

A Distributed, Energy-Efficient and QoI-Aware Framework for In-Network Processing

Sepideh Nazemi[†], Kin K. Leung[†] and Ananthram Swami[‡]

[†]Imperial College, London, UK. Email: {s.nazemi10, kin.leung}@imperial.ac.uk

[‡]Army Research Laboratory, Adelphi, MD, USA. Email: ananthram.swami.civ@mail.mil

Abstract—In-network processing (INP) is a promising method that allows aggregation of data while it is being transferred along the communication paths as a means to optimize the utilization of network resources without violating the quality of information (QoI) requirements. Given the large amount of data existing in dynamic environments, the optimization of INP requires a distributed framework that can adapt easily to network changes and user requirements. In this work, we develop the principle for designing a distributed mechanism in order to determine and control INP. Specifically, the proposed framework can decide, in a distributed way, which nodes along the communication paths optimally perform INP, with consideration of operational energy consumption and QoI requirements for achieving global optimal INP. The significance of the proposed distributed method is that it requires each node to make independent decisions locally for data aggregations, thus naturally enhance robustness and efficiency against network and data load dynamics. Extensive numerical results are presented to confirm the validity of the proposed approach.

I. INTRODUCTION

Large information networks such as wireless sensor networks (WSN) are increasingly employed in a variety of applications, ranging from healthcare and education to military operations. The development of complex network technology, the growth of smart environments, and the combination of communication devices equipped with various sensing capabilities generate huge amount of data. These data must be transferred and processed to create useful information for either single or multiple end-users often via multi-hop transmissions. However, practical constraints on the network, such as communication range, bandwidth limitations, and dynamic environments cause the transferring of the huge volume of data generated from many sources to be very challenging, if not impossible. However, In-Network Processing (INP) has opened a new door to possible solutions for optimizing the utilization of network resources. INP methods primarily aim to aggregate (e.g., compression, fusion and averaging) data from different sources with the objective of reducing the data volume for further transfer, thus reducing energy consumption and increasing the network lifetime [1].

Data aggregation approaches may be classified as lossy and lossless. For lossy data aggregation, packets with data coming from different sources are aggregated into a single packet by intermediate nodes along the communication path. However, the original data cannot be fully recovered from the aggregated packet in contrast to that of lossless data aggregation which it is possible to fully extract the original data from aggregated

data packet. In this work, we do not focus on the data aggregation function. Instead, we assume that aggregation process is performed by a known aggregation function in the network.

When INP is applied in an information network, it is crucial to consider how such data processing affects the quality of information (QoI) at the receiving end, which is represented by multi-dimensional metrics [2], e.g., information accuracy, completeness, reliability and timeliness. How INP should be carried out for satisfactory QoI at the user level remains an open research issue. For example, [3]–[6] investigate the QoI and introduce various models to maximize a utility function of QoI. [7] introduce a distributed model to satisfy the user QoI requirement. Nichols [8] introduces a centralized optimization model utilising the QoI notion to prioritize traffic in directional network. However, they do not consider INP or the impact of further INP process on aggregated information.

Eswaran et al. [9] apply the network utility maximum (NUM) framework to determine the optimal compression and fusion factors for INP and which nodes in a path serve as the optimal places for performing the data processing. However, they do not include realistic QoI attributes in their work, as they assume a utility function to represent all of the attributes. Chen et al. [10] explore joint scheduling and data aggregation by assuming that full aggregation is always possible with an additive utility. Silberstein et al. [11] present a fully distributed method for implementing many-to-many aggregation in a sensor network that minimizes the communication cost by optimally balancing a combination of multicast and in-network aggregation. However, they only consider binary aggregation options, i.e., no aggregation and full aggregation which may deteriorate the QoI at the end user. Both [10], [11] use integer programming techniques for making data aggregation decisions. However, they do not consider any QoI metrics.

Considering dynamic approaches, Kennedy et al. [12] focus on the data itself for designing a dynamic in-network data aggregation protocol, while Jiao et al. [13] propose their full data aggregation protocol based on dynamic routing, in which data packets are sent to the nodes with the longest data aggregation queue. However, again they did not take into account any QoI metrics. Anandkumar et al. [14] take a stochastic approach to model an energy-efficient distributed inference in a random fusion network. They analyse the spatial correlation among nodes presented by a dependency graph. However, the inference policy in [14] still needs the centralized network information and they do not consider any QoI metric.

Our work consists of two technical components: QoI and INP. Although many studies have been conducted on both of these technical issues, a gap still exists in the research regarding how to dynamically control INP to satisfy QoI requirements specified by the end users. For this reason, the

Research was partially sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. Furthermore, the authors would like to thank Dr. Stuart Farquhar of DSTL for many helpful discussions and comments.

objective of this paper is to introduce a distributed framework for INP where intermediate nodes aggregate data dynamically for energy efficiency, while satisfying user QoI requirements.

Providing a realistic QoI function and designing a method such that end user can specify a required QoI and retaining the QoI in a satisfactory level while processing and transferring them are very challenging especially in a high dynamic and complex network such as WSN. In such dynamic environment, fully distributed methods are very desirable due to the ability to readily adapt to the network changes and constraints and deal with huge amount of data generated in this environment. In sharp contrast to [3]–[14], we concentrate in this work on designing a fully distributed framework to facilitate controlling of INP process at intermediate nodes while satisfying the end user requirements. For this purpose, as a starting point, we consider the amount of information that user needs to receive as a QoI parameter. Note that required amount of data can present the accuracy metric of QoI [3]. In addition, a realistic objective function, the total energy consumption of the network is taken into account. We prove that under a set of sufficient assumptions, the INP process can be fully distributed while performing closely to the optimality.

The rest of the paper is organized as follows. We formulate the problem in Section II. In Section III, the proposed distributed framework is presented and the evaluation results and conclusions are presented in Sections IV and V, respectively.

II. PROBLEM FORMULATION

A. Scenario, QoI metrics and Assumptions

As an example, consider a WSN monitoring the level of chemical contamination in an area. A user approaches the area and wants to know the level of chemical contamination of a specific part of the area. The user sends a query to the network specifying his area of interest and the amount of information that user needs to receive (e.g., number of data packets) as a QoI requirement threshold.

We acknowledge that beside the required amount of information, other parameters of QoI may also be important in this scenario (e.g., timeliness). The significance of this paper is that we propose the first efficient approach for designing a fully distributed and QoI-aware framework in the field of INP. The problem with more complicated parameter settings and assumptions will be left to our future work.

We assume that a data aggregation tree (of nodes over which the requested information is processed and routed toward the end user) is formed after the user queries the network. Furthermore, we assume that the total energy consumption of a node consists of the energy spent in receiving p_R , computing p_C and transmitting p_T its data. Among these operations, data transmission typically uses more energy than the others [9]. Data generated in the information networks has a large amount of redundancy due to the spatial and temporal correlation among sensors. Therefore, it is possible to aggregate (e.g., compress, fuse, average) data as a means of reducing energy consumption for transmission and reception, without sacrificing QoI for the end users. On the other hand, the more the data are aggregated, the higher the computational energy cost is. Therefore, an energy trade-off exists among the energy that each node spends for data reception, transmission and computation.

Let the data processing algorithm in use aggregates received

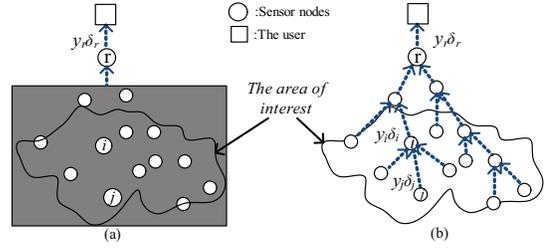


Fig. 1. (a) INP architecture; (b) Aggregation tree formed in responding to the query by the end user.

data into a smaller volume of data for transmission at each node. Therefore, a key for the energy trade-off is the data reduction rate (denoted by δ between 0 and 1), which is the ratio of the amount of aggregated data to that of all input data at a node. The optimal energy trade-off is determined by choosing the optimal reduction rate at each node.

B. A global optimization problem

We formulate the problem of finding the optimal data reduction rates at all nodes as an optimization problem. Specifically, the optimal reduction rate at each node minimizes the total energy consumption in the whole network, while satisfying the user QoI requirements. Our goal is to choose the data reduction rates at all nodes involved in order to minimize the total energy consumption, while ensuring that the amount of aggregated data for the end user exceeds a specified QoI threshold γ . To illustrate, Fig. 1 (a) sketches the INP architecture where the gray box denotes the network of interest, and Fig. 1 (b) presents the corresponding aggregation tree along which the data from source nodes are routed and processed toward node r and finally to the end user. The optimisation problem is as follows:

$$\begin{aligned} \min_{\{\delta_i\}} \quad & \sum_{i=1}^n p_i(\delta_i, y_i), \\ \text{s.t.} \quad & \delta_r y_r \geq \gamma \end{aligned} \quad (1)$$

where n is the total number of nodes in the network aggregation tree and $p_i(\delta_i, y_i)$ is the total energy consumption which is a function of the volume of input data y_i and the data reduction rate δ_i at node i . Since node r is responsible for delivering the required information to the user, the constraint $y_r \delta_r \geq \gamma$ specifies the minimum volume of aggregated data that the user requires from all source nodes in the area of interest, shown in Fig. 1 (a).

We refer to (1) as a global optimization (GO) problem where each node in the network has the optimal data reduction rate in order to minimize the total energy consumption while satisfying the users QoI requirement. Even though the problem has only a single constraint, data reduction rate must be chosen optimally at every node so that the constraint for end user can be satisfied.

The GO problem in (1) is a variant of 0/1 knapsack problem [15]. We investigate on the hardness of the GO problem under two different scenarios. In the first scenario, we assume that computation cost at each node is far more expensive than its transmission cost and in the second scenario, the transmission cost is considered to be far more expensive than the computation cost at each node.

Under the first scenario, the GO problem is NP-hard, since 0/1 knapsack problem can be reduced to the GO problem as follows. Assume a single level aggregation tree rooted at node r as an instance of the GO problem (see Fig. 2). The total

energy consumption of node i consists of the energy spending in receiving p_R , computing p_C and transmitting p_T of its data. Without loss of generality, assume $p_R = 0$. In addition, if node i chooses to forward all ($\delta_i = 1$) or none ($\delta_i = 0$) of its data, the computation cost at node i equals zero. Under the first scenario, it is beneficial for node i to forward all or none of its data instead of aggregating them. That is, node i only has two choices of sending all ($\delta_i = 1$) or none ($\delta_i = 0$) of its data; Therefore, the only effective energy consumption cost will be transmission cost. Since node r is responsible for delivering information received from its children (i.e., $y_r = \sum_{i \in C_r} \delta_i y_i$, where C_r is a set of node r 's children) to the end user, its reduction rate must be equal to one ($\delta_r = 1$), otherwise the constraint in the GO problem will not be satisfied. The transmission energy cost is relative to the amount of information that node i sends (e.g., $p_T = \alpha_i y_i \delta_i$). Therefore, the problem is to find a set of $\delta_i \in \{0, 1\}$ such that $\sum_{i \in C_r} \alpha_i y_i \delta_i$ attains its minimum value, while $\sum_{i \in C_r} y_i \delta_i \geq \gamma$, where $\alpha_i y_i$ and y_i are value and weight of item i respectively. Therefore, if there exists a deterministic algorithm that can solve the instance of the GO problem in polynomial time, then 0/1 knapsack problem is also solvable, contradicting the prior-knowledge of 0/1 knapsack problem being NP-hard.

In the second scenario, given that transmission cost is far more expensive than computation cost, the GO problem converts to the fractional knapsack problem. However, although the greedy approach [15] or dynamic programming method can be applied to approximate the global solution for a single level data aggregation tree, the interdependency among nodes when a multi-level data aggregation tree is considered impedes use of these techniques in our problem. In general, the GO problem is an NP-hard problem, since the probability of having the first scenario or multi-level data aggregation tree is not zero.

Therefore, intractability and interdependency among nodes motivate the need for a distributed method where no centralized operator is required and complex network structures, such as multi-level data aggregation tree shown in Fig. 1 (b) can be handled easily. To this end, we will propose our fully distributed framework, a solution to the GO problem. Proposed solution performs closely to the optimality with $O(n)$ complexity which not only provides solution to the intractable GO problem where the first scenario and multi-level data aggregation tree structures are considered, but also improves on the complexity of well-known greedy technique complexity $O(n \log n)$.

III. DISTRIBUTED, ENERGY-EFFICIENT AND QOI-AWARE APPROACH

Given intractability of the GO problem, high interdependency among nodes and the desirability of fully distributed solutions for our applications, we introduce our model the Local Constrained Optimization framework (to be referred to as LCO model) as follows:

$$\begin{aligned} & \min_{\{\delta_i\}} \sum_{i=1}^n p_i(\delta_i, y_i) \\ & \text{s.t.} \\ & \delta_i y_i \geq \gamma \\ & y_i = \sum_{j \in C_i} \delta_j y_j, \text{ for } i = 1, \dots, n \end{aligned} \quad (2)$$

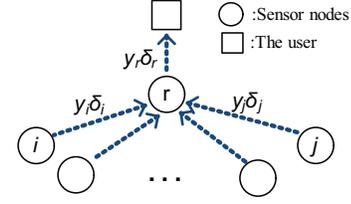


Fig. 2. An instance of the GO problem.

The significance of introducing (2) is explained as follows. It turns out that under a reasonable set of assumptions, the order of the minimization and summation operators in (2) can be switched, while yielding the same optimal solution as follows.

$$\begin{aligned} & \min_{\{\delta_i\}} \sum_{i=1}^n p_i(\delta_i, y_i) \\ & \text{s.t.} \\ & \delta_i y_i \geq \gamma \\ & y_i = \sum_{j \in C_i} \delta_j y_j, \text{ for } i = 1, \dots, n \end{aligned} \quad (3)$$

We refer to (3) as Constrained Optimization (CO) problem. While (3) and (1) show a global optimization model, the difference between the CO problem (3) and the GO problem (1) is that in the CO problem, each node has its own QoI constraint, while the GO problem has only one constraint for the root node r .

It is important to note under the following set of assumptions the LCO model can provide exact solution to the CO problem. That is, the total energy consumption in the network can be minimized by selecting the optimal data reduction rate at each node based on local information at each node, thus introducing a fully distributed optimization system! We shall develop conditions (assumptions) under which the LCO leads to such a global optimal solution as follows.

Let p_i denotes the total energy consumption of node i as

$$p_i = p_{iR} + p_{iC} + p_{iT}, \quad (4)$$

where p_{iR} , p_{iC} and p_{iT} denote the energy spent in receiving, aggregating and transmitting data by node i , respectively.

p_{iR} is a function of the amount of data y_i that node i receives from its children nodes. let e_R be the energy consumed in receiving one unit of data. So, we assume

$$p_{iR} = f(y_i) = e_R y_i. \quad (5)$$

Consider that data aggregation at node i yields a data reduction rate of δ_i . Given the amount of received data y_i at node i , the energy consumption for transmission p_{iT} at the node i is assumed to be

$$p_{iT} = g(y_i, \delta_i) = e_T y_i \delta_i, \quad (6)$$

where e_T is the transmission energy consumption per unit of data.

Typically, the greater degree of data aggregation (i.e., the smaller amount of aggregated data), the higher the energy consumption for computation at the node. Therefore, p_{iC} is assumed to be a function of δ_i and y_i . Furthermore, the influence of the reduction rate δ_i on p_{iC} is highly dependent on the type of data being aggregated, the aggregation functions and the characteristics of computation hardware. For instance, processing high-quality video consumes more energy than processing data such as temperature or humidity values. To adequately capture the dependence, the energy consumption for computation at node i is assumed to be

$$p_{iC} = k(y_i, \delta_i) = e_C y_i q_i(\delta_i), \quad (7)$$

where e_C is the energy consumption for processing one unit of data, y_i is the amount of received data and $q_i(\delta_i)$ is a scaling function for energy consumption of computation which is a decreasing differentiable function of the reduction rate δ_i . For example, $q_i(\delta_i) = (\frac{1}{\delta_i} - 1)$ for $\delta_i > 0$.

Furthermore, as shown in the data aggregation tree in Fig.1 (b), the amount of data that node i receives is

$$y_i = \sum_{j \in C_i} y_j \delta_j, \quad (8)$$

where C_i denotes the set of children nodes of node i .

There may be a concern that the linear model in (5) to (7) is unable to adequately adjust to all the characteristics of communication and computation in the network (e.g., coding and processing functions); however, as a general assumption in [9], [16] we assume a linear model here. We will investigate non-linear cost models in our future work.

With these assumptions, we present the following theorem which introduces sufficient conditions under which the order of the minimization and summation operators in (3) can be switched while preserving optimality.

Theorem. *Given the energy consumption represented by (5) to (7) and the amount of received data in (8) for each node in a single level data aggregation tree, the LCO model (2) is equivalent to the CO (3) as follows:*

$$\begin{aligned} \min_{\{\delta_i\}} \sum_{i=1}^n p_i(\delta_i, y_i) &= \sum_{i=1}^n \min_{\delta_i} p_i(\delta_i, y_i) \\ \text{s.t. } \delta_i y_i &\geq \gamma & \text{s.t. } \delta_i y_i &\geq \gamma \\ y_i &= \sum_{j \in C_i} \delta_j y_j & y_i &= \sum_{j \in C_i} \delta_j y_j \\ \text{for } i &= 1, \dots, n & \text{for } i &= 1, \dots, n \end{aligned} \quad (9)$$

Proof. Duo to space limitation we develop our proof for a single child. The same argument will be applied for multiple children case. Considering the the LHS of (9), we have

$$\begin{aligned} \min_{\{\delta_i\}} \sum_{i=k}^{k+1} p_i(\delta_i, y_i) \\ \text{s.t. } \delta_k y_k &\geq \gamma \\ \delta_{k+1} y_{k+1} &\geq \gamma \\ y_k &= \delta_{k+1} y_{k+1} \end{aligned} \quad (10)$$

Substituting (4) to (7) into (10) and forming the Lagrangian give the following:

$$\begin{aligned} L = & y_k(e_{R_k} + e_{C_k} q_k(\delta_k) + e_{T_k} \delta_k - \mu_k \delta_k) + \mu_k \gamma \\ & + y_{k+1}(e_{R_{k+1}} + e_{C_{k+1}} q_{k+1}(\delta_{k+1}) + e_{T_{k+1}} \delta_{k+1} \\ & - \mu_{k+1} \delta_{k+1}) + \mu_{k+1} \gamma + \lambda(y_{k+1} \delta_{k+1} - y_k). \end{aligned} \quad (11)$$

Considering λ as a complicating variable, (11) can be reorganized as follows:

$$\begin{aligned} L = & \left(y_k(e_{R_k} + e_{C_k} q_k(l_k) + e_{T_k} \delta_k - \mu_k \delta_k) + \mu_k \gamma - \lambda y_k \right) \\ & + \left(y_{k+1}(e_{R_{k+1}} + e_{C_{k+1}} q_{k+1}(\delta_{k+1}) + e_{T_{k+1}} \delta_{k+1} \right. \\ & \left. - \mu_{k+1} \delta_{k+1}) + \mu_{k+1} \gamma + \lambda y_{k+1} \delta_{k+1} \right). \end{aligned} \quad (12)$$

If we fix the λ , (12) becomes decomposable in δ_k, μ_k and δ_{k+1}, μ_{k+1} , i.e., $L = f_k(\delta_k, \mu_k, \lambda) + f_{k+1}(\delta_{k+1}, \mu_{k+1}, \lambda)$. Therefore, we have

$$\begin{aligned} \min_{\delta_k, \delta_{k+1}, \mu_k, \mu_{k+1}} L &= \min_{\delta_k, \mu_k} f_k(\delta_k, \mu_k, \lambda) \\ &+ \min_{\delta_{k+1}, \mu_{k+1}} f_{k+1}(\delta_{k+1}, \mu_{k+1}, \lambda). \end{aligned} \quad (13)$$

As a result we have

$$L(\lambda) = g_k(\lambda) + g_{k+1}(\lambda), \quad (14)$$

where

$$\begin{aligned} g_k(\lambda) = & \min_{\delta_k, \mu_k} \left(y_k(e_{R_k} + e_{C_k} q_k(\delta_k) + e_{T_k} \delta_k - \mu_k \delta_k) \right. \\ & \left. + \mu_k \gamma + \lambda y_k \right). \end{aligned}$$

$$\begin{aligned} g_{k+1}(\lambda) = & \min_{\delta_{k+1}, \mu_{k+1}} \left(y_{k+1}(e_{R_{k+1}} + e_{C_{k+1}} q_{k+1}(\delta_{k+1}) \right. \\ & \left. + e_{T_{k+1}} \delta_{k+1} - \mu_{k+1} \delta_{k+1}) + \mu_{k+1} \gamma + \lambda y_{k+1} \delta_{k+1} \right). \end{aligned}$$

In order to calculate λ , we need to form the dual problem as follows.

$$\max_{\lambda} (g_k(\lambda) + g_{k+1}(\lambda)). \quad (15)$$

We can use the sub-gradient method to calculate the optimal λ . In this method, the dual variable λ is updated by the following iteration step:

$$\lambda_{m+1} = \lambda_m + a_m (\nabla g_k(\lambda) + \nabla g_{k+1}(\lambda)), \quad (16)$$

where a_m is the step size and m indicates the number of iterations.

The sub-gradient of g_k at λ is $-y_k$, and the sub-gradient of g_{k+1} at λ is $+y_{k+1} \delta_{k+1}$. Substituting the sub-gradients into (16) yields $\lambda_{m+1} = \lambda_m + a_m (y_{k+1} \delta_{k+1} - y_k)$. However, $\lambda_{m+1} = \lambda_m$ due to the tree structure and the assumption introduced by (8). That means λ never changes in the iteration function and initial λ_0 is the optimal λ for dual problem. Therefore, it can be treated as a constant. As a result, the minimum values and parameters of g_k and g_{k+1} are not effected by λ and therefore, g_k and g_{k+1} can be calculated separately. Therefore,

$$\min_{\{\delta_i\}} \sum_{i=k}^{k+1} p_i(\delta_i, y_i) = \min_{\delta_k} p_k(\delta_k, y_k) + \min_{\delta_{k+1}} p_{k+1}(\delta_{k+1}, y_{k+1}). \quad (17)$$

■

As a result of the theorem, the optimal solution for (3) under assumptions (4) to (8) can be obtained in a distributed manner with linear time complexity where each node selects its data reduction rate to minimize its energy consumption based on local information forming a fully distributed optimization solution.

The importance of introducing the LCO model is that it can provide very close approximate solution to the GO problem. Although imposing the individual constraint at each node may lead to consuming more energy in the network, under set of practical network settings and assumptions the LCO model performs very close to the optimality. In the next section, we evaluate and verify the performance of the proposed framework, LCO, under different parameter settings.

IV. EVALUATIONS

In this section, we present results for evaluating distributed LCO modelled in (2). We assume that the full binary aggregation tree rooted at node 1 is formed for the request of information by a user. The full binary tree is considered for the purpose of model verification. However, the LCO model can be applied to any K-ary aggregation tree. The function $q_i(\delta_i) = \frac{1}{\delta_i} - 1$, $\delta_i > 0$ is considered to reflect the impact of the data aggregation process on energy consumption. The packet size and the data transmission rate are set 128 bytes and 250 kbps, respectively. We consider two different settings. In the first setting, Case 1, the energy consumption for data reception, transmission and computation are assumed to be equal (i.e., $e_R = e_T = e_C = 0.00024J$) and in the second setting, Case 2, $e_C = \frac{1}{2} e_T$ (i.e., $e_R = e_T = 0.00024J$, $e_C = 0.00012J$). QoI threshold γ is assumed to be 5 data packets for both cases. We compare the LCO results to the optimal value of global optimization (GO) in (1).

Fig. 3 reports the total energy consumption of the network in these two cases. Imposing individual constraints on each node and forcing them to satisfy the user QoI threshold leads to a gap between energy consumption of the GO and LCO models.

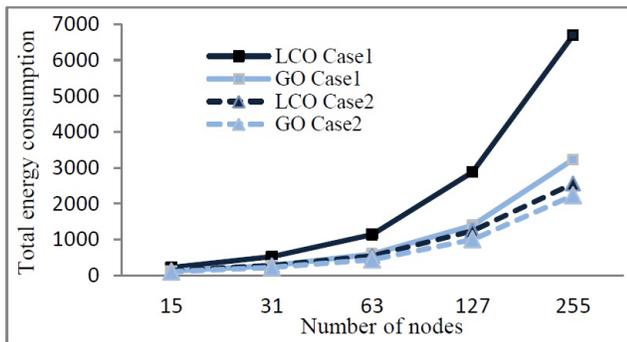


Fig. 3. Impact of computation and transmission cost on total energy consumption.

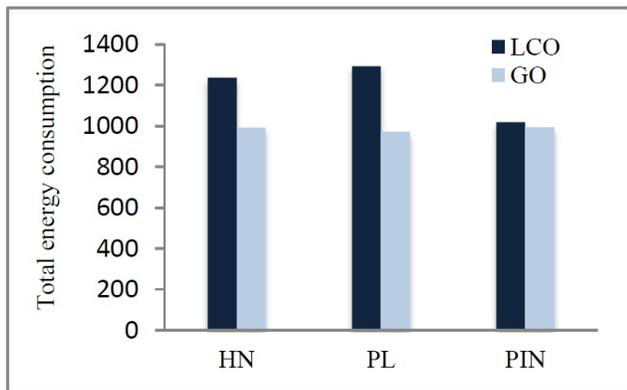


Fig. 4. Total energy consumption comparison of 127 nodes under three different settings.

In Case 1 which is an extreme worst case, nodes under the LCO model choose to send almost all of their data due to the expensive computation cost (i.e., select a large data reduction rate), and thus all of them have to transmit huge volume of data. However, in the GO model, due to the centralized and global decision, each node achieves the optimal reduction rate and consumes a less amount of energy than the LCO model. In Case 2, in spite of forcing individual constraints on each node, the LCO model performs very close to the GO value. The reason for this exciting result is, due to less expensive computation cost, each node aggregates more data and as a result, consumes less energy for data transmission.

Remark: In contrast to Case 1, parameters applied to Case 2 represent a practical network setting, i.e., computation cost is generally less expensive than transmission cost. Therefore, under a general network setting presented by Case 2, the LCO performs very close to the global optimal value.

Moreover, we examine the heterogeneous case where nodes are not identical in the LCO model. We consider the setting of ($e_T = e_R = 0.00024$) for all nodes and change the e_C parameter for Intermediate Nodes (INs) and Leaf Nodes (LNs). In Fig. 4, the bar labelled PIN (Powerful Intermediate Node) shows the result when e_C of INs are less expensive than that of LNs. It can be clearly seen that the LCO performs very close to the optimality provided by the GO. This is because, INs can gather more data and aggressively reduce data due to the less expensive computational cost. The bar labelled PL (Powerful Leaves) presents the result related to the inverse setting of PIN (i.e., e_C of LNs are less expensive than that of INs). The HN

(Homogeneous Node) bar shows the result of having homogeneous nodes under parameter setting of Case 2 experiment.

As the results illustrate, the LCO model shows a promising performance relative to the GO value. We also analysed the performance bound of the LCO. However, due to the space limitation, we omit this result here.

V. CONCLUSION

We formulated the problem of in-network processing (INP) as a non-linear optimization problem in order to minimize the total energy consumption by all associated nodes while satisfying QoI requirements. Since this problem cannot be solved efficiently by existing approaches due to intractability and high interdependency among nodes, these inspired a fully distributed system for our applications, for which we introduced a distributed LCO framework, a Local Constrained Optimization solution. Under a set of reasonable assumptions, we have proven that the optimal data reduction rate can be determined by each node based on local information in a fully distributed manner. Computer simulations have been conducted to illustrate the effectiveness of the LCO framework. Numerical results show that the new framework, LCO, can perform very close to the global optimum for parameter settings where communication costs more than computation.

We plan to extend the framework to consider information requests from multiple users as well as other QoI parameters.

REFERENCES

- [1] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "In-network aggregation techniques for wireless sensor networks: a survey," *IEEE Wireless Communications*, 2007.
- [2] C. Bisdikian, L. M. Kaplan, M. B. Srivastava, D. J. Thornley, D. Verma, and R. I. Young, "Building principles for a quality of information specification for sensor information," in *FUSION*, 2009.
- [3] E. Ciftcioglu, A. Yener, R. Govindan, and K. Psounis, "Operational information content sum capacity: Formulation and examples," in *FUSION*, 2011.
- [4] E. N. Ciftcioglu and A. Yener, "Quality-of-information aware transmission policies with time-varying links," in *MILCOM*, 2011.
- [5] E. N. Ciftcioglu, A. Yener, and M. J. Neely, "Maximizing quality of information from multiple sensor devices: The exploration vs exploitation tradeoff," 2013.
- [6] R. Uргаonkar, E. N. Ciftcioglu, A. Yener, and M. J. Neely, "Quality of information aware scheduling in task processing networks," in *WiOpt*, 2011.
- [7] V. Sachidananda, A. Khelil, D. Umap, M. Majuntke, and N. Suri, "Trading transport timeliness and reliability for efficiency in wireless sensor networks," in *ICNSC*, 2013.
- [8] R. A. Nichols, "Qoi in dtm-based directional networks," in *2009. SARNOFF'09*, 2009.
- [9] S. Eswaran, J. Edwards, A. Misra, and T. F. LaPorta, "Adaptive in-network processing for bandwidth and energy constrained mission-oriented multihop wireless networks," *IEEE Transactions on Mobile Computing*, 2012.
- [10] F. Chen, M. P. Johnson, A. Bar-Noy, and T. F. L. Porta, "Convergecast with aggregatable data classes," in *SECON*, 2012.
- [11] A. Silberstein and J. Yang, "Many-to-many aggregation for sensor networks," in *ICDE*, 2007.
- [12] O. Kennedy, C. Koch, and A. Demers, "Dynamic approaches to in-network aggregation," in *ICDE*, 2009.
- [13] Z. Jiao, R. Fengyuan, H. Tao, and L. Chuang, "Data aggregation protocol based on dynamic routing in wireless sensor networks," in *CMC*, 2009.
- [14] A. Anandkumar, A. Swami, J. E. Yukich, and L. Tong, "Energy scaling laws for distributed inference in random fusion networks," *IEEE Journal on Selected Areas in Communications*, no. 7, 2009.
- [15] M. M. Guntzer and D. Jungnickel, "Approximate minimization algorithms for the 0/1 knapsack and subset-sum problem," *Operations Research Letters*, 2000.
- [16] L. M. Feeney and M. Nilsson, "Investigating the energy consumption of a wireless network interface in an ad hoc networking environment," in *INFOCOM*, 2001.