i \) and \( E_j \) are independent):

\[
QoI[E]_{\vec{r}}(\theta) = 1 - \prod_{j=1}^{n}(1 - QoI[E_j]_{\vec{r}}(\theta)) \tag{6}
\]

We write \( QoI[E](\theta) \) when the sending rates \( \vec{r} \) are obvious from the context.

3 Optimisation Problems

The optimisation problems that we consider in this note, are the following:

**Problem 1. MinCost** – Find a subset of the sensors that minimises the cost while maintaining an acceptable quality of information:

**Instance.** A set of sensors \( S \) and an event expression \( E \). For each sensor \( i \), a sending rate \( r_i \in N \), a hop count \( h_i \in N \), and constants \( \alpha_i > 0 \) and \( \beta_i > 0 \) such that \( 1/2 \leq q_i \leq 1 \). A minimum quality of information value \( Q > 0 \).

**Problem.** Find a subset \( \theta \subseteq S \) such that \( QoI[E](\theta) \geq Q \) and such that \( Cost(\theta) \) is minimised.

**Problem 2. MaxQoI** – Find a subset of the sensors that maximises the quality of information while maintaining an acceptable cost:

**Instance.** A set of sensors \( S \) and an event expression \( E \). For each sensor \( i \), a sending rate \( r_i \in N \), a hop count \( h_i \in N \), and constants \( \alpha_i > 0 \) and \( \beta_i > 0 \) such that \( 1/2 \leq q_i \leq 1 \). A maximum cost \( C \in N \).

**Problem.** Find a subset \( \theta \subseteq S \) such that \( Cost(\theta) \leq C \) and such that \( QoI[E](\theta) \) is maximised.

**Problem 3. MaxQoIBalancing** – Given a partition of the sensors, find for each group in the partition a subset such that the overall quality of information is maximised while maintaining an acceptable cost for each subset as well as an acceptable overall cost:

**Instance.** A set of sensors \( S \) and an event expression \( E \). For each sensor \( i \), a sending rate \( r_i \in N \), a hop count \( h_i \in N \), and constants \( \alpha_i > 0 \) and \( \beta_i > 0 \) such that \( 1/2 \leq q_i \leq 1 \). A partition \( S_1, \ldots, S_k \) of \( S \). A maximum cost \( C \in N \). For each \( S_j \), a local maximum cost \( C_j \in N \).

**Problem.** Find subsets \( \theta_j \subseteq S_j \) such that \( Cost(\theta_j) \leq C_j \) for all \( j \), \( Cost(\theta) \leq C \), and such that \( QoI[E](\theta) \) is maximised, where \( \theta = \bigcup_{j=1}^{k} \theta_j \).
Problem 4. MaxLifetime – Find an allocation of sensors into groups that maximises the number of groups while ensuring that each group provides an acceptable quality of information:

INSTANCE. A set of sensors $S$ and an event expression $E$. For each sensor $i$, a sending rate $r_i \in \mathbb{N}$ and constants $\alpha_i > 0$ and $\beta_i > 0$ such that $1/2 \leq q_i \leq 1$. A minimum quality of information value $Q > 0$.

PROBLEM. Find a subset $S' \subseteq S$ and a partition $\theta_1, \ldots, \theta_k$ of $S'$ such that $QoI[E](\theta_j) \geq Q$ for all $j$ and such that $k$ is maximised.

Problem 5. MaxMinQoI – Find an allocation of sensors into a given number of groups such that the quality of information for the group with the lowest such value, is maximised:

INSTANCE. A set of sensors $S$ and an event expression $E$. For each sensor $i$, a sending rate $r_i \in \mathbb{N}$ and constants $\alpha_i > 0$ and $\beta_i > 0$ such that $1/2 \leq q_i \leq 1$. A constant $k \in \mathbb{N}$.

PROBLEM. Find a subset $S' \subseteq S$ and a partition $\theta_1, \ldots, \theta_k$ of $S'$ such that

$$\min_{1 \leq j \leq k} \{QoI[E](\theta_j)\}$$

is maximised.

In MaxLifetime and MaxMinQoI, if a problem instance has more than one optimal solution, one might additionally consider optimising the subset $S'$, for instance by choosing the solution with the smallest $\text{Cost}(S')$. If such a further optimisation of $S'$ is not a concern, one may simplify the problems to finding a partition $\theta_1, \ldots, \theta_k$ of $S$ rather than a partition of $S'$. We have however chosen to keep the original problem formulations.

4 Decision Problems

In order to provide complexity results for the problems in Section 3, we first state the corresponding decision problems.

The decision versions of Problems 1 and 2 are identical:

Problem 6. MinCostDec and MaxQoIDec

INSTANCE. A set of sensors $S$ and an event expression $E$. For each sensor $i$, a sending rate $r_i \in \mathbb{N}$, a hop count $h_i \in \mathbb{N}$ and constants $\alpha_i > 0$ and $\beta_i > 0$ such that $1/2 \leq q_i \leq 1$. A minimum quality of information value $Q > 0$ and a maximum cost $C \in \mathbb{N}$.

QUESTION. Is there a subset $\theta \subseteq S$ such that $QoI[E](\theta) \geq Q$ and $\text{Cost}(\theta) \leq C$?

If we for each sensor $i$ abstract the cost and quality of information to constants $c_i$ and $p_i$ respectively, and restrict our focus to atomic events, we get the following formulation of Problem 6, where $pf(\theta) = \frac{\Pi_{i \in \theta} p_i}{\Pi_{i \in \theta} c_i \Pi_{i \in \theta} (1 - p_i)}$ is the fusion of the $p_i$’s:
Problem 7. Bayesian Profit

INSTANCE. A set of items $U$. For each item $i$, a profit $p_i$ where $1/2 \leq p_i \leq 1$, and a cost $c_i \in N$. A minimum profit $P > 0$ and a maximum cost $C \in N$.

QUESTION. Is there a subset $\theta \subseteq U$ such that $pf(\theta) \geq P$ and $\sum_{i \in \theta} c_i \leq C$?

We simplify the decision versions of Problems 3–5 similarly, noting that the decision versions of Problems 4 and 5 are identical:

Problem 8. MaxQoIBalancingDec

INSTANCE. A set of items $U$. For each item $i$, a profit $p_i$ such that $1/2 \leq p_i \leq 1$, and a cost $c_i \in N$. A partition $U_1, \ldots, U_k$ of $U$. A minimum profit $P > 0$ and a maximum cost $C \in N$. For each $U_j$, a local maximum cost $C_j \in N$.

QUESTION. Are there subsets $\theta_j \subseteq U_j$ such that $pf(\theta_j) \geq P$, $\sum_{i \in \theta_j} c_i \leq C_j$ for all $j$, and $\sum_{i \in \theta} c_i \leq C$, where $\theta = \bigcup_{j=1}^k \theta_j$?

Problem 9. MaxLifetimeDec and MaxMinQoIDec

INSTANCE. A set of items $U$. For each item $i$, a profit $p_i$ such that $1/2 \leq p_i \leq 1$. A minimum profit $P > 0$. A constant $k \in N$.

QUESTION. Is there a subset $V \subseteq U$ and a partition $\theta_1, \ldots, \theta_k$ of $V$ such that $pf(\theta_j) \geq P$ for $j = 1, \ldots, k$?

5 Complexity of the Problems

Our aim is to prove that all the decision problems in Section 4 are NP-complete. We base our proofs on the known NP-complete decision problems Knapsack [5] (often named 0-1 Knapsack) and Dual Bin Packing [6]:

Problem 10. Knapsack

INSTANCE. A set of items $U$. For each item $i$, a profit $p_i \in N$ and a weight $w_i \in N$. A minimum profit $P \in N$ and a maximum weight $W \in N$.

QUESTION. Is there a subset $V \subseteq U$ such that $\sum_{i \in V} w_i \leq W$ and $\sum_{i \in V} p_i \geq P$?

Problem 11. Dual Bin Packing

INSTANCE. A set of items $U$. For each item $i$, a profit $p_i \in N$. A minimum profit $P \in N$. A constant $k \in N$.

QUESTION. Is there a partition of $U$ into $k$ sets $U_1, \ldots, U_k$ such that $\sum_{i \in U_j} p_i \geq P$ for $j = 1, \ldots, k$?
In the proofs below we use tilde ($\tilde{X}$) to emphasize that an entity $X$ belongs to the transformed problem.

**Lemma 1.** Bayesian Profit is NP-complete.

*Proof.* Transformation from Knapsack: Let

$$
\tilde{U} = U
\tilde{p}_i = 1 - \frac{1}{1+2^{\theta_i}}
\tilde{c}_i = w_i
\tilde{P} = \frac{1}{1+2^{\theta_i}}
\tilde{C} = W
$$

Since $p_i \in N \Rightarrow 2^{p_i} \geq 1$, we have $0 \leq \frac{1}{1+2^{\theta_i}} \leq 1/2$, and consequently $1/2 \leq \tilde{p}_i \leq 1$. $\theta \subseteq U$ is a witness for Knapsack if and only if it is a witness for the transformed problem: $\sum_{i \in \theta} w_i = \sum_{i \in \theta} c_i$, so $\sum_{i \in \theta} w_i \leq W$ iff $\sum_{i \in \theta} \tilde{c}_i \leq \tilde{C}$. Further, $\tilde{p}_i = \frac{2^{p_i}}{1+2^{\theta_i}}$, so $\frac{1-\tilde{p}_i}{\tilde{p}_i} = 2^{-p_i}$. Thus,

$$
pf(\theta) = \frac{\prod_{i \in \theta} \tilde{p}_i}{\prod_{i \in \theta} \tilde{p}_i + \prod_{i \in \theta} (1-\tilde{p}_i)} = \frac{1}{1+\prod_{i \in \theta} \frac{1-\tilde{p}_i}{\tilde{p}_i}} = \frac{1}{1+\prod_{i \in \theta} 2^{-\theta_i}} = \frac{1}{1+2^{-\sum_{i \in \theta} \theta_i}}.
$$

Consequently,

$$
\sum_{i \in \theta} p_i \geq P \iff 2^{-\sum_{i \in \theta} p_i} \leq 2^{-P} \iff pf(\theta) \geq \frac{1}{1+2^{-\theta_i}} = \tilde{P}.
$$

□

**Lemma 2.** MinCostDec and MaxQoIDec are NP-complete.

*Proof.* Transformation from Bayesian Profit: Let

$$
\tilde{S} = U
\tilde{E} = \text{any atomic event}
\tilde{\alpha}_i = 1
\tilde{r}_i = c_i
\tilde{h}_i = 1
\tilde{Q} = P
\tilde{C} = C
$$

In addition we choose the value of $\tilde{\beta}_i$ so that $\tilde{q}_i = p_i$ (thus, $1/2 \leq \tilde{q}_i \leq 1$). The actual choice of $\tilde{\beta}_i$ depends on the definition of $q(\alpha, \beta, r)$:

$$
\tilde{\beta}_i = -1/(c_i \log_b p_i) \quad \text{if } q(\alpha, \beta, r) = \alpha b^{-\frac{1}{r}},
\tilde{\beta}_i = -\log_b(1-p_i)/c_i \quad \text{if } q(\alpha, \beta, r) = 1 - \frac{1}{\alpha} b^{-\beta r}.
$$

Let $\theta \subseteq U$. Then $\theta$ is a witness for Bayesian Profit if and only if it is a witness for the transformed problem: $\sum_{i \in \theta} \tilde{h}_i \tilde{r}_i = \sum_{i \in \theta} c_i$, so $\sum_{i \in \theta} c_i \leq C$ iff $\text{Cost}(\theta) \leq \tilde{C}$. Since $\tilde{q}_i = p_i$ and the event is atomic,

$$
QoI[\tilde{E}](\theta) = \prod_{i \in \theta} p_i / (\prod_{i \in \theta} p_i + \prod_{i \in \theta} (1-p_i)) = pf(\theta),
$$

and thus $pf(\theta) \geq P$ iff $QoI[\tilde{E}](\theta) \geq \tilde{Q}$. □
Lemma 3. MaxQoIBalancingDec is NP-complete.

Proof. The transformation from Bayesian Profit is straightforward: Let
\[
\begin{align*}
\tilde{U} &= U, \\
\tilde{p}_i &= p_i, \\
\tilde{c}_i &= c_i, \\
\tilde{k} &= 1, \\
\tilde{U}_1 &= U, \\
\tilde{P} &= P, \\
\tilde{C} &= C, \\
\tilde{C}_1 &= C
\end{align*}
\]
Then \( \theta \subseteq U \) is a witness for Bayesian Profit iff it is a witness for the transformed problem. \( \square \)

Lemma 4. MaxLifetimeDec and MaxMinQoIDec are NP-complete.

Proof. Transformation from Dual Bin Packing: Let
\[
\begin{align*}
\tilde{U} &= U, \\
\tilde{p}_i &= 1 - \frac{1}{1 + 2^{p_i}}, \\
\tilde{k} &= k, \\
\tilde{P} &= k \frac{1}{1 + 2 - p}
\end{align*}
\]
Since \( 2^{p_i} \geq 1 \), \( 1/2 \leq \tilde{p}_i \leq 1 \). Since \( \frac{1 - \tilde{p}_i}{p_i} = 2^{-p_i} \), we get that for all \( \theta \subseteq S \), \( pf(\theta) = \frac{1}{1 + 2 - \sum_{i \in \theta} p_i} \). Consequently, for all \( \theta \subseteq S \), \( \sum_{i \in \theta} p_i \geq P \) iff \( pf(\theta) \geq \frac{1}{1 + 2 - P} = \tilde{P} \).

Let \( \theta_1, \ldots, \theta_k \) be a witness for the transformed problem (with \( V = \bigcup_{j=1}^k \theta_j \)). Then for all \( j \), \( pf(\theta_j) \geq \tilde{P} \), and consequently \( \sum_{i \in \theta_j} p_i \geq P \) for all \( j \). Let \( \theta_0 = U \setminus \bigcup_{j=1}^k \theta_j \) and let
\[
\theta'_i = \begin{cases} 
\theta_1 \cup \theta_0 & \text{if } i = 1, \\
\theta_i & \text{if } i > 1.
\end{cases}
\]
Then for all \( j > 1 \), \( \sum_{i \in \theta'_j} p_i = \sum_{i \in \theta_j} p_i \geq P \). Further, \( \sum_{i \in \theta'_j} p_i \geq \sum_{i \in \theta_i} p_i \geq P \), so \( \theta'_1, \ldots, \theta'_k \) is a witness for Dual Bin Packing.

Let \( \theta_1, \ldots, \theta_k \) be a witness for Dual Bin Packing. Then for all \( j \), \( \sum_{i \in \theta_j} p_i \geq P \), and consequently \( pf(\theta_j) \geq \tilde{P} \) for all \( j \), so \( \theta_1, \ldots, \theta_k \) is a witness for the transformed problem (with \( V = U = \bigcup_{j=1}^k \theta_j \)). \( \square \)
6 Conclusion

As the decision problems are provably NP-complete, the corresponding optimisation problems are NP-hard.

References


