A Centralized Energy-Efficient Model for Increasing the Information Gathered in Wireless Sensor Networks

By

Sushyanth Pamulaparthy

Bachelor of Technology in Information Technology
Jawaharlal Nehru Technological University
Hyderabad, Telangana, India
2013

Submitted to the Faculty of the
Graduate College of
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
Master of Science
July, 2015

A CENTRALIZED ENERGY-EFFICIENT MODEL FOR INCREASING
THE INFORMATION GATHERED IN WIRELESS SENSOR NETWORKS

Thesis Approved:

Dr. Christopher Crick

Thesis Adviser

Dr. Eric Chan-Tin

Dr. Johnson Thomas

ACKNOWLEDGMENTS

'Life may be tough, but I've got a God that's tougher'. Firstly, I would like to thank God for all his blessings, helping me complete my thesis. I would like to express my sincerest gratitude towards my adviser, Dr. Christopher Crick, who has supported me throughout my thesis with his creative ideas and invaluable suggestions. I could not have asked for a better adviser and certainly would not have completed the thesis without his time, support and guidance.

I would like to take this opportunity to also thank my parents, brother, sister-in-law and sisters for the suggestions, support and attention they have provided me throughout my life. Last but not the least, I thank my friends for their unwavering support.

Name: Sushyanth Pamulaparthy

Date of Degree: July, 2015

Title of Study: A Centralized Energy-Efficient Model for Increasing the Information Gathered in Wireless Sensor Networks

Major Field: Computer Science

A Wireless Sensor Network (WSN) consists of numerous sensor nodes spread over a wide area to gather information and transmit it to a sink node, which then sends it to the end user for analysis. Such networks have a wide range of applications in areas like health, military, security, wildlife monitoring, etc. These sensors gather huge amounts of data, not all of which is important to the end user. Only a few sensors at any given point of time have valuable information. Identifying such sensors consistently can help us increase the Value of Information (VoI) of a system. We know that sensors generally have limited resources in terms of memory capacity, power supply and communication bandwidth. Hence, it is important to take energy-efficiency into consideration while implementing an approach.

In order to address the above problems, we propose a centralized network model which makes use of Information Entropy to determine the theoretical upper bound on the VoI available in a network. We also provide a probabilistic sensor selection algorithm to consistently select the most informative sensor, enabling the model to increase the amount of information gathered from a network. Simulation results show that our approach gathers more information, especially at low ping rates, when compared with several state-of-the-art models. To our knowledge, this is the first implementation of a VoI-based approach for a centralized WSN that is intended to efficiently maximize the amount of information gathered from the system.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

A Wireless Sensor Network is a collection of a huge number of sensor nodes that are spread over a wide area and are responsible for sensing, storing and transmitting information. A typical sensor node consists of a radio, responsible for transmitting and receiving information; a battery, which is the main energy source; a micro-controller, which is a small computer on a chip, responsible for performing computation and processing; an analog circuit and a sensor interface.

Research on WSNs has been quite prevalent over the past decade as they can be easily deployed in harsh environmental conditions, they are cheap and are easy to maintain. More recently, advancements in wireless technologies and wireless communication has led to widespread use of sensors in applications like health, military, security, agriculture, and wildlife monitoring. With such advancements, research has also spread into several domains and has led to the development of several architectures for sensor arrangement [1]; made deployment of sensors easier; made them cheaper and made them more scalable than ever. WSN sensor architectures are commonly implemented in either centralized or distributed fashion. In a centralized network, the sink node (which is usually considered to have unlimited resources in terms of processing and energy) has complete control of the network and is responsible for sensor deployment and resource allocation. This means there is less burden on the sensor nodes and no self-organization costs are inposed on them. On the other hand, in a distributed network, sensors govern themselves and are responsible for deployment, allocation of resources and communication with neighboring sensors to transmit information to the sink node.

Let us consider an example scenario where a number of cheap, dumb sensors are spread over an agricultural field in a distributed manner to monitor agricultural factors like amount of water required for cultivation, amount of fertilizer required, depth at which a seed must be sown, etc. A decentralized arrangement for this network has the following flaws: first, each sensor communicates with its neighboring sensor(s) to transmit its information and, at the same time, it is involved in passing information of other sensors' data [1]. This means that the number of communications being performed by each sensor in such a setting is potentially greater than the network's energy budget can afford, as we know that data transmission consumes more energy than any other sensor network activity [2]. Hence, such an approach will deplete the sensors' energy in short order, and the expensive process of retrieval and redeployment must take place. Second, certain sensors located close to the sink get exhausted quicker in comparison to other sensors, as most sensors' data needs to pass through one of these sensors [3]. Hence, for scenarios like the one mentioned above and many others like monitoring of buildings, it is often advisable to opt for a centralized architecture like the one depicted in Figure 1.1, which eliminates the flaws mentioned above.

However, in order to take advantage of the benefits mentioned above, we will have to deal with several problems associated with WSNs. Probably the most important of such problems is to improve the energy-efficiency of the sensors within the WSN. This is because the sensors ordinarily have small, low-capacity batteries as their only energy source. Once a battery is drained it can be very expensive to replace it, and more importantly, we will lose the ability of the network to give us an accurate measure

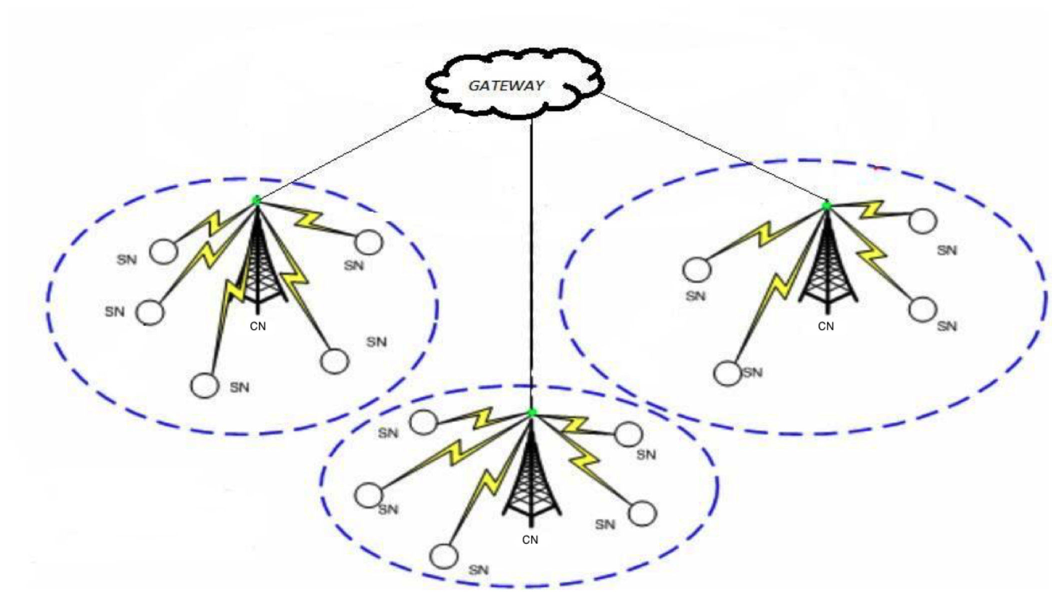

Figure 1.1: A centralized WSN solution to WSNs on an energy budget

of the sensing environment during the period when the sensor is offline. Hence, it is crucial that we prolong the lifetime of sensors by improving the energy efficiency of a sensor network deployment scheme. Over the years, several approaches have been proposed to address this problem.

Another problem lies in the fact that not all the sensors at every moment of time provide the best information. Most of the time, sensors send redundant information, which is non-informative in terms of entropy. Many approaches like data fusion, data aggregation, clustering etc. have been proposed to reduce redundant data [4] [5]. Also, several approaches have been proposed in order to maximize the amount of information gathered from a system [6]. This problem has also been proven as NP-Hard [7].

In this paper, we propose an approach that makes use of Information Entropy to determine the theoretical upper bound on the VoI available in a network and to increase the amount of information gathered from the network. We achieve this by identifying the most informative sensors consistently and query/ping them more often on comparison to others, thereby increasing the overall information gathered from the system.

In order to show energy-efficiency, our approach varies the rate of sampling to prove that it gathers more information from the sensors even at low sampling rates, instead of transmitting information to the central node whenever a change is seen – which eventually leads to making excessive, unnecessary transmissions and depleting the sensors' energy more quickly. The simulation results shown below are concrete proof that our approach (in comparison with two state-of-the-art models) gathers more information at every ping rate taken into consideration. These results are further evidence for the fact that our model gathers more information per ping and is energy-efficient in comparison.

CHAPTER 2

REVIEW OF LITERATURE

The main goal of a WSN is to collect data from the sensing environment and send it to the end user for analysis. Research to maximize this information being collected has led to development of several models [8]. Generally, in order to increase the overall information to be gathered, the sensors will have to sense more and transmit that sensed information to the sink. We know that energy conservation of sensor nodes is vital in WSNs and we cannot afford the communication costs to be high, as it will lead to depletion of the network sooner than expected. There have only been a handful of papers published considering both these issues together. Our approach is the first to implement a VoI-based technique in centralized networks, in order to address increasing the information gathered while simultaneously reducing overall communications. Before we take a look at our technical approach, we will look at some of the recent work done in these areas.

Various techniques are employed among multi-hop communication approaches, such as sensor shutdowns and censoring, are usually used to increase the energy-efficiency in WSNs. We should keep in mind that energy management in WSNs involves not only reducing the energy consumption of a single sensor node but also maximizing the lifetime of the entire network.

Data aggregation is one approach, which considers the problem of sending redundant information to the sink and aims to reduce such information reaching the sink. Several data aggregation techniques can be seen in [5], where authors show how redundant data proves to be expensive in terms of system performance, energy consumption and congestion. However, authors in [9] have proven that aggregation implies performing more complex operations than simply relaying traffic, and this can lead to an increase, rather than a decrease, in the overall energy consumption. Also, higher aggregation could be costly in terms of loss of information.

In [10], authors propose a distributed entropy-based data aggregation model with the aim of sending only surprising information to the sink node. This is one work closely related to our approach as authors use propose local and global probability models for computing entropy. We, on the other hand, calculate entropy of the data and use a probability approach to increase the amount of information gathered from the system by reducing the overall transmissions. In [10], authors use the entropy computed in data aggregation and from [9], we know that aggregation is not easy to handle. In contrast, we use the entropy calculated to consistently select the most informative sensor at every instant thereby increasing the overall information gathered.

In [7], authors propose an adaptive algorithm based on "Adaptive Compressive Sensing" to obtain an accurate approximation of the sensing field with minimum energy consumption as possible. Here, the sink node makes projections to the sensor nodes if the readings transmitted by them are not satisfactory. Authors jointly optimize the routing and compression to obtain optimal measurement, which greatly increases the complexity. On the other hand, our VoI based approach uses entropy to firstly, determine the theoretical maximum amount of information available and secondly, to randomly identify the informative sensors based on their weighted VoI. From the results shown below, we show that our model is on course to reach the theoretical maximum computed which means we will have a very accurate measure of the environment being sensed depending on the number of transmissions.

Clustering is another approach where several techniques have been proposed, particularly to increase the energy-efficiency. The main goal is to reduce the number of transmissions by sending information collected by each sensor of a cluster to a particular sensor, called a cluster head (CH). CHs are usually responsible for processing, filtering, aggregating and transmitting non-redundant information to the sink node [11] [12] [13]. Authors in [14] propose a clustering approach in a centralized WSN to increase the energy-efficiency. Here, CHs are chosen based on the amount of energy available to the sensors. This means that every sensor within the network has to have on-board processing abilities, which may impose higher costs on the network's energy budget than it can afford. Moreover, allocation of additional resources to the elected CH is a problem, as it should have the abilities to perform additional tasks and transmit information to the sink node.

On the other hand, our approach assumes the CN to be responsible for performing the computations, leaving the sensor nodes to gather information and send it to the CN upon request. We increase the amount of information gathered from the system by calculating the entropy of the data and using it to randomly query the sensors based on the current weighted VoI. Moreover, our approach does not consider any inter-cluster communications to be involved thereby further reducing the transmissions and hence saving energy. We compare the implementation of our model to the model proposed in [14], where performance improvement can be clearly observed.

Recently, self-censoring has become a major focus of research. This is a distributed, in-network processing technique where sensors function independently and are responsible for performing computations to determine whether to send data to the sink node or not [15] [16] [17]. In this scheme, sensors only send measurements that are deemed sufficiently informative, thereby reducing the number of transmissions being made by the sensors. In [18], authors implement a sleep/wake mechanism in the state-of-the-art self-censoring technique to enhance the energy efficiency. Here, a sleeping and censoring combined scheme is proposed to jointly optimize the energy consumption cost under the optimal sleeping constraint and the censoring thresholds. However, authors in [19] have proved that sensor start-up can consume more energy than information transmission. Using this demonstration, we can say that each sensor consumes energy for gathering information, performing computation, sensor start-up or wake-up and for transmitting information, hence depleting the node faster than normal.

Our approach, in comparison, considers a model wherein sensors consume energy only for sensing and transmitting information to the sink node. Our approach considers the information content of every sensor's readings, calculates the entropy for the information, and uses a probabilistic approach to select the most informative sensor consistently.

In [15], a censoring approach is used to reduce the number of transmissions to the sink. They propose a model based on the likelihood ratio (LR). Sensors only transmit information to the sink node if the information gathered by the sensor exceeds the LR determined. LR is calculated using the following equation

$$LR = (Sensitivity)/(1 - Specificity) \tag{2.1}$$

This can be explained as the dual of the ratio of the probability of false positives and false negatives.

Our model, on the other hand, uses a VoI-based technique using entropy to determine information in the system. This entropy determines the sensor to be selected for querying, and our model increases the information gathered using fewer communications (or "pings"), therefore improving the energy-efficiency. Simulation results show

how our approach outperforms this model by gathering more amount of information per ping.

CHAPTER 3

METHODOLOGY

As we have seen in the previous section, research incorporating both information extraction and energy-efficiency in WSNs is limited, particularly in centralized networks. In this paper, we consider a WSN organized in a centralized fashion, where all the sensor nodes are deployed over a wide area. There can be multiple central nodes (denoted by CN) depending on the area and number of sensor nodes being deployed. However, we will be considering only one CN for the networks we are dealing with in this paper. In this paper, we assume that all sensors have the same sensing range and consume the same energy for performing activities like data transmission and reception. Also, we assume that the identity and position of each sensor are fixed and known both to the CN and the sensor itself. Additionally, we consider the CN to have unlimited resources in terms of energy and processing abilities. Finally, we assume that energy consumption for sensing is negligible, which is a reasonable consideration [20].

In this paper, we will compute the amount of information extracted by examining the total number of transmissions required to collect this information from the sensing field. This work distinguishes itself from existing works [7] in that it uses a probability-based sensor selection scheme to increase the information gathered and to improve the energy-efficiency of the entire network.

3.1 Theoretical Maximum VoI ($\mathbf{TM_{VoI}}$)

In order to evaluate the information that our approach is extracting from the network, it is important to compute a theoretical maximum for the amount of information present within the system. Doing so also provides an idea of the accuracy of a given model's sensing field. We obtain the TM_{VoI} using an entropy-based approach, as shown in Algorithm 1. To calculate the TM_{VoI}, we assume that the CN is able to query each of the sensor nodes within the network *whenever* the sensor detects new information. We then sum the calculated entropies associated with those values. Entropy provides the average amount of information present within a message. TM_{VoI} is computed using the following algorithm.

Algorithm 1 Determining Theoretical Maximum VoI (TM_{VoI})

Input: Set of Sensors, SN

Output: Theoretical Maximum VoI, TM_{VoI}

```
for i = 0 to SN.length do
    data[i] = (SN[i])                              ▷ Initially, request data from each sensor in SN
    H(data[i]) = P(data[i])*log_e(1/P(data[i]))    ▷ Compute entropy H of each sensor
    I[i] = H(data[i])                              ▷ Save the computed entropy in I[i]
    TM_VoI = TM_VoI + I[i]                         ▷ Finally, add computed information to TM_VoI
end for
```

3.2 Probability Model

After computing the TM_{VoI} of a network, we use our probability model based on weighted random selection to preferentially select the most informative sensors at each communication opportunity. Using this approach, we do not select information from all the sensors at each instant of time; instead we will most often select the most informative sensor at that instant. The main goal of our work is to gather maximum possible information from a network even at a low transmission rate, thereby enabling end users to get a good approximation of the network being monitored. At the same time, we demonstrate our model's energy-efficiency on comparison to other existing models. Our implementation is clearly outlined in the following algorithm.

Algorithm 2 Probability-based sensor selection

Input: Set of sensors, SN
Output: Total information in system, I

Step 1: Calculate entropy at every instant
for $i = 0$ to $SN.length$ **do**
  $data[i] = (SN[i])$ ▷ Initially, we request data from each sensor in SN
  $H(data[i]) = P(data[i]) * \log_e(1/P(data[i]))$ ▷ Compute entropy H of each sensor
  $I[i] = H(data[i])$ ▷ Save the computed entropy in I[i]
end for
Step 2: Probability model for sensor selection at every instant
$I_{\text{total\_inst}}(P) = \sum_{i=1}^{SN} I[i]$ ▷ Compute the sum of entropies of all sensors in network
$R = random(0, I_{\text{total\_inst}})$ ▷ Generate a random number between 0 and $I_{\text{total\_inst}}$
Step 3: Randomly select the information-weighted sensor
for $i = 0$ to $SN.length$ **do**
  if R <I[i] **then**
    $I = I + I[i]$ ▷ Select sensor i
  else
    $R = R - I[i]$
  end if
end for

The general idea of Algorithm 2 is to select a sensor to gather information from at each time point in time by rolling a die. We generate a random number between 0 and the current VoI gathered from the network, and use it to select a sensor to query, weighted by the information content of that sensor. Initially, the CN assumes a uniform distribution of information value across all of the network's sensors, and queries them all equally often. Sensors transmit their current information state to the CN. The CN now computes the entropy of the data received from each of the sensors using the formula stated in Algorithm 2. At the next time point, CN repeats step 1 and computes entropy. Now, the CN implements the probability model to select a sensor at this instant. It does so by adding up the entropies of each sensor. This step is similar to determining the discrete cumulative density function (CDF). Next, we calculate a random number ranging between 0 and the sum of entropies. Then, starting from sensor 1, the CN subtracts the entropy of sensor 1 from the random number if random number is greater than the entropy of sensor 1. If not greater, then select sensor 1. Else, we move to the next sensor and subtract entropy from the random number, and repeat this process until a sensor's entropy is greater than the reducing random number's value. In that case, we choose that sensor to query, and add its reading to the overall information of the network. A great benefit of using this approach is that it ensures that every sensor is being attended to and the entire network is monitored.

Now, let us consider the following example to understand the functioning of our algorithms. To determine the $\mathrm{TM}_{\mathrm{VoI}}$, the CN is responsible for identifying sensors where a change has been observed and then querying it to retrieve the information. The CN then computes the entropy for the data and adds it to the overall VoI gathered so far. Assume a network consisting of 5 sensors that are randomly deployed to monitor temperatures. At a particular instant, each sensor senses the environment and stores the information as follows: [21.92, 25.84, 17.08, 11.14, and 38.16]. The CN queries each of these sensors and computes the entropy, say, [3, 2, 1, 0, 1], and sums each entropy value to calculate a $\mathrm{TM}_{\mathrm{VoI}}$ of 7. This procedure is repeated at every time step.

Now, using the same example, let us see how our probability model functions. After we calculate the summation of values for a particular time step, we generate a random number R between 0 and 7 (the total information), say 3.24. Now we scan through the array of sensors, starting at 1, and select a sensor when its entropy is greater than the random number is generated. However, if a sensor's entropy is found to be greater, we subtract it from the random number and move to the next sensor. So, in the above example, we do not select 1, as its entropy is less than the random number. Sensor 2 on the other hand is selected as its entropy of 1 is greater than 0.24. Next, we query sensor 2's current reading and add it to the overall information gathered from the network.

As you can see, using this algorithm, we will preferentially select particularly informative sensors at each time step, while ensuring that the entire network is monitored. This is extremely useful, especially when dealing with event-based applications such as earthquake monitoring. The main motivation behind our work has been energy-efficiency, and with applications increasingly being built with a strict energy budget in mind, such as in agriculture or building monitoring, an implementation that maximizes the amount of information gathered per ping is important. In the next section we show how our model outperforms other state-of-the-art models.

CHAPTER 4

FINDINGS

In order to analyze the performance of our approach proposed in this paper, we developed a simulator to compare our work against a clustering approach proposed for a centralized network [14] and a distributed censoring approach proposed in [15]. The sensor network on which we conducted experiments consists of sensor nodes that are static and homogeneous, and there is only one static CN which has access to an unlimited amount of energy. We also assume that nodes are deployed randomly, forming a high-density network. Performance is measured by the quantitative metrics of amount of information gathered and the number of transmissions taken to gather this information. Plotting graphs using these two measures will not only tell us about the information being gathered but will also demonstrate the energy-efficiency of the models in consideration. This is because energy consumption is proportional to the number of transmissions performed [2].

To show the results, we made use of two data sets. First, we use a data set collected from a simple single-hop wireless sensor network deployment of four TelosB sensors [21]. The data consists of humidity and temperature measurements collected during a 6-hour period at intervals of 5 seconds. Second, the Intel Research Lab data [22] contains data collected from 54 sensors deployed in the Intel Research Lab measuring several factors such as humidity, temperature, light etc. Only the measurements of temperature were considered from both these data sets.

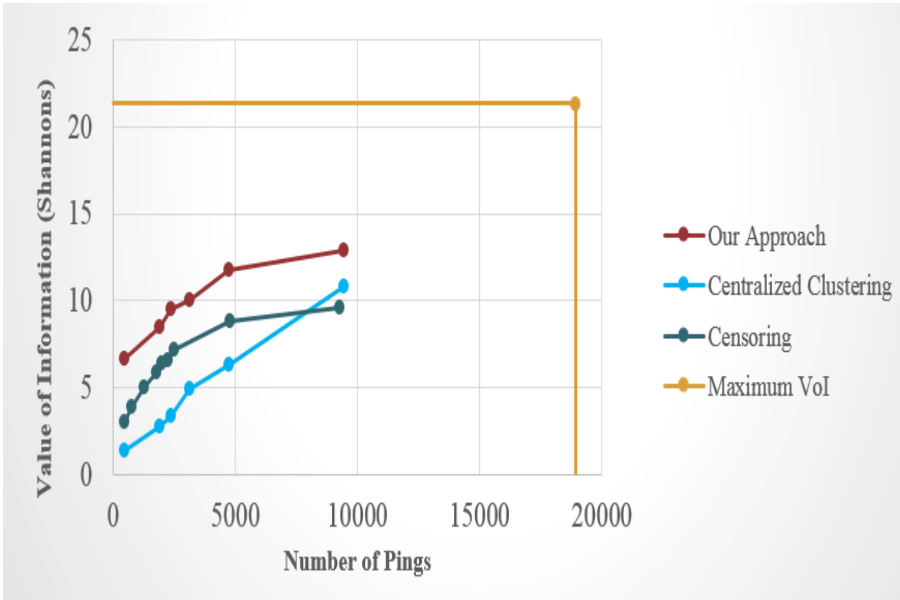

Figure 4.1: Information gathered vs Number of pings —TelosB sensor data

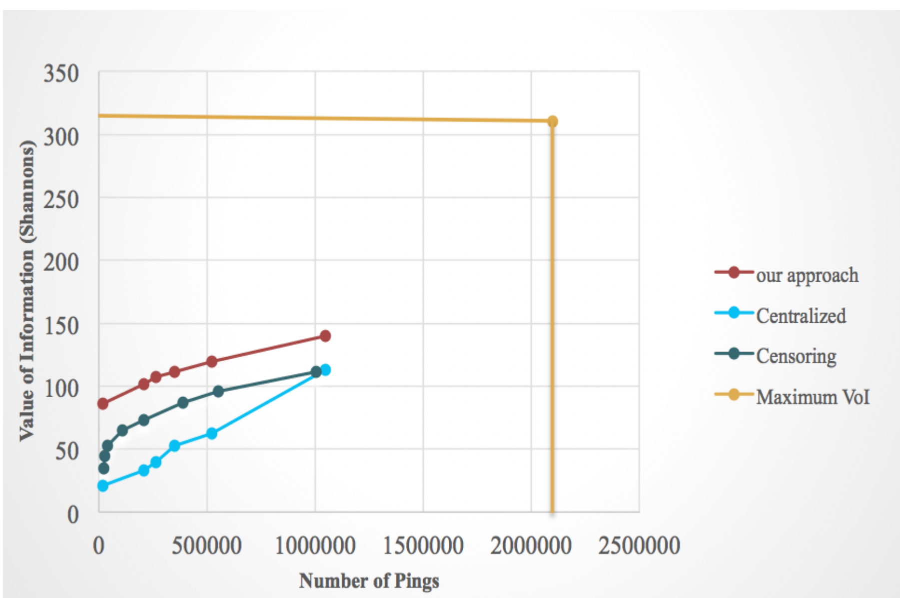

Figure 4.2: Information gathered vs Number of pings —Intel Research Lab data

In the first experiment, we considered data from the TelsoB sensors [21], and as stated in the algorithm above, we calculate the entropies for each measurement of this data using entropy equation. Next, we calculate the $\mathrm{TM}_{\mathrm{VoI}}$ available in the data set using Algorithm 1. This will help us understand how each model will rank against the overall information available in the network. Next, using Algorithm 2, we run the simulator with this data set by varying the number of pings. The variation in pings gives us a measure of how each model is performing at different rates of sampling. Obviously, the more a network is sampled, the more information can be collected, but when we are dealing with networks that are on an energy budget, we cannot afford sensors that perform excessive transmissions, as it would deplete the network much faster than expected.

Figure 4.1 illustrates the performance of all the models in consideration, in terms of the amount of information gathered at different ping rates for data gathered by the TelosB sensors in [21]. We can clearly see that our model outperforms both the models at each variation of the ping rate. In our second experiment, we considered the Intel Lab data [22], and followed the same procedure outlined in Section 3.2: initially determining the theoretical maximum VoI available in a system and later choosing the informative sensors using algorithm 2. From Figure 4.2, we again notice our model gathering more information at every variation in the ping rate, compared to the other two models.

Now, let us take a look at the reasons behind the performance variation and why our model would be an ideal implementation in order to increase the amount of information being gathered. Although both the models chosen for comparison intend to increase the per-ping information gathered while reducing the number of transmissions, they were developed with the intent of discarding the information gathered by sensors in most cases. Considering [14], the cluster head (CH), after gathering information from each sensor within the cluster, performs data aggregation and discards most of the data, sending only part of the information gathered. Moreover, this approach consumes extra transmissions at every time step, as each CH again needs to transmit information to the sink. [15] on the other hand, uses Likelihood Ratio (LR, defined as the dual of the ratio of the probability of false positives and false negatives.) to transmit information to the sink. Sensors found to have LR with enough information report information to the sink node. Our experiments showed that there were many instances where LR was found to be low which resulted in excessive transmissions. Also, implementing such a decentralized model means that sensors will have to deal with the problems reviewed in Section 1. We also feel that if the computed LR were quantized, the information gathered would be higher for this approach than computed. The performance of our model is entirely dependent on the probability model proposed in Section 3.2. The algorithm, at every time step, is able to choose particularly informative sensors, thereby gathering more information on comparison to other models.

The following implication can also be made from the graphs. As we know, more information can be gathered from a network if it is sampled more often, meaning that one would expect to see an asymptotic graph with the previous statement taken into consideration. This can be seen in all of the models in both of the experiments.

In our third experiment, we considered the Intel data set, and queried all of the 48 sensors considered by slowing the ping rate down to 1/2 of the original. This ping rate was chosen as a good illustration of the network performance in terms of the amount of information gathered. A graph was then plotted against individual sensors and the number of times each sensor was chosen at different time steps. Figure 4.3 shows the ranking of the sensors (from left to right) in terms of the number of times each sensor was selected for querying. This plot shows the consistency of the model in

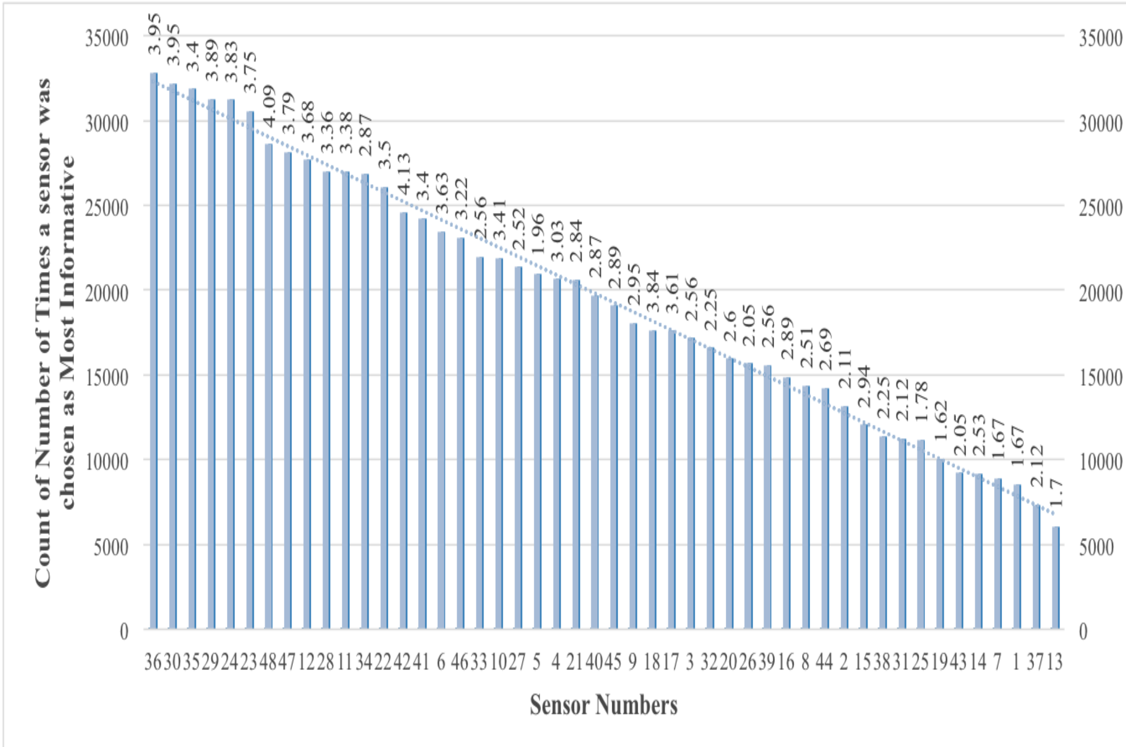

Figure 4.3: Ranking of the sensors according to the number of times chosen as Most-Informative

choosing the sensors at each instant. Figure 4.3 is further proof to the results above that our model correctly queries sensors at an appropriate rate, given the amount of information predicted for each sensor. We can also observe that the sensors gathered a proportional amount of information with respect to the times they were sampled. For example, sensor 36, which was chosen highest times (approx. 32000) gathered close to 4 Shannons of information, while sensor 26 gathered 2.05 Shannons over approximately 17000 samples. The trend line in the graph shows the average number of queries sent to each sensor.

In our final experiment, considering the same assumptions as in our previous experiment, we plot a graph against sensor number and the times each sensor was chosen as most informative but, this time to show the overall VoI gathered by them. Figure 4.4 shows the ranking of the sensors (from left to right) in terms of the amount of information gathered by them during a run. Our model always randomly selects a

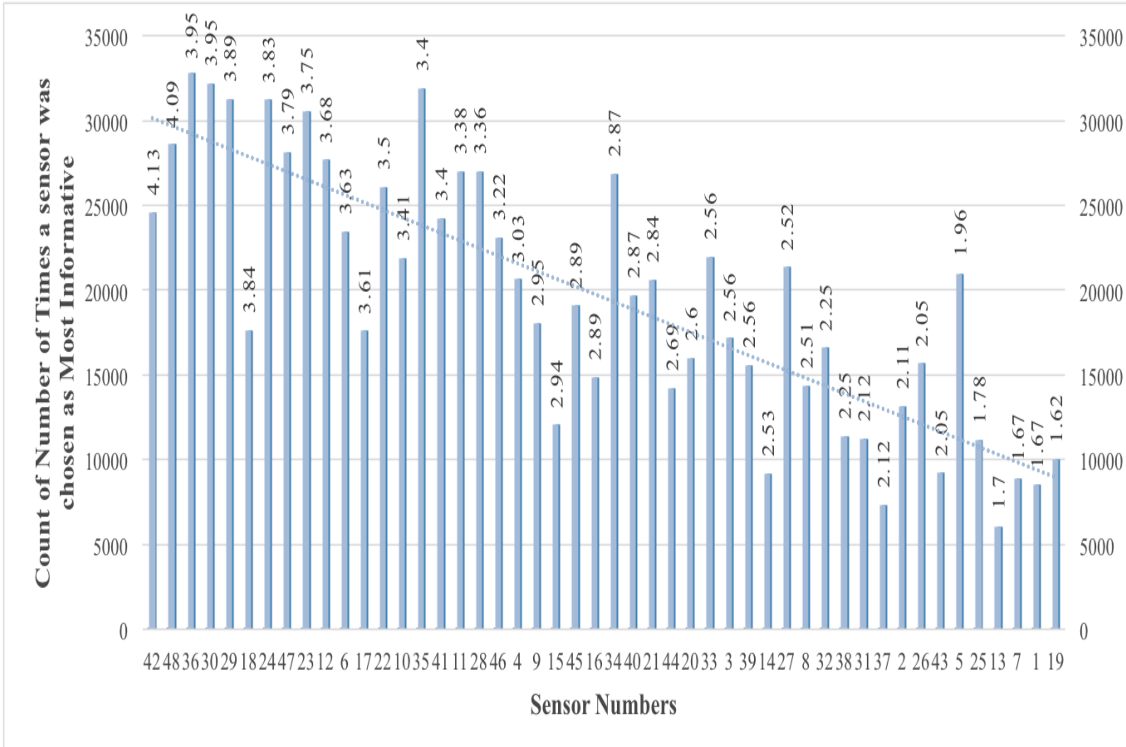

Figure 4.4: Ranking of the sensors according to amount of Information Gathered

sensor based on the current VoI gathered, not on the theoretical maximum VoI (which is not accessible to an on-line algorithm). If that were the case, we would have seen a result similar to figure 4.3. However, as the VoI distribution changes constantly with every query at every time step, we do not see a smooth slanting curve. For the same reason, we can also observe that certain sensors have been queried less often, although in the end they provided a greater amount of information, and vice-versa. However, the trend line in the graph is proof that, sensor selection is proportional to the number of times each sensor is being queried. It also shows the average amount of information gathered by each sensor. From these analyses, we can clearly see how our model outperforms other state-of-the-art models, both in terms of information collection and energy efficiency.

CHAPTER 5

CONCLUSION

In this paper, we started off by proposing a model that is best suitable for WSNs that are going to be built on an energy budget. We stated several reasons why a centralized architecture with a high-energy central node would be beneficial in comparison to a distributed architecture. Next, we proposed a novel entropy-based VoI approach for a centralized network to determine the theoretical maximum VoI present within a network. We also present a probability-based sensor selection algorithm to consistently select informative sensors, enabling the model to increase the amount of information gathered from a network. Moreover, by varying the sampling rate, we have shown that our model can extract a high amount of information even at very low sampling rates.

The performance of our approach was compared to two state-of-the-art techniques (clustering for a centralized WSN and distributed self-censoring) currently being employed to increase both the amount of information being gathered and the energy efficiency. We developed a simulator emulating all three models, and results clearly show that our approach gathers more information per ping, implying that it is more energy-efficient than the comparison models. Hence, we conclude that our model provides an energy-efficient approach for increasing the information gathered from a centralized network.

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

[2] M. G. C. Torres, *Energy consumption in wireless sensor networks using gsp*. PhD thesis, University of Pittsburgh, 2006.

[3] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *Wireless Communications, IEEE Transactions on*, vol. 1, no. 4, pp. 660–670, 2002.

[4] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer communications*, vol. 30, no. 14, pp. 2826–2841, 2007.

[5] R. Rajagopalan and P. K. Varshney, "Data aggregation techniques in sensor networks: A survey," 2006.

[6] S. Ji, J. S. He, and Z. Cai, "Data gathering in wireless sensor networks," in *The Art of Wireless Sensor Networks*, pp. 535–565, Springer, 2014.

[7] C. T. Chou, R. Rana, and W. Hu, "Energy efficient information collection in wireless sensor networks using adaptive compressive sensing," in *Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on*, pp. 443–450, IEEE, 2009.

[8] A. Abuarqoub, M. Hammoudeh, and T. Alsboui, "An overview of information extraction from mobile wireless sensor networks," in *Internet of Things, Smart Spaces, and Next Generation Networking*, pp. 95–106, Springer, 2012.

[9] L. Galluccio, S. Palazzo, and A. T. Campbell, "Efficient data aggregation in wireless sensor networks: an entropy-driven analysis," in *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pp. 1–6, IEEE, 2008.

[10] A. Sinha and D. K. Lobiyal, "Performance evaluation of data aggregation for cluster-based wireless sensor network," *Human-Centric Computing and Information Sciences*, vol. 3, no. 1, pp. 1–17, 2013.

[11] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *System sciences, 2000. Proceedings of the 33rd annual Hawaii international conference on*, pp. 10–pp, IEEE, 2000.

[12] S. Lindsey and C. S. Raghavendra, "Pegasis: Power-efficient gathering in sensor information systems," in *Aerospace conference proceedings, 2002. IEEE*, vol. 3, pp. 3–1125, IEEE, 2002.

[13] O. Younis and S. Fahmy, "Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *Mobile Computing, IEEE Transactions on*, vol. 3, no. 4, pp. 366–379, 2004.

[14] S. D. Muruganathan, D. C. Ma, R. I. Bhasin, and A. Fapojuwo, "A centralized energy-efficient routing protocol for wireless sensor networks," *Communications Magazine, IEEE*, vol. 43, no. 3, pp. S8–13, 2005.

[15] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: A low-communication-rate scheme for distributed detection," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 32, no. 2, pp. 554–568, 1996.

[16] R. Jiang and B. Chen, "Fusion of censored decisions in wireless sensor networks," *Wireless Communications, IEEE Transactions on*, vol. 4, no. 6, pp. 2668–2673, 2005.

[17] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances," *Signal Processing Magazine, IEEE*, vol. 29, no. 3, pp. 101–116, 2012.

[18] S. Maleki, A. Pandharipande, and G. Leus, "Energy-efficient distributed spectrum sensing for cognitive sensor networks," *Sensors Journal, IEEE*, vol. 11, no. 3, pp. 565–573, 2011.

[19] S. Cho and A. P. Chandrakasan, "Energy efficient protocols for low duty cycle wireless microsensor networks," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, vol. 4, pp. 2041–2044, IEEE, 2001.

[20] W. Hu, C. T. Chou, S. Jha, and N. Bulusu, "Deploying long-lived and cost-effective hybrid sensor networks," *Ad Hoc Networks*, vol. 4, no. 6, pp. 749–767, 2006.

[21] S. Suthaharan, M. Alzahrani, S. Rajasegarar, C. Leckie, and M. Palaniswami, "Labelled data collection for anomaly detection in wireless sensor networks," in *Intelligent sensors, sensor networks and information processing (ISSNIP), 2010 sixth international conference on*, pp. 269–274, IEEE, 2010.

[22] P. Bodik, C. Guestrin, W. Hong, S. Madden, M. Paskin, and R. Thibaux, "Intel Lab Data," 2004.

VITA

Sushyanth Pamulaparthy

Candidate for the Degree of

Master of Science

Thesis: A Centralized Energy-Efficient Model for Increasing the Information Gathered in Wireless Sensor Networks

Major Field: Computer Science

Biographical:

Personal Data: Born in Hyderabad, Telangana, India on February 6, 1992

Education:

Received the B.Tech degree from Jawaharlal Nehru Technological University Hyderabad, Telanagana, India 2013

Completed the requirements for the degree of Master of Science with a major in Computer Science at Oklahoma State University in July, 2015.