A Virtual Sensor Scheduling Framework for Heterogeneous Wireless Sensor Networks

Wen Hu, Damien O'Rourke, Branislav Kusy, Tim Wark
CSIRO ICT Centre, Brisbane, Australia
Email: {wen.hu, damien.o'rourke, brano.kusy, tim.wark}@csiro.au

***Abstract*—We investigate the problem of scheduling sensor node up-times to maximize the utility of the data they collect while operating within their resource constraints. We show that the optimal scheduling algorithm can improve data utility by more than 70% compared to naive schedules. We consider a suite of sensors with different capabilities and resource demands and represent their subsets as virtual sensors. For each virtual sensor, we calculate its optimal data fusion parameters and evaluate the sensors' performance in a given environment. The selection of virtual sensors best suited to collect data in a given environment can be modeled as an Integer Linear programming problem, and we study three different algorithms to solve the problem efficiently. We evaluate the performance of virtual sensor scheduling algorithms by extensive simulation. We show that even though the naive greedy scheduling approaches work well in some scenarios, none of them are able to match our best scheduling algorithm consistently, under varying environmental conditions and sensor resources.**

I. Introduction

In this paper we focus on heterogeneous sensor networks where it is assumed many nodes are dealing with high sample-rate signals [1], however the techniques developed could apply to any class of high-sample rate sensor networks. Despite recent hardware advances, interpretation of more complex, high-sample rate samples remains a difficult and energy consuming task due to the sheer volume of the data involved. Data collection and analysis of scalar sensors, such as Passive InfraRed (PIR) sensors, requires a few orders of magnitude *less* energy than that of multimedia data, such as 2D images, however the discrimination capability of such sensor types is typically quite poor leading to high probabilities of error when used in isolation. One solution is for individual sensors to be scheduled at different times, where the combination or fusion of sensors turning on concurrently maximizes the likely utility or performance given the energy consumption required at that point. Data fusion theory [2] shows that it is possible to improve the event detection capability of a system by combining (fusing) the output of local sensors deployed in the field. In other words, by fusing the output of existing physical sensors, a new set of *virtual* sensors is effectively formed that can have greater performance for the same energy cost. The focus of this work is to find the optimal schedules for multi-node heterogeneous class systems such that the overall utility of the system, subject to energy constraints, is maximized.

Our approach to the problem described is to propose a theoretical model that treats combinations of low and high power physical sensors as *virtual sensors*. The virtual sensor definition is quite general and can possibly span multiple nodes and a subset of their physical sensors. It covers data processing of the raw sensor readings for each included physical sensor, fusion of data both within and across nodes, as well as communication between different physical nodes. In addition to being able to quantify accuracy and efficiency of any virtual sensor, our model can be used to find optimal virtual sensors in a freely deployed sensor network. We introduce three efficient approximation algorithms (heuristics) based on Greedy and Tabu searches [3] that solve the problem for larger systems efficiently. In particular, we explore how the scheduling of virtual sensors affects the performance of the network and demonstrates which are the critical system parameters we need to consider. In summary, the primary contributions of this paper are as follows:

- The development of a novel virtual sensor abstraction that includes physical sensors, data processing, and data fusion of the sensors.
- An optimization model to find optimal virtual sensor schedules in deployed heterogeneous networks, as well as an efficient schedule framework that is based on the virtual sensor abstraction and optimal virtual sensor schedule model to study the performance of these classes of systems
- Validation of the benefits of the proposed framework by extensive empirically motivated simulation case studies.

II. Architecture

Our formulation of this problem is based on some key assumptions as follows:

Discrete time intervals: Over the lifetime of a network, we assume that time can be split into a series of discrete time slots. The length of the slots is application dependent: accuracy of the event classification defines the lower bound and granularity of the event detection limits the upper bound. A single virtual sensor is defined in each slot and virtual sensors cannot change within the slot. We assume a common sensing period t, with each sensor generating samples and making a local decision as to whether an interesting event occurred. We schedule the duty cycles of the sensor over T time slots. Accordingly, we index time to be integer-valued, with $0 \leq t \leq T$, where t refers to the t^{th} time slot.

Event types are defined before deployment: We assume that a finite set of events are defined prior to a network's deployment, for example, specific birdcalls of interest. This assumption implies that the proposed framework is limited to event types that are known in advance and is therefore unsuitable for the detection of unknown events.

Predictable event arrival patterns: We assume the event arrival times are independent and that their distribution is known in advance. For this work we assume a Poisson distribution of the number of events arriving per time slot (see section IV); this leads to an exponential distribution for the actual arrival times of those events. The use of a Poisson distribution in this sense is well known and has been observed empirically

978-1-4799-0537-9/13/$31.00

in outdoor deployments [4]. The single unknown parameter of the Poisson and exponential distributions is referred to as λ and is the mean arrival rate of events. An example of such a distribution could be the daily pattern of events which show similar values of λ at various time intervals (e.g. traffic conditions at peak and off-peak times) .

Areas of interest are application/event dependent: We assume that the events of interest are, in theory, detectable by the available physical sensors. If an event is always out of range of all physical sensors, i.e. never detectable, then we do not consider it at all.

Based on these assumptions, our system architecture for determining the scheduling of sensors is shown in Figure 1. A key aspect of the architecture is that the pre-deployment steps of forming fusion rules, and assigning optimal virtual sensors and their scheduling, are kept separate to reflect the fact that these are indeed independent functions.

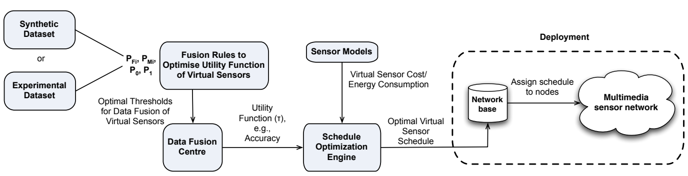


Fig. 1. System architecture. P_0, P_1: the prior probabilities of *events not of interest* and *events of interest* respectively. p_{m_i} and p_{f_i}: the probabilities of a *miss* and *false alarm*, respectively.

In the pre-deployment phase, we use either synthetic or experimental data to determine appropriate fusion rules for the various combinations of physical sensors which can make up a virtual sensor. The intuition is that the fusion rules define the optimal set of sensors for each event as well as the algorithm to combine the sensor readings.

Once fusion rules have been determined for each virtual sensor, we find an optimal schedule to turn the sensors on using our optimization engine. The optimization step requires sample data that is indicative of the event patterns occurring in the actual deployment. The process for assigning this optimal schedule is described in detail in Section III. Finally, once the schedule of the virtual sensors has been determined, it is provided to a network base which will upload it to a network, allowing all nodes to follow their assigned virtual sensor schedule for each cycle.

III. Optimal Scheduling

A. Notation

We model a tiered WSN as a set of nodes $\mathbf{N} = \{n_i : 1 \leq i \leq N\}$ and a set of physical sensors $\mathbf{S} = \{s_j : 1 \leq j \leq M\}$. We represent the initial energy resources at each node by $\mathbf{E} = \{e_i, 1 \leq i \leq N\}$. Physical sensors and nodes form a many-to-1 relationship: many physical sensors (possibly of the same type) may be located on the same node.

We also define an indicator function that identifies active physical sensors as follows:

$$\beta_{ij}^t = \begin{cases} 1 & \text{if a physical sensor } s_j \text{ at node } n_i \text{ is turned} \\ & \text{on at time } t, \\ 0 & \text{otherwise.} \end{cases}$$

A physical sensor s_j consumes ϵ_j joules of energy over a time slot, if turned on. This includes any energy required to process the raw sensor data; on aggregate, a node n_i consumes $\sum_{t=1}^{T} \sum_{j=1}^{M} \beta_{ij}^t \epsilon_j$ joules.

B. Virtual Sensors and Cost

Definition 1: A virtual sensor v_k, where $0 \leq k \leq V$ and $V = 2^M$ is the total number of possible virtual sensors in the network, is defined as a subset of $\mathbf{S}$ along with an associated fusion function.

We further define an indicator function:

$$\gamma_{ij}^k = \begin{cases} 1 & \text{if a virtual sensor } v_k \text{ includes a physical} \\ & \text{sensor } s_j \text{ at node } n_i, \\ 0 & \text{otherwise.} \end{cases}$$

For example, if virtual sensor v_3 is comprised of the second InfraRed sensor (IR) located on node n_4 and the first Image Sensor (IS) on node n_7, we can model this association as: $\gamma_{4,\mathrm{IR}_2}^3 = \gamma_{7,\mathrm{IS}_1}^3 = 1$. Our model can therefore represent situations in which more than two sensors of the same type are located at one node.

We define c as the communication cost for the fusion algorithm that combines sensor readings taken at different nodes. In this model, we limit the fusion algorithm to associate with the nodes within one-hop neighborhood as the communication range of a node is significantly larger than its sensing range. For example, common off-the-shelf PIR sensors have ranges on the order of 10 m, whereas the AT86RF212 has a communication range of more than 100 m in an outdoor environment [5]. Therefore, the communication cost (local broadcast) is fixed for all nodes. If we define δ_i^t as:

$$\delta_i^t = \begin{cases} 1 & \text{if node } n_i \text{ communicates at time } t, \\ 0 & \text{otherwise.} \end{cases}$$

Then, the communication cost of node n_i is $\sum_{t=1}^{T} \delta_i^t c$. Finally, the energy cost of node n_i is:

$$\sum_{t=1}^{T} \sum_{j=1}^{M} \beta_{ij}^t \epsilon_j + \sum_{t=1}^{T} \delta_i^t c. \tag{1}$$

We define another variable, x_k^t as:

$$x_k^t = \begin{cases} 1 & \text{if virtual sensor } v_k \text{ is turned on at time } t \\ 0 & \text{otherwise.} \end{cases}$$

Turning a virtual sensor v_k on at time t, implies that all the physical sensors associated with v_k are turned on at time t. We don't allow two virtual sensors to be turned on in the same time slot. If required, the user should define a new virtual sensor that consists of all physical sensors in the two virtual sensors.

C. Utility Functions

Definition 2: The utility function τ_k of a sensor v_k is defined as the probability of successful operations of the sensor over one time slot.

Depending on the applications, examples of τ_k could be:

1) $1 - p_e$
2) Sensitivity: $T_p/(T_p + F_n)$
3) Specificity: $T_n/(T_n + F_p)$
4) Accuracy: $(T_p + T_n)/(T_p + T_n + F_p + F_n)$

p_e is the probability of error given by $P_0 p_f + P_1 p_m$. The values T_p, T_n, F_p, and F_n represent the number of true and false positives and negatives. For the evaluation presented in this paper we chose to use $1 - p_e$.

D. Optimization

We model our scheduling mechanism as a *0,1* Integer Linear Programming (ILP) problem. The objective of the optimization is to determine the scheduling of virtual sensors so as to maximize the overall utility of a heterogeneous WSN. The utility of a WSN is the summation of all utilities for the virtual sensors over time T. This is because there is *at most* one event and *at most* one virtual sensor turned on in any time slot. It should be noted that these constraints will not pose any practical issues. In the first case we can simply consider the concept of *atomic* and *compound* events as proposed in [6]. An atomic event is one in which exactly one object having one or more attributes is involved in exactly one activity. A compound event is simply the composition of two or more different atomic events. The overall net effect means that we can abstract out the event details to consider only one event per time slot - whether it be atomic or compound. In the second case we justify only one virtual sensor per time slot by reiterating the attribute of a virtual sensor as stated in section II: A virtual sensor can be made up of *any* combination of physical sensors - from one to all, or anywhere in between. The problem can therefore be formulated as:

Maximise:

$$\sum_{t=1}^{T}\sum_{k=1}^{V} x_k^t \tau_k \quad (2)$$

subject to:

$$\sum_{j=1}^{M}\sum_{t=1}^{T}\sum_{k=1}^{V} \gamma_{ij}^k x_k^t \epsilon_j + \sum_{t=1}^{T} \delta_i^t c \le e_i, \forall i \quad (3)$$

$$\delta_i^t \le \sum_{j=1}^{M}\sum_{k=1}^{V} \gamma_{ij}^k x_k^t, \forall i, t \quad (4)$$

$$\sum_{j=1}^{M}\sum_{k=1}^{V} \gamma_{ij}^k x_k^t \le \delta_i^t B, \forall i, t \quad (5)$$

$$\sum_{k=1}^{V} x_k^t \le 1, \forall t \quad (6)$$

$$x_k^t \in \{0,1\}, \forall t, k \quad (7)$$

$$\delta_i^t \in \{0,1\}, \forall i, t. \quad (8)$$

Constraint (3) ensures that the energy consumption is less than the available energy resources at each sensor node. Constraints (4) and (5) ensure that node n_i consumes energy for communication once, and once only, at each time slot t if at least one physical sensor s_j, located at the node, is turned on at t. B is a very large constant such that $B > \sum_{j=1}^{M}\sum_{k=1}^{V} \gamma_{ij}^k x_k^t, \forall i, t$. Finally, constraints (7) and (8) define the scope of variables x_k^t, and δ_i^t respectively.

E. Heuristic Algorithms

Since the ILP problem introduced in Section III is NP-complete, it is very inefficient to solve it to obtain the optimal solution. As an example, we found that the maximum number of time slots for a nine node network that the *state-of-the-art* commercial optimization package, ILOG CPLEX[1], can handle efficiently is only six. Thus, the results produced by CPLEX are not very helpful for real network deployments. To address this issue we develop and compare three heuristic algorithms allowing us to efficiently obtain approximate solutions to the optimization problem. The three heuristic algorithms we developed are based on Greedy and Tabu-search [3]. We describe each one in turn.

1) Greedy 1: maximum performance first: At each time slot t, Greedy 1 tries to turn on the virtual sensor v_k with the maximum utility τ_k, while meeting the energy constraints of all the nodes associated with v_k.

2) Greedy 2: Maximum performance per unit energy first: At each time slot t, Greedy 2 tries to turn on the virtual sensor v_k with the maximum ratio of utility to energy-cost (i.e., $\frac{\tau_k}{E_k}$), while meeting the energy constraints of all nodes associated with v_k.

3) Tabu search: The Tabu search heuristic is a variant of the Local search [3]. We tested a number of different ways of defining local virtual sensors, and our experience shows that the following works best: during a local search, Tabu search varies the on/off schedules of the virtual sensors for any two time slots.

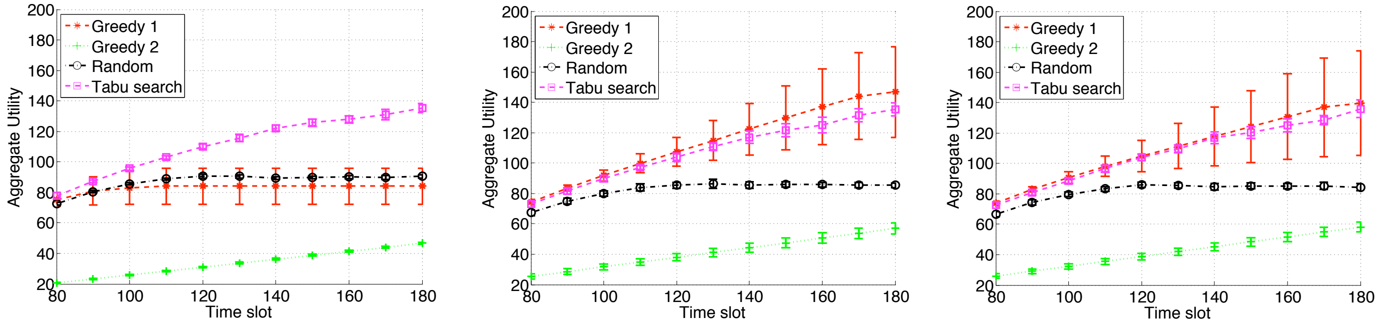
IV. Evaluation

We developed a WSN simulation to evaluate our optimization framework. In this simulation, we used $N = 3$ nodes and $M = 9$ physical sensors. The initial energy of each node $\epsilon = 5,000J$. The energy consumption figures in this table were measured on our wireless sensor platform, which includes that of a low power micro-controller (used to control the PIR sensors) and a high power DSP (used to control the audio and image sensors) as they need to be turned on to support sensing, local data processing, and fusion tasks The PIR sensor, Audio sensor and image sensor consume 1.78, 50 and 59 joules energy per time slot respectively. The simulation returns a utility for every possible virtual sensor given a Poisson distributed set of events.

We generated the utilities τ for each sensor. As noted in section III-C, the specific function we used in this case was $1 - p_e$, where p_e is the probability of error. Specific values of P_0 and P_1 were randomly chosen and used to generate the events. As noted earlier we generated events with a Poisson distribution. We carried out 10 trials of the simulation with two separate event types per trial: event A and event B. The only difference between these event types was their arrival rates, each of which was randomly generated for each trial. We then considered 5 cases per trial. The differences in each case was to do with the specified user event of interest: Case 1 - event A AND event B; Case 2 - event A; Case 3 - event B; Case 4 - event A AND event B; Case 5 (sensors were homogeneous). For the benefit of space, we didn't show the results of Case 4 and 5 because they showed similar trends as the other cases.

We calculated the mean and standard deviation over the 10 trials of the aggregate utility of the sensed data. We determined the optimal fusion weights and thresholds for every possible sensor combination (which for 9 sensors amounted to $2^9 = 512$ combinations). The sensors' outputs were combined using

[1] http://en.wikipedia.org/wiki/CPLEX, accessed 6th May, 2012.



(a) Case 1 (b) Case 2 (c) Case 3

Fig. 2. Network utilities versus the number of time slots in a large scale network.

these weights and the output of the fusion function was then compared with the ground truth in order to calculate $1 - p_e$.

Figures 2(a) to 2(c) show the network utility results versus the number of time slots (T) in three different cases. Clearly, the network utility increases as the number of time slots increase in all the cases. However, the figures also show that the network utility can be improved by choosing a better virtual sensor schedule. For example, in Fig. 2(b), when the number of time slots is 180, the mean network utility increases from approximately 82 to 143, an increase of 72%, by using the schedule generated by Greedy 1 instead of the naive random schedule.

It is interesting to note that the Tabu search performs very close to or the same as the best results in all figures; in particular, Figure 2(a) shows that the Tabu search produces a result that is approximately 70% better than the second best result when the number of time slots is 180, which justifies the introduction of this meta-heuristic.

Overall, heuristic solutions, e.g., tabu-search and Greedy 1, produced significantly better results compared with their naive random counterpart. The performance of Tabu-search is more stable (the results have less variance) than that of Greedy 1.

V. Related work

Although the authors in [7] discuss different fusion rules, they don't provide a generic method for choosing one. The past two decades however, have seen numerous research studies both in multi-sensor data fusion [8] and distributed detection [9]. We make use of the optimal fusion rule when creating virtual sensors [8].

Assuming a fusion strategy to create a virtual sensor has been chosen, the next question we are interested in is how to schedule the sensors. Optimal sensor scheduling algorithms to maximize network utilities have been investigated in the literature [10].The optimal network coverage problem in rechargeable sensor networks has been studied in [11], and the optimal sensor schedule problem has been considered in [12] in sensor networks powered by batteries. However, they treat each sensor individually and don't consider inter-sensor collaborations, not to mention inter-node, data fusion in the problem formulation. To the best of our knowledge, this is the first paper which investigates the optimal sensor schedule problem with inter-sensor and inter-node data fusion.

VI. Conclusion

We investigated the problem of sensor scheduling in heterogenous WSNs. We represented a group of sensors in a WSN as a virtual sensor, studied the utilities of each virtual sensor using an optimal data fusion algorithm, and attempted to optimize the utilities of the WSN based on those of the virtual sensors. We modeled the maximum performance in the WSN problem as an ILP problem, and implemented three heuristic algorithms based on Greedy and Tabu-search to solve the optimization problem efficiently. We evaluated the performance of virtual sensor scheduling algorithms by extensive simulations.

References

[1] T. Wark, W. Hu, P. Corke, J. Hodge, A. Keto, B. Mackey, G. Foley, P. Sikka, and M. Bruenig, "Springbrook: Challenges in developing a long-term, rainforest wireless sensor network," in *Intelligent Sensors, Sensor Networks and Information Processing. ISSNIP 2008. International Conference on*, pp. 599–604.

[2] Varshney, "Multisensor data fusion," *Electronics & Communication Engineering Journal*, vol. 9, no. 6, pp. 245 – 253, 1997.

[3] F. Glover and M. Laguna, *Tabu Search*. Norwell, MA, USA: Kluwer Academic Publishers, 1997.

[4] D. K. Dawson and M. G. Efford, "Bird population density estimated from acoustic signals," *J Appl Ecol*, vol. 46, no. 6, pp. 1201–1209, Jan 2009.

[5] M. Afanasyev, D. O'Rourke, B. Kusy, and W. Hu, "Heterogeneous traffic performance comparison for 6lowpan enabled low-power transceivers," in *Proceedings of the 6th Workshop on Hot Topics in Embedded Networked Sensors*, ser. HotEmNets '10, 2010.

[6] P. K. Atrey, M. S. Kankanhalli, and R. Jain, "Information assimilation framework for event detection in multimedia surveillance systems," *Multimedia Systems*, vol. 12, no. 3, pp. 239–253, Jan 2006.

[7] W. Wang, V. Srinivasan, K.-C. Chua, and B. Wang, "Energy-efficient coverage for target detection in wireless sensor networks," *IPSN '07: Proceedings of the 6th international conference on Information processing in sensor networks*, Apr 2007.

[8] P. K. Varshney, *Distributed Detection and Data Fusion*. Spinger-Verlag, 1997.

[9] R. Tenney and N. Sandell, "Detection with distributed sensors," *IEEE Transactions on Aerospace ...*, Jan 1981. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4102537

[10] P. Vicaire, T. He, Q. Cao, T. Yan, G. Zhou, L. Gu, L. Luo, R. Stoleru, J. A. Stankovic, and T. F. Abdelzaher, "Achieving long-term surveillance in vigilnet," *ACM Trans. Sen. Netw.*, vol. 5, no. 1, pp. 1–39, 2009.

[11] W. Hu, N. Bulusu, C. T. Chou, S. Jha, A. Taylor, and V. N. Tran, "Design and evaluation of a hybrid sensor network for cane toad monitoring," *ACM Trans. Sen. Netw.*, vol. 5, no. 1, pp. 1–28, 2009.

[12] Y. Gu, J. Hwang, T. He, and D. H.-C. Du, "usense: A unified asymmetric sensing coverage architecture for wireless sensor networks," in *ICDCS '07: Proceedings of the 27th International Conference on Distributed Computing Systems*, 2007.