Optimal, Secure Cluster Head Placement Through Source Coding Techniques in Wireless Sensor Networks

Tata Jagannadha Swamy, Garimella Ramamurthy, and Padmalaya Nayak

***Abstract*— In many applications of wireless sensor networks (such as military communications), secure communication, message delay minimization and energy efficiency are crucial. Such requirements constrain special or Important Cluster Head (ICH) placement over the network architecture modeled by a tree. The optimal important cluster head placement problem is formulated and solved using source coding results (providing minimum possible delay and security through prefix-free paths over the tree). Also, through simulations energy efficiency of the proposed approach is established. The reported research is naturally applicable for many applications of Wireless Sensor Networks (WSNs) such as Body Area Networks (BANs).**

***Index Terms*— Wireless sensor networks, cluster head, important cluster head, prefix-free path, source coding, Kraft's inequality.**

I. Introduction

THE rapid growth of wireless communication technology has led to the development of low power, low cost and tiny sensor nodes. Each tiny node has the capacity to sense, process, and to transmit the sensed data to base station. Randomly deployed tiny nodes form a network for data transmission. These Networks have become extremely popular due to the large number of applications in the areas of intrusion detection, habitat, environmental monitoring, etc. However, some of the limitations of sensor nodes are mainly in storage capacity, power capacity, and short communication range [1]. These limitations can be overcome by some efficient sensor node placement schemes and hierarchical implementations [2].

Secure data transmission and low energy consumption is possible through choices of Cluster Heads (CHs) in the wireless sensor field [3]. The entire sensor field is divided into several small fields known as clusters and each one is headed by a cluster head [16]. Among cluster heads, some are important and others are ordinary CHs. These important cluster heads transfer data between a base station and other cluster heads.

In Wireless Sensor Networks (WSNs), the cluster head placement problems have not been adequately studied in the literature, while secure placement of nodes in sensor field has been at the core of research [7], [8]. Motivation : Wireless Sensor Networks find many applications in fields such as Military Communication, healthcare etc.

- Security Constraint: Specifically, in the case of communications among military personal, certain messages can be received by officers of certain cadre and above (in the military hierarchy) only.
- Optimization Constraint: Also, it is clear that messages will have delay constraint and must be received by officers in real time.

In this letter, we propose a novel optimal approach for the sensor placement in WSNs. The main goal is to minimize average depth of Important Cluster Heads from the base station by reducing the number of hops. Further, we ensure message security and make, the paths from Base Station (BS) to Important Cluster Heads to be prefix-free. The contributions are, (i) Minimize the average depth of an ICH from the BS in terms of hop-count, (ii) Provide an optimal hierarchy in the network for secure data transmission. (iii). Relate CH placement and source coding technique. The rest of the letter is organized as follows: In Section II, the problem specification with the help of Secure Cluster Head placement in the network field is discussed. In Section III, the relationship between secure cluster head placement and optimal source coding techniques in WSNs are discussed and later followed by conclusion in section IV.

II. System Model

The entire system structure is based on the approaches made for an optimal, energy efficient and secure wireless sensor network. To minimize the average depth of an ICH from the BS in terms of hop-count, to provide an optimal hierarchy in the network for secure data transmission, and to relate CH placement and source coding technique are the three main proposals required for an efficient network design.

- Based upon hierarchy among military personnel, importance values are assigned/decided
- The total number of important cluster heads is decided by a number of officers in the military hierarchy

We realized that the above motivation can be understood by modeling the real-time communication problem as a source coding problem. Specifically the important cluster heads are chosen as the prefix free nodes (code words) on the military hierarchy so that the security constraint is met.

In this context, a tree based structure is selected and the depth of a tree is the path with maximum number of hops between BS and leaf nodes and the depth of BS in a tree structure is zero. The number of leaves present in a

Manuscript received September 26, 2019; revised October 17, 2019; accepted November 2, 2019. The associate editor coordinating the review of this letter and approving it for publication was Y. Wu. *(Corresponding author: Tata Jagannadha Swamy.)*

T. J. Swamy is with the Department of Electronics and Communication Engineering, Gokaraju Rangaraju Institute of Engineering and Technology (GRIET), Hyderabad 500090, India (e-mail: tatajagan@gmail.com).

G. Ramamurthy is with the Department of Computer Science and Engineering, Mahindra École Centrale, Hyderabad 500043, India.

P. Nayak is with the Department of Computer Science and Engineering, Gokaraju Rangaraju Institute of Engineering and Technology (GRIET), Hyderabad 500090, India (e-mail: padma_nayak@yahoo.com).

Digital Object Identifier 10.1109/LCOMM.2019.2953850

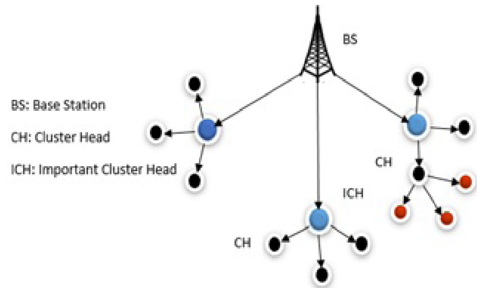

Fig. 1. Wireless sensor network D-ary tree architecture.

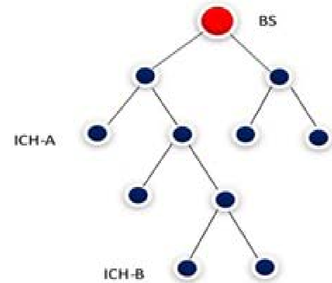

Fig. 2. Example of a binary tree structure with WSNs.

full tree can be calculated by D^d, Where 'd' is the depth of the tree and 'D' represents the type of tree structure (e.g. $D = 2$ implies binary tree). Total number of nodes in a binary tree structure is $N = 2^{d+1} - 1$ and the total number of leaves is equal to nodes at the last level of the tree. For any D-ary tree, the total number of nodes in the structure is given by $N = \dfrac{D^{d+1} - 1}{D - 1}$. Fig. 1 is an example of hierarchical sensor network architecture representing CHs, ICHs and BS. Throughout the letter, wherever required, we consider fully binary trees. The results also apply for fully D-ary trees.

A. Importance Values of Cluster Heads: In a sensor field, selection of important cluster head is essential. In order to signify this distinction, a value of importance is assigned to some cluster heads in the organization based on its internal energy and its closeness to the base station. According to the importance value, ICHs are labeled as more important than ordinary cluster heads in the hierarchical structure. For instance, a node with a hop count of zero has no superiors and only has subordinates. It implies that these nodes can act as the base station in the sensor field. It is important to note that cluster heads with a higher importance value in a sensor field should have a lower hop count. These nodes are more vital members in the sensor field and should be closer to the base station. It can be easily observed that hop count and importance values are inversely related. The normalized equation for the importance probabilities is shown in equation (1).

$$p_i = \frac{v_i}{\sum_{i=1}^{c} v_i} \tag{1}$$

where p_i is the probability of importance and v_i is the importance value of i^{th} important cluster head and 'c' is the total number of ICHs. Now consider a binary tree representing sensor field consisting of cluster heads. Prefix (or prefix-free or instantaneous) code is a code in which no codeword is a prefix of any other codeword. Prefix codes are uniquely and instantly decodable [9]. PROBLEM: Place the important cluster heads in a binary tree such that the average path length (in terms of hop count) is as small as possible and the paths from root node to ICHs are prefix-free.

In Fig. 2, node labeled as ICH-A has depth of two, whereas the node labeled as ICH-B has depth of four from the Base Station. A large importance probability directly implies a small depth of member node or a smaller hop count, which suggests that the member is more important to the base station. It is only logical for important cluster heads be near the base station of the sensor field. So, the idea is to define the average depth of all ICHs in the hierarchy of the sensor field to be the sum of the product of their respective depths and importance values, as shown in equation (2).

$$\bar{n} = \sum_{i=1}^{k} p_i n_i \tag{2}$$

where p_i is the importance probability of i^{th} cluster head. n_i is the number of hops of i^{th} cluster head from the base station and $\bar{n}$ is the average path length. An immediate inference to this is: If p_i is large, n_i can be made small and vice versa.

B: Prefix-free path and Tree structure: A path is defined to be a sequence of edges from a start vertex or member to an end vertex or member and each node is associated with ICH or codeword in the sensor field hierarchy. A D-ary tree hierarchy will produce a D-ary codeword. Here it was selected as a binary tree with $D = 2$. Starting at the Base Station, we append a '1' to the edge leading to the right sub-tree or a '0' for the edge leading to the left sub-tree. Once the end member node is reached, the sequence of steps with 1's and 0's will represent the code word for that member. The prefix-free condition is significant because no two ICHs have a codeword with the same prefix-free path. If the base station wants to communicate with an important cluster head, then it must choose the specific prefix-free path only.

C: Inherent Security Provided by Prefix-free path: Organization (hospital, military etc…) determines which nodes get Higher Important Values (HIV) and which nodes get lower (LIV). Goals: i) Transfer of delay sensitive sensed data from sensors (cluster heads) to the base station i.e. Optimization of average hop count from important cluster heads for Base Station. ii) The communication must be secure in the sense that information from ICH's of higher importance value should not reach those ICH's of lower importance values i.e. when the base station wants to send cryptographic keys, the key meant for CH with Lower Importance Value (LIV) can be heard by a CH of Higher Importance Value (HIV). But the key meant for any CH with higher importance value (HIV) cannot be heard by CHs of lower importance value. Thus the security of WSN is automatically enhanced.

III. Secure Cluster Head Placement

Let X be a random variable and $X_1, X_2 \ldots X_k$ Be statistically independent and identically distributed (IID) variables, having a common probability mass function (generated from a common source). $n_1, n_2, \ldots, n_k$ are the hop counts with probabilities $p_1, p_2, .., p_k$. We are interested in the mean length of path in terms of number of hops from a leaf nodes i.e. the number of hops required for packets to reach the base station [13]. For security and energy concerns, minimum number of hops is preferable for packet data transmission [6], [12].

A. Source Coding Problem: In [14], it is considered that any prefix-free code with the codeword length/hop-distances must satisfy the inequality shown in equation (3) [4].

$$\sum_{i=1}^{k} D^{-n_i} \leq 1 \tag{3}$$

Consider a prefix-free binary tree in which each node has two children and certain nodes of the tree represent the code words. The prefix condition on the code words implies that no codeword is an ancestor of any other codeword on the tree. A full binary tree of depth n_{max} has $2^{n_{max}}$ leaves. The channel codeword to a source symbol has the codeword length n_i. A prefix-free code can be constructed by using Kraft's inequality. We now discuss the lower bound on the average codeword length i.e. $\bar{n} = \sum_{i=1}^{k} p_i n_i$ yielding optimal codeword lengths

$$n_i = -log_D p_i \tag{4}$$

These non-integer codeword lengths yield average or expected codeword lengths as per equation (5)

$$\bar{n} = \sum_{i=1}^{k} p_i n_i = -\sum_{i=1}^{k} p_i log_D p_i = H_D(X) \tag{5}$$

The expected average codeword length $\bar{n}$ of any instantaneous D-ary code for a random variable X is greater than or equal to the entropy D-ary $H_D(X)$, that is,

$$\bar{n} \geq H_D(X) = \frac{H(X)}{log_D} \tag{6}$$

and with equality if and only if $D^{-n_i} = p_i$.

By choosing the codeword lengths (i.e. hop count from root node) suitably, kraft's inequality is satisfied. Hop count essentially determines the 'average delay' through the associate normalized importance values (i.e. probabilities i.e. 'p_i' Source coding provides achievable average delay for communication (using well known lower bound). Thus, the modeling problem boils down to the source coding problem. Specifically the optimization problem follows objective function $J(n_1, n_2, \ldots \ldots n_n) = \sum_{i=1}^{n} p_i n_i$ (where $n_i's$ are hopcounts and $p_i's$ are normalized importance values).

Constraint: $\sum_{i=1}^{M} D^{-n_i}$ where 'D' corresponds to balanced D-ary tree. In the presentation, correspondence to source coding is made clear even at the cost of redundant explanation (e.g. Explanation of security constraint and optimization problem).

B. Secure Cluster Head Placement with Huffman Coding: It has been discussed in section II that an optimally secure hierarchy is one in which the average hop count from the base station to ICHs is minimized. The smaller the distance, the more optimal it is. In a well-organized sensor field, the base station would prefer the more important ICHs to be closer than

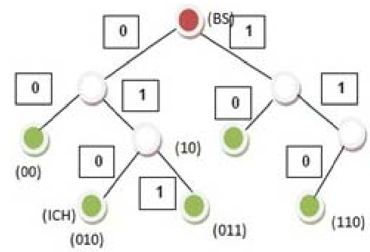

Fig. 3. Example of Huffman coding.

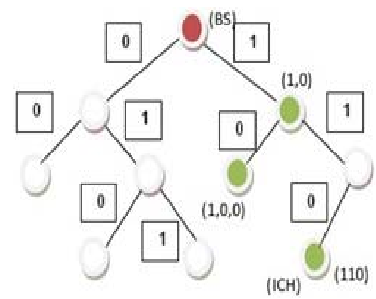

Fig. 4. Example of non-prefix code.

ICHs of lower importance. The problem of optimally placing ICHs is quite crucial in wireless sensor networks. To provide importance probabilities for all ICHs and to place them at different depths in the sensor field hierarchy, Huffman coding can be used. Considering the importance probabilities of ICHs, we can organize an optimal placement pattern in a sensor field which would also be unique. Huffman coding algorithm produces a hierarchical structure [13], [14] similar to the one shown below in Fig. 3. The average distance from the base station node to each of the cluster head nodes is minimized and a prefix-free path exists to all ICHs. Fig. 3 shows, no two important cluster head nodes share their entire path with another cluster head node.

Once the ICH is chosen at a certain depth, no other ICH can be chosen either in its sub-tree or from its subordinate cluster heads. In fig. 3, as already ICH was placed at codeword of '$1-0$', there cannot be another ICH at codeword of '$1-0-0$', because this violates the condition, i.e. No code word should be a prefix of another, and hence it becomes impossible to be uniquely decoded. It is important to note that after placing ICHs using Huffman coding, we can achieve an optimally secure hierarchy i.e. A prefix-free path to all important cluster heads, an optimal distance from the base station to ICHs according to importance probabilities (a uniquely decodable code scheme). Example of non-prefix code as shown in Fig.4.

C. Energy Efficiency and Security with Source Coding: In the routing protocol the data is propagated from nodes with higher hop count from the BS to the nodes with lower hop count (using a tree data structure). If a node at higher hop count receives a packet from lower hop count node [5], [10], it simply drops it, thus flooding is prevented, saves energy. By using prefix- free path, security of sensitive data (transmitted) is handled efficiently. This scheme assumes that there is hierarchical structure among the nodes. Only nodes with higher importance value have access to important data.

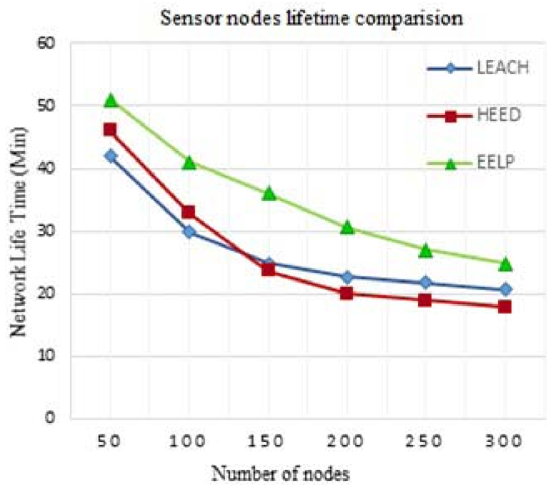

Fig. 5. Nodes vs node lifetime with different protocols.

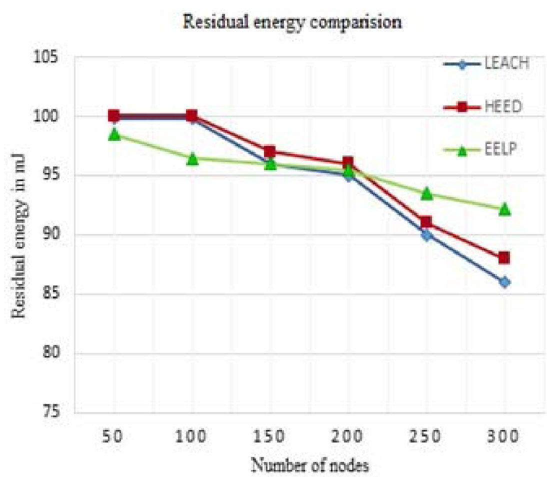

Fig. 6. Nodes-vs-residual energy with different protocols.

Lower importance nodes are prevented from accessing such information. In popular schemes like LEACH (Low Energy Adaptive Clustering Hierarchy) and HEED (Hybrid Energy Efficient Distributed) [11], [15], the data are flooded along all directions using Cluster Heads. It was shown that the above proposed protocol, termed by us as EELP (Energy Efficient Leveling Protocol), flooding of data is prevented from spreading over the entire tree structure. Thus the EELP saves energy, thereby increasing the network lifetime. In the following overall sensor network life time using EELP is compared to the popular protocols.

The simulation results of the proposed protocol are compared with the popular protocols in Fig. 5, and Fig. 6. In Fig. 5, the node lifetime with respect to the number of nodes are shown. If the node number is less, node or network lifetime is high with all three protocols, but it gradually decreases when the node number is increasing. EELP is showing better performance than other two protocols along with decrease in the network lifetime. Fig. 6 shows that the nodes residual energy with respect to active nodes in the network. Here, if the node number is less, EELP has the highest residual energy and will be decreasing gradually, when the number is increasing. Here EELP is showing better results than LEACH and HEED. Of all the analysis and simulations, EELP is more efficient than LEACH and HEED in terms of network lifetime, node residual energy. From the results and comparisons EELP is more energy efficient protocol.

IV. Conclusion

In this letter, we proposed a new protocol for optimal cluster head selection and node placement in WSNs. For Energy Efficiency, secure routing and network lifetime improvement, Important Cluster Heads (ICHs) are selected by providing some importance value to Cluster Heads and hierarchy among the CHs are maintained with the help of prefix-free path. A relationship is derived between secure cluster head placement and source coding and it is derived based on Kraft's inequality and Huffman coding. Results show that, the proposed EELP increases the network life time, providing data security and Energy Efficiency compared to LEACH and HEED protocols.

References

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

[2] M. Younis and K. Akkaya, "Strategies and techniques for node placement in wireless sensor networks: A survey," *Ad Hoc Netw.*, vol. 6, no. 4, pp. 621–655, Jun. 2008.

[3] M. Dhanaraj and C. S. R. Murthy, "On achieving maximum network lifetime through optimal placement of cluster-heads in wireless sensor networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2007, pp. 3142–3147.

[4] J. Mohorko, U. Pesovic, P. Planinsic, and Z. Cucej, "Source coding technique for energy efficient IEEE 802.15.4 wireless sensor networks," in *Proc. IEEE 18th Int. Conf. Softw., Telecommun. Comput. Netw.*, Sep. 2010, pp. 71–75.

[5] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Proc. IPSN*, 2003, pp. 162–177.

[6] D. U. Maheswari, M. Meenalochani, and S. Sudha, "Influence of the cluster head position on the lifetime of wireless sensor network—A case study," in *Proc. IEEE 2nd Int. Conf. Next. Gen. Comput. Technol.*, Oct. 2016, pp. 378–381.

[7] J. Guo, E. Koyuncu, and H. Jafarkhani, "A source coding perspective on node deployment in two-tier networks," *IEEE Trans. Commun.*, vol. 66, no. 7, pp. 3035–3049, Jul. 2018.

[8] E. Ekici, S. Vural, J. McNair, and D. Al-Abri, "Secure probabilistic location verification in randomly deployed wireless sensor networks," *Ad Hoc Netw.*, vol. 6, no. 2, pp. 195–209, 2008.

[9] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Trans. Inf. Theory*, vol. TIT-23, no. 3, pp. 337–343, May 1977.

[10] F. Kiani, E. Amiri, M. Zamani, T. Khodadadi, and A. A. Manaf, "Efficient intelligent energy routing protocol in wireless sensor networks," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 3, 2015, Art. no. 618072.

[11] S. G. Neogi, A. A. Bhaskar, and P. Chakrabarti, "Energy efficient hierarchy-based clustering routing protocol for wireless sensor networks," *Int. J. Comput. Appl.*, vol. 95, no. 13, pp. 1–8, 2014.

[12] L. Zheng, K. Zhang, and D. Zhang, "An energy efficient dynamic address encoding scheme for wireless sensor network," *Int. J. Distrib. Sensor Netw.*, vol. 10, no. 3, 2014, Art. no. 476562.

[13] M. J. Golin and G. Rote, "A dynamic programming algorithm for constructing optimal prefix-free codes with unequal letter costs," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1770–1781, Sep. 1998.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[15] M. A. Perillo and W. B. Heinzelman, "Wireless sensor network protocols," in *Fundamental Algorithms and Protocols for Wireless and Mobile Networks*. Boca Raton, FL, USA: CRC Press, 2005.

[16] T. J. Swamy, T. Sandhya, and G. Ramamurthy, "Link evaluation of uniform grid based wireless sensor networks to base station with leveling and clustering," in *Proc. Int. Conf. Comput. Commun. Inform.*, Jan. 2013, pp. 1–5.