A Sybil Attack Detection Scheme for a Centralized Clustering-based Hierarchical Network

Mian Ahmad Jan, Priyadarsi Nanda, Xiangjian He
School of Computing and Communications
University of Technology, Sydney
Australia

Ren Ping Liu
Wireless and Networking Laboratory
CSIRO, Sydney
Australia

***Abstract*—Wireless Sensor Networks (WSNs) have experienced phenomenal growth over the past decade. They are typically deployed in remote and hostile environments for monitoring applications and data collection. Miniature sensor nodes collaborate with each other to provide information on an unprecedented temporal and spatial scale. The resource-constrained nature of sensor nodes along with human-inaccessible terrains poses various security challenges to these networks at different layers. In this paper, we propose a novel detection scheme for Sybil attack in a centralized clustering-based hierarchical network. Sybil nodes are detected prior to cluster formation to prevent their forged identities from participating in cluster head selection. Only legitimate nodes are elected as cluster heads to enhance utilization of the resources. The proposed scheme requires collaboration of any two high energy nodes to analyze received signal strengths of neighbouring nodes. The simulation results show that our proposed scheme significantly improves network lifetime in comparison with existing clustering-based hierarchical routing protocols.**

***Index Terms*—Wireless Sensor Network, Sybil Attack, Base Station, Cluster, Cluster Head**

I. Introduction

Wireless Sensor Networks (WSNs) consist of small sensor nodes working together to monitor and obtain data about an environment [1]. Sensor nodes have limited resources in terms of energy, computation, storage, transmission range and available bandwidth. They are typically deployed in a remote or hostile location and are left unattended to perform monitoring and reporting tasks. Therefore, limited resources of nodes need to be utilized efficiently in order to prolong network lifetime and obtain better throughput. These networks have been successfully deployed in a wide range of applications such as military surveillance, health care and environmental monitoring are few to mention [2].

Most WSNs are deployed for mission-critical tasks for an unspecified duration of time [3]. Therefore, security considerations need to be in place at the time of network design. The resource-constrained nature of these networks coupled with their unique characteristics, such as dynamic topology, in-network processing, error-prone communication links and scalability makes security provisioning challenging and complicated. In addition, these networks are left unattended without human intervention and base station supervision. Instead, sensor-collected data is harvested intermittently by a base station [4]. Since data are retained on individual sensors, securing these data is both important and challenging. Sensor nodes operating in unattended environments face a higher risk of security breaches. If any one of these nodes is compromised, its sensitive data and security parameters will be retrieved by an adversary to participate in malicious activities.

These networks face a diverse range of security challenges at various layers. For example, it is very challenging to detect Sybil attacks, where an adversary forges fake identities to legitimate nodes. An adversary may either fabricate such identities or steal them from legitimate nodes by disabling them permanently [5]. A single physical node may forge multiple identities to influence the outcome of data aggregation, fair resource allocation and voting on suspicious nodes [6].

The unstructured and distributed environment along with broadcast nature of communication in WSNs suits well to Sybil attacks. Various protection mechanisms have been developed to guard nodes against this type of attacks. A voting-based protection approach allows the nodes to determine if identities of a suspected node are legitimate or not [5]. However, a Sybil node may use its forged identities to vouch for each other and influence the outcome of voting. In [7], an authentication-based detection scheme was proposed which allowed the nodes to use certificates and shared encryption keys to detect Sybil attacks. The proposed scheme is computationally complex and requires significant resources on each node. In [8], the authors adopted a probabilistic approach using neighbourhood information. They argued that it was highly improbable for two nodes to have exactly the same set of neighbours in a densely deployed WSN. In [9], a *received signal strength indicator* (RSSI) approach was proposed which required coordination of at least four nodes to detect Sybil attacks.

All of detection techniques mentioned above are designed for data-centric routing protocols in which flooding is used to regulate traffic flow. Flooding allows intermediate nodes to broadcast data and control packets on their ways to base station from source nodes [10]. Duplicate packets keep circulating in the network which causes excessive energy consumption, delay, congestion, implosion and overlapping [11].

In this paper, we propose a lightweight Sybil attack detection scheme for a centralized clustering-based hierarchical network. Clustering approach significantly prolongs network lifespan by avoiding direct communication among nodes and a base station [12]. Our proposed scheme has two main objectives. First, we design a Sybil attack detection scheme which requires collaboration of only two nodes. Second, we implement our scheme for a centralized clustering-based hierarchical network to prevent Sybil nodes from participating in cluster head selection as these nodes are capable of forming multiple virtual clusters using their forged identities. Therefore, our proposed approach is lightweight in terms of Sybil attack detection and efficient in terms of prolonging network lifetime, cluster head selection, energy consumption, packet loss rate and packet delivery ratio.

The rest of the paper is organized as follows. In Section II, related work from literature for Sybil attack detection and clustering-based hierarchical routing protocols is provided. In Section III, we present a brief description of our proposed scheme followed by experimental work in Section IV. Finally, the paper is concluded and directions for future research are provided in Section V.

II. Related Work

In this section, we provide related research works on Sybil attack detection and applications to clustering-based hierarchi-

cal routing protocols using our proposed scheme.

Low-Energy Adaptive Clustering Hierarchy (LEACH) [13] was designated as a pioneer protocol among clustering-based hierarchical routing protocols. LEACH partitions a sensor field into small geographical regions known as clusters. Each cluster has a cluster head node which collects and aggregates data from member nodes and transmits to a base station. The protocol operates in rounds and nodes take turn to become cluster heads in subsequent rounds for uniform distribution of energy load. The problem with LEACH protocol is the probabilistic selection of cluster heads using random number generation. Each node, n, chooses a random number between 0 and 1. If this number is less than the threshold value, $T(n)$, defined in Equation 1, the node is elected as a cluster head for the current round.

$$T(n) = \begin{cases} \frac{k_{opt}}{1-k_{opt}(r mod(\frac{1}{k_{opt}}))}, & if\ n \in G, \\ 0, & otherwise. \end{cases} \quad (1)$$

Here, k_{opt} is the optimal number of cluster heads in each round, r is the current round and G is the set of nodes that have not been elected as cluster heads in the past $\frac{1}{k_{opt}}$ rounds. The probabilistic selection of cluster heads has a potential risk of low energy nodes being elected as cluster heads in subsequent rounds. Moreover, Equation 1 cannot guarantee an optimal number of cluster heads in each round. In [14], the authors argued that cluster heads need to be elected based on the residual energy of the nodes. They suggested the inclusion of residual energy of nodes in Equation 1. However, it will not solve the problem because cluster heads are still elected using a random number generation. To solve this problem, nodes need to be elected by a central controller or base station. In [15], the authors proposed a centralized approach for cluster head selection. Nodes having remaining energy greater than the average residual energy are elected as cluster heads in each round. However, it is highly probable that there will be a large number of such nodes in each round which will result in too many cluster heads. In [16], we proposed a centralized scheme which elected an optimal percentage of cluster heads (5 percent of total nodes). Each round results in balanced clusters which enhance network stability, scalability and data aggregation. Moreover, the proposed approach reduces network load, energy consumption and congestion.

In WSNs, communication over an error-prone wireless channel exposes nodes to various types of malicious activities. One of them is a Sybil attack, where an adversary forges multiple identities to mislead legitimate nodes into believing that they are having many neighbours. In [5], the authors proposed a radio resource testing approach for detecting forged identities. They assumed that a sensor node was incapable of simultaneous transmission or reception on a single radio. Moreover, a physical node may forge multiple identities but is incapable to use a single channel for these identities at a given time. Apart from radio resource testing, they also proposed a key validation approach for random key predistribution. However, it requires excessive resources of a node, is computationally complex and requires ample amounts of memory space. In [8], the authors proposed a scheme based on the assumption that probability of two nodes having exactly the same set of neighbours was extremely low provided that a network had high node density. They argued that forged identities typically had the same set of neighbours because they were associated with the same physical device. Therefore, presence of a malicious node can easily be detected by checking neighbourhood of the suspected victim of a Sybil attack. In [9], the authors proposed an RSSI-based solution for Sybil attack detection. They argued that even though RSSI was a time-varying parameter and unreliable in nature, using RSSI ratio from multiple receivers may be used for Sybil attack detection. In [17], the authors proposed an identity-based detection scheme for Sybil and spoofing attacks in IEEE 802.11 and WSNs. The proposed scheme uses a detector to identity malicious activities of malevolent entities capable of adjusting their transmission power. The detector locates the positions of these entities and eradicates them from network participation.

In WSNs, the detection techniques for Sybil attacks are typically designed for data-centric and location-based routing protocols. These protocols mostly rely on flooding [18] for routing data from source nodes to a base station. The broadcast nature of these protocols generates too many duplicate packets enroute to a base station. The above techniques for Sybil attack detection achieve their goal at the expense of excessive delay, congestion, packet duplication and energy consumption. To overcome the shortcomings of flooding, gossiping was proposed [19]. In gossiping, each node transmits a packet randomly to one of its neighbours. Gossiping ensures that each node receives a single copy of packet being sent. However, random selection of a neighbouring node is a risky task because a neighbour may be a Sybil node.

In this paper, we propose a lightweight scheme for Sybil attack detection. The proposed scheme requires coordination of any two high energy nodes and performs detection using signal strength of received packets. We use clustering-based hierarchical architecture to detect forged identities of an adversary. Each node transmits control packets to its two nearest high energy nodes. The control packets contain residual energy and identity of a node. Both high energy nodes calculate signal strength of the received packets and exchange it using a half-duplex communication channel to calculate RSSI ratio. After a certain amount of time, the same operation is performed to calculate a new RSSI ratio using signal strength of received packets from the same node. If the new ratio is equal to the one previously calculated and identities of the node in received packets are also different, it means that the node has forged its identities. Each node in the network undergoes a similar operation for identity verification. The goal of our proposed scheme is to prevent Sybil nodes from participation in cluster head selection. If a Sybil node is elected as cluster head, it will wreak havoc in a network by forming virtual clusters and using its forged identities as a cluster heads for each cluster.

III. Detection of Sybil Attack in a Centralized Clustering-based Hierarchical Network

In our proposed scheme, sensor nodes are classified according to their energy levels at the time of network deployment. Each node is either an ordinary sensing node or high energy node. The ordinary sensing nodes are equipped with 2 joules while high energy nodes are having 5 joules of energy. High energy nodes are of only 5 percent of total nodes to balance network cost and assist the base station in Sybil attack detection and in relaying vital information. In our scheme, only ordinary sensing nodes are eligible to be elected as cluster heads.

To minimize the consequences of a malicious activity, Sybil nodes are barred from cluster head selection. In doing so, network stability, efficiency and energy consumption are enhanced. If a Sybil node is elected as cluster head, it will exhaust network resources by forging multiple identities to legitimate nodes. In Figure 1, a Sybil node forges five different identities to sensor nodes in its vicinity. In WSNs, each node has the ability to adjust its transmission power to reach a far distant node [20]. A single Sybil node can form multiple clusters with each one of its identity as a separate cluster head.

In a distributed LEACH protocol, the base station has little control over cluster formation and cluster head selection. Each node (including Sybil nodes) can elect itself as a cluster head based on a generated random number, so LEACH is highly vulnerable to Sybil attacks. Each self-elected Sybil node can form multiple clusters, maliciously manipulating and aggregating data and transmitting to a base station. Instead, error-prone redundant data may be delivered to the base station at the expense of actual data [21]. High energy nodes have a vital role in Sybil attack detection and are constantly monitored by a base station. High energy nodes refrain themselves from cluster formation and participation in cluster head selection.

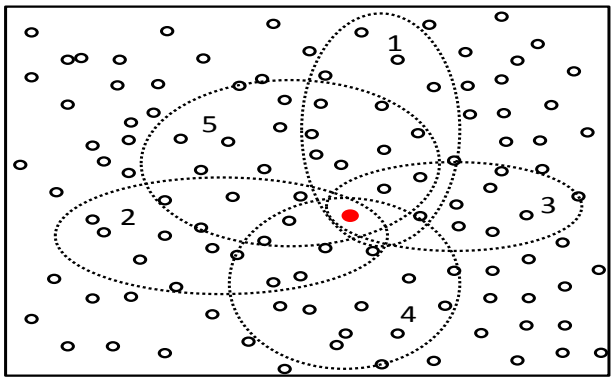

Fig. 1. A Single Sybil Node Forming Multiple Clusters

In view of the above discussion, we propose a novel Sybil attack detection scheme for a centralized clustering-based hierarchical network. Initially, high energy nodes identify Sybil nodes and report them to a base station to avoid their participation in cluster head selection. It ensures that only legitimate nodes can be elected as cluster heads in each round. Next, the base station elects an optimal number of cluster heads in each round based on an average energy threshold level. To the best of our knowledge, our proposed scheme is the first for Sybil attack detection in clustering-based hierarchical network. A brief overview of our proposed scheme is presented here.

A. Sybil Attack Detection

We use the concept of an RSSI for detection of Sybil attacks. A variable number of Sybil nodes (having 2 joule of energy) with multiple forged identities are injected before the start of each round. The objective of our scheme is to prevent their participation in cluster head selection.

Initially, each node broadcasts control packets to its two nearest high energy nodes as shown in Figure 2. This message contains its identity and residual energy. Theorem 5 in [22] argued that if atleast four sensor nodes monitor radio signals from a neighbouring node, it will not be able to hide its location. However, for a resource-constrained WSN, it is computationally complex task which requires abundant of network resources. To reduce the processing complexity, we propose a lightweight scheme for Sybil attack detection which requires coordination of any two high energy nodes.

Suppose that, high energy nodes, $hen1$ and $hen2$, receive control packets from node i at time t_1. If the identity of node i in control packets is x, then the RSSI, R^x_{hen1}, is calculated by $hen1$ using Equation 2.

$$R^x_{hen1} = \frac{P_t k}{d^\alpha_{hen1}}. \tag{2}$$

Here, P_t is the transmitted power, k is constant, d_{hen1} is the Euclidean distance between node i and $hen1$, and α is the

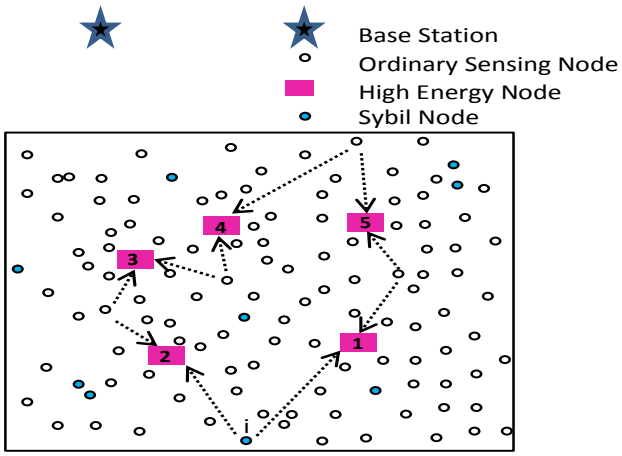

Fig. 2. High Energy Nodes Collaboration for Sybil Attack Detection

path-loss exponent. The value of α depends on the deployed environment. Its value is 2 for free-space, 1.6 to 1.8 for line-of-sight connection and 4 to 6 for buildings with obstructions [23]. The value of α for a free-space environment is computed using Equation 3. Here, λ is the wavelength of a radio signal.

$$\alpha(dB) = 10\log_{10}(\frac{\lambda}{4\pi d_{hen1}})^2. \tag{3}$$

The transmitted power, P_t, is related to received power, P_r, as shown in Equation 4.

$$P_t = \frac{P_r}{(1/d_{hen1})^\alpha}. \tag{4}$$

The location of node i with respect to $hen1$ can be computed by solving the Euclidean distance given in Equation 5.

$$d_{hen1} = \sqrt{(x_{hen1} - x_i)^2 - (y_{hen1} - y_i)^2}. \tag{5}$$

Solving Equations 3, 4 and 5 and substituting their values in Equation 2 enable $hen1$ to calculate the RSSI value, R^x_{hen1}. At this point, $hen1$ creates its own control packet and appends the value of signal strength, R^x_{hen1}, in it and transmits to its nearest high energy node, $hen2$. Recall that $hen2$ has received a similar control packet from node i at time t_1 and has calculated the value of R^x_{hen2} using a similar procedure as $hen1$. Next, $hen2$ calculates the radio signal strength ratio as shown in Equation 6.

$$\frac{R^x_{hen2}}{R^x_{hen1}} = (\frac{P_t k}{d^\alpha_{hen2}})/(\frac{P_t k}{d^\alpha_{hen1}}). \tag{6}$$

Further evaluation results in

$$\frac{R^x_{hen2}}{R^x_{hen1}} = (\frac{d_{hen1}}{d_{hen2}})^\alpha \quad and \quad t = t_1. \tag{7}$$

At time, $t_1 + t_0$, node i again broadcasts control packets with a different identity, y. High energy nodes, $hen1$ and $hen2$ perform similar operations as before and coordinate with each other to calculate the radio signal strength ratio at $hen2$ as shown in Equation 8.

$$\frac{R^y_{hen2}}{R^y_{hen1}} = (\frac{d_{hen1}}{d_{hen2}})^\alpha \quad and \quad t = t_1 + t_0. \tag{8}$$

At this point of time, $hen2$ compares the ratios obtained at time t_1 and $t_1 + t_0$. If the difference between these ratios is very close to zero as indicated in Equation 9, then $hen2$ concludes that a Sybil attack has occurred.

$$\frac{R^x_{hen2}}{R^x_{hen1}} - \frac{R^y_{hen2}}{R^y_{hen1}} \approx 0. \tag{9}$$

A single physical node, i has forged two identities, x and y, to its nearest high energy nodes at different time intervals. As the radio signal strength ratios are equal, location is in fact the same for alleged multiple identities. The complete process of Sybil attack detection is shown in Algorithm 1.

Algorithm 1 Detection of Sybil Attack

Input: E_i, n, ID_N, s, m, α, k
Output: {Sybil *or* non-Sybil}
`syb=round(rand(1)*s)+1;` ▷ Sybil generation
`id=round(rand(1)*m);` ▷ Generate m identities
▷ Next, each node is associated with high energy nodes
for $i = 1$ `to` N **do** ▷ N=n+s
  for $b = 1$ `to` 5 **do** ▷ Five high energy nodes
    Calculate Euclidean distance between i and b, d_i^b
    Sort d_i in ascending order to get two nearest high energy nodes, b' and b'', where b', $b'' \in b$
  At time, t_1
    `SEND` $(E_i, ID_i), \forall\, i \in N$ ▷ Each node sends its control packets to b' and b''
    `Calculate` $R_{b'}$ ▷ Check identity of i
    `Calculate` $R_{b''}$ ▷ Check identity of i
  ▷ $R_{b'}$, $R_{b''}$ are the received signal strengths at b' and b''
    b' `transmits` $R_{b'}$ `to` b''
    `Calculate` $R_{b''}/R_{b'}$ ▷ Calculated at b''
  At time, $t_1 + t_0$
    `Repeat step 9-13` ▷ Check identity of i
    `Compare ratios` ▷ Obtained at time, t_1 and $t_1 + t_0$
    if *Ratios are equal* ***and*** *having similar identities for* i **then**
  Node i is **Sybil**
    else
  Node i is **non-Sybil**

B. Centralized Clustering-based Hierarchical Protocol

In Figure 2, each high energy node monitors its nearest neighbours for a possible Sybil attack. Upon detection, Sybil nodes are reported to a base station located outside a sensor field. Each high energy node creates a control packet containing residual energy and identities of ordinary sensing nodes along with forged identities of detected Sybil nodes and transmits to a base station which makes the final decision on cluster head selection. The base station maintains two queues, one for blacklisted Sybil nodes and one for ordinary sensing nodes. It monitors the status of both queues at regular intervals.

The procedure of Sybil attack detection is repeated at the start of each round before cluster formation and cluster head selection. Once a Sybil node is detected, it is blacklisted to withhold its participation in cluster head selection. Clearly, there is a trade-off between the cost of Sybil attack detection and energy consumption of high energy nodes. High energy nodes remain active before the start of each round to detect new Sybil nodes. Furthermore, they avoid communication with already blacklisted Sybil nodes to preserve their energy levels.

The base station evaluates residual energy of ordinary sensing nodes to derive an average energy threshold, E_{avg}, as shown in Equation 10.

$$E_{avg} = \sum_{i=1}^{i=n} \frac{E_i}{n}. \tag{10}$$

Here, n is the total number of ordinary sensing nodes in a network and is equal to N-s, where, N is the set of all nodes including s Sybil nodes and E_i is the residual energy of an ordinary sensing node.

An ordinary sensing node having residual energy greater than average energy threshold is eligible for cluster head selection. However, it is probable that there will be a large number of such nodes in each round. These nodes are potential candidates for cluster heads in a particular round as shown in Figure 3. It is the job of base station to elect a desire number of cluster heads among candidate nodes. In our scheme, the following criteria are used for cluster head selection:

- Residual energy of a candidate is greater than or equal to average energy threshold.
- Candidate is not elected as cluster head during the past $\frac{1}{p}$ rounds.
- Two or more candidates located in a same geographical region are evaluated based on their residual energy and previous history of selection.

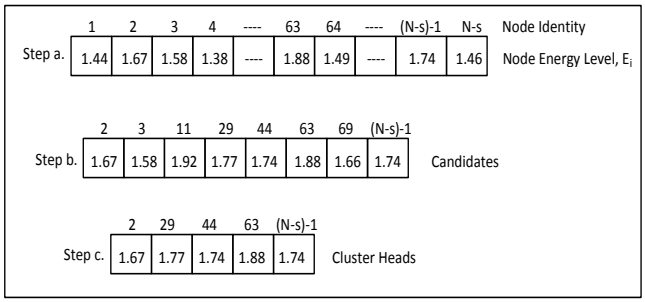

Fig. 3. Cluster Head Selection Procedure

In Figure 3, the base station stores residual energy of ordinary sensing nodes in a queue. It then computes the average residual energy, E_{avg} (1.5 joules in this case). An ordinary sensing node having residual energy greater than or equal to 1.5 joules is nominated as a candidate for cluster head. All candidates are evaluated according to the specified criteria. If two or more candidates are located in a same geographical region, they are evaluated according to their residual energies and their selections as cluster heads in the past $\frac{1}{p}$ rounds. For example, residual energy of node 2 is lower as compared to node 11 but the latter was elected in the past $\frac{1}{p}$ rounds in Figure 3.

In our proposed scheme, the optimal percentage of cluster heads are 5 percent for a network of 100 nodes. An optimal percentage of such nodes is one major factor that influences the performance of clustering-based hierarchical WSNs. A cluster head consumes more energy on aggregating data and relaying vital information to base station and performs general route maintenance and some other similar tasks [24]. If a small set of cluster heads are elected, network lifetime will degrade because these nodes will spend extra energy in data aggregation and long-haul transmission to base station. On the other hand, the selection of more cluster heads will make a clustering network rather inefficient and ineffective.

The base station elects an optimal percentage of cluster heads for a particular round and broadcasts nomination packets containing their identities. Each cluster head advertises itself to nearest neighbouring nodes which evaluate the received signal strength from multiple cluster heads. A neighbouring node associates itself with a cluster head having the strongest signal strength to form a cluster as shown in Figure 4. The selection of cluster head and formation of cluster is known as set-up phase. The completion of set-up phase is followed by initiation of steady-state phase during which each cluster head allocates time division multiple access (TDMA) slots within its cluster for sharing the transmission medium. This concept of slot allocation enables sensing nodes to remain inactive for most of their lifetime and at the same time avoids contention for transmission over a wireless link. Each cluster head collects data within its cluster in a particular round and transmits it to the nearest high energy node. A cluster

head may transmit data directly to a base station, however, in view of limited resources of a node, data is delivered to the nearest high energy node which in turn transmits to the base station. In doing so, energy load is uniformly distributed which prolongs the network lifetime. The flowchart of Figure 5 shows the complete process of set-up and steady-state phases for a particular round.

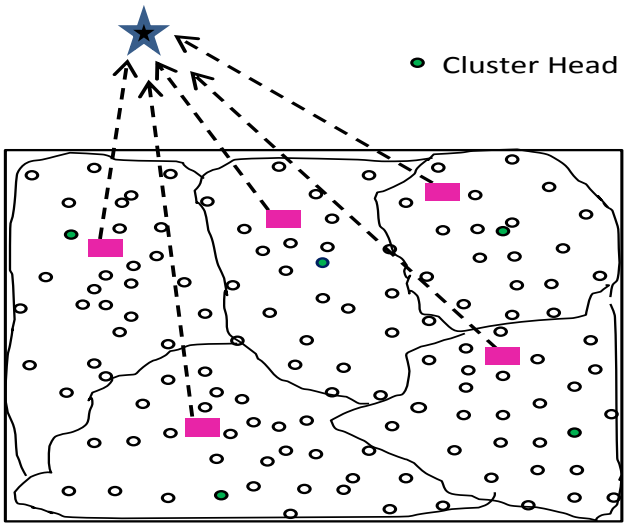

Fig. 4. Set-up and Steady-state Phases

Next, we calculate the energy consumption of various nodes during set-up and steady-state phases. The energy consumption of an ordinary sensing node in a particular cluster depends on its distance from its respective cluster head and is computed using Equation 11.

$$E(k, d_{CH}) = kE_{elec} + E_{amp}(k, d_{CH}). \tag{11}$$

Here, $E(k, d_{CH})$ is the energy consumption in transmitting k-bits packet to a cluster head over a distance d_{CH}, E_{elec} is the energy dissipated by radio of a node and E_{amp} is the energy dissipated by its amplifier in achieving an acceptable signal-to-noise ratio (SNR), $\frac{E_b}{N_0}$. The value of E_{amp} depends on the distance between an ordinary sensing node and its cluster head. If the distance, d_{CH}, is less than crossover distance, d_c, a free-space propagation model is used, otherwise a multipath fading model is used [16]. The values of $E(k, d_{CH})$ with respect to the free-space model and the multipath model are computed using Equation 12.

$$E_{Tx}(k, d_{CH}) = \begin{cases} kE_{elec} + k\epsilon_{fs}d_{CH}^2, & d_{CH} < d_c, \\ kE_{elec} + k\epsilon_{mp}d_{CH}^4, & d_{CH} \geq d_c. \end{cases} \tag{12}$$

Here, ϵ_{fs} and ϵ_{mp} are the energy consumption of an amplifier in the free-space propagation model and the multipath fading model respectively [16]. The crossover distance is computed using Equation 13.

$$d_c = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}}. \tag{13}$$

The free-space propagation model assumes an ideal condition for transmission in which there is a line-of-sight connection between an ordinary sensing node and its cluster head. In contrast, a radio signal reaches by two or more different paths in the multipath fading model due to reflection, refraction and obstacles between the nodes.

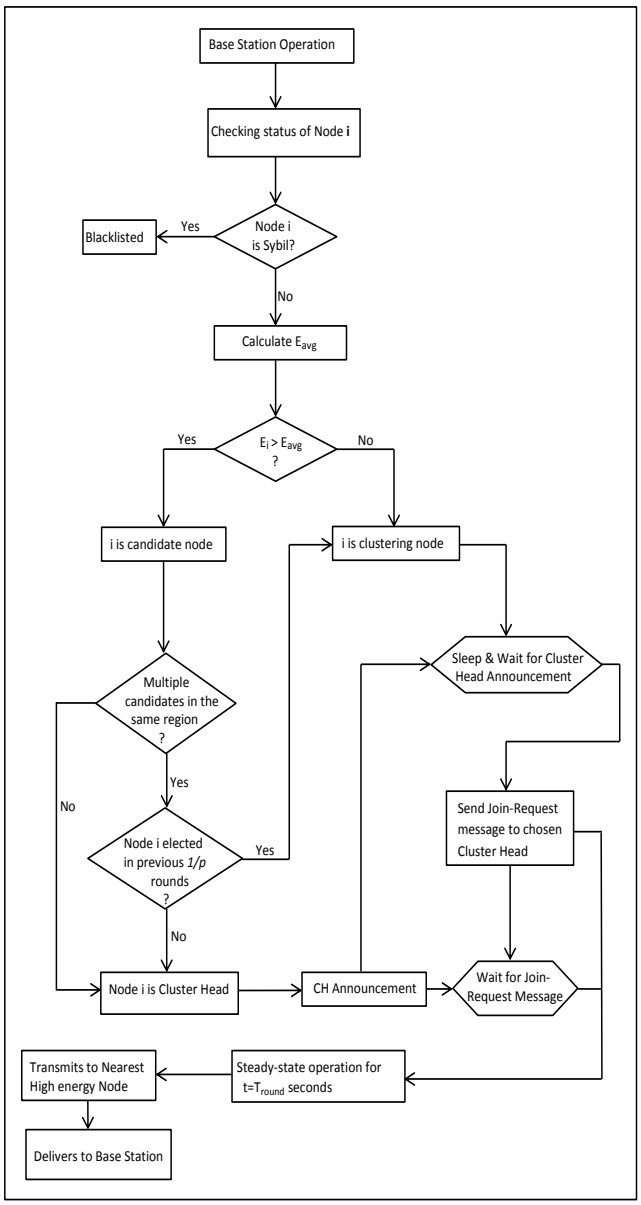

Fig. 5. Centralized Clustering-based Hierarchical Concept

Once a cluster head receives data from all its clustering members, it aggregates the data to reduce its size without compromising its quality. The aggregated data is further transmitted to a nearest high energy node for ultimate transmission to a base station. The energy consumption of a cluster head is significantly higher than a non-cluster head node and is calculated using Equation 14.

$$E_{CH} = \begin{cases} kE_{elec}\frac{n}{k_{opt}} + kE_{DA}\frac{n}{k_{opt}} + k\epsilon_{fs}d_{HEN}^2, d_{HEN} < d_c, \\ kE_{elec}\frac{n}{k_{opt}} + kE_{DA}\frac{n}{k_{opt}} + k\epsilon_{mp}d_{HEN}^4, d_{HEN} \geq d_c. \end{cases} \tag{14}$$

Here, E_{DA} is the energy consumption in data aggregation, k is the message size, k_{opt} is the optimal number of cluster heads and d_{HEN} is the distance between a cluster head and its nearest high energy node and its value determines the type of model (free-space or multipath) to be used by a cluster head in calculating its energy consumption. Our proposed scheme is based on the idea of balanced-clustering technique in which the number of cluster heads is equal to the number of clusters

[25]. Optimal number of cluster heads, k_{opt}, ensures that each round will have balanced clusters in which there will be one cluster head per cluster.

The energy consumption of high energy nodes differs from one to another. Before the start of each round, ordinary sensing nodes and Sybil nodes transmit their control packets to their two nearest high energy nodes at two different time intervals. All five high energy nodes are involved in computationally complex task of Sybil attack detection and the strength calculation of received signals based on the control packets.

In our scheme, hen1 calculates only the received signal strength values of incoming packets. However, the actual decision about the type (Sybil or non-Sybil) of a node is taken by hen2. Clearly, hen2 consumes more energy because of the additional task of finding the type of a node. Therefore, each high energy node is classified as either a received signal strength calculator ($rssc$) or a Sybil detector (sd). In view of the above discussion, hen1 is an $rssc$ while hen2 is an sd for node i mentioned in Section III. It is important to mention here that an sd node performs dual functionality of signal strength calculation and Sybil detection. The energy consumption of an $rssc$ node is calculated using Equation 15.

$$E_{rssc} = \begin{cases} E_{elec} \times 2\sum_{i=1}^{x} ctr_i + \epsilon_{fs}\sum_{i=1}^{x} ctrpk_i \; d_{nHEN}^2 \\ +k\epsilon_{fs}d_{CH-HEN}^2 + k\epsilon_{fs}d_{BS}^2, \qquad d_{nHEN} < d_c, \\ \\ E_{elec} \times 2\sum_{i=1}^{x} ctr_i + \epsilon_{mp}\sum_{i=1}^{x} ctrpk_i \; d_{nHEN}^4 \\ +k\epsilon_{mp}d_{CH-HEN}^4 + k\epsilon_{mp}d_{BS}^4, \qquad d_{nHEN} \geq d_c. \end{cases} \tag{15}$$

Here, ctr is the control packet sent by each node, x. d_{nHEN} is the distance between the $rssc$ and its nearest sd and d_{CH-HEN} is the distance between the cluster head and its nearest high energy node. Recall that each cluster head transmits its data to the nearest high energy node for ultimate transmission to the base station. Each node transmits two control packets to its two nearest high energy nodes to determine its type. Furthermore, each $rssc$ transmits a control packet to its nearest sd which contains the received signal strength value. The size of the control packet, ctr, is much smaller than the data packet, k.

The energy consumption of an sd node is calculated using Equation 16.

$$E_{sd} = \begin{cases} E_{elec} \times 2\sum_{i=1}^{x} ctr_i + E_{elec}\sum_{i=1}^{x} ctr_i + \epsilon_{fs}\sum_{i=1}^{i=x} \\ ctr_i d_{BS}^2 + k\epsilon_{fs}d_{CH-HEN}^2 + k\epsilon_{fs}d_{BS}^2, \quad d_{BS} < d_c, \\ \\ E_{elec} \times 2\sum_{i=1}^{x} ctr_i + E_{elec}\sum_{i=1}^{x} ctr_i + \epsilon_{mp}\sum_{i=1}^{i=x} \\ ctr_i d_{BS}^4 + k\epsilon_{mp}d_{CH-HEN}^4 + k\epsilon_{mp}d_{BS}^4, \quad d_{BS} \geq d_c. \end{cases} \tag{16}$$

The extra energy consumed by the sd node is due to the received control packets from its counterpart $rssc$ node. Furthermore, an sd node also consumes energy in transmitting control packets containing the identities and energy levels of ordinary sensing nodes along with the forged identities of detected Sybil nodes.

IV. Experimental Results and Analysis

In this section, we provide a series of simulation results for our proposed scheme. Our network comprises of n ordinary sensing nodes in a 100×100 square meter area. A variable number (denoted by s) of Sybil nodes with multiple identities are injected in each round. The Sybil nodes and ordinary sensing nodes have the same residual energy of 2 Joules. Table I shows the parameters of our experimental work.

TABLE I
Parameters and their Values

Simulation Parameters	Values
Number of ordinary sensing nodes, n	100
Number of high energy nodes, HEN	5
Energy consumed by electronic component, E_{elec}	50 nJ/bit
Energy consumed in data aggregation, E_{DA}	5 nJ/bit/packet
Energy consumed by amplifier in free-space, ϵ_{fs}	100 pJ/bit/m^2
Energy consumed by amplifier in multipath, ϵ_{mp}	0.013 pJ/bit/m^4
Length of data packet, k	2000 bits
Length of control packet, ctr	40 bits
Number of rounds, r	25000
Down-sampling rate of signal	500

Next, we evaluate our scheme in terms of number of detected Sybil nodes, total number of candidates and optimal selection of cluster heads, energy consumption, network lifetime, packet loss rate and packet acceptance ratio.

A. Detection of Sybil Nodes

In our proposed scheme, a random number of Sybil nodes having a variable number of forged identities are injected in the network before the start of each round. It is the job of high energy nodes to refrain Sybil nodes from cluster head selection. In Figure 6, the total number of Sybil nodes and their average number of forged identities detected in each round are shown.

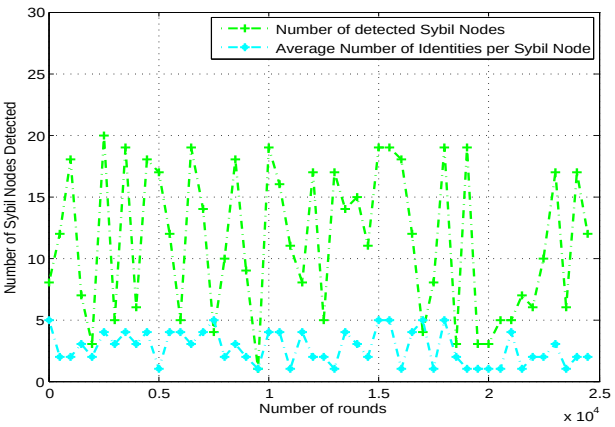

Fig. 6. Detection Rate of Sybil Nodes and their Forged Identities

In Figure 6, the number of Sybil nodes are as high as 20 and the average number of their forged identities have reached upto 5 in certain rounds. It would have an adverse impact on the outcome of voting, data aggregation and fair resource utilization if these nodes had gone undetected and were elected as cluster heads.

B. Total Number of Candidates and Cluster Heads

Our proposed scheme prevents Sybil nodes from participation in cluster head selection process. If the detection scheme is not in place, Sybil nodes may elect themselves as potential candidates for cluster heads as shown in Figure 7.

Figure 7 depicts the significance of high energy nodes in Sybil attack detection. In comparison with Figure 6, it is obvious that the majority of Sybil nodes are capable to nominate themselves as potential candidates for cluster heads. However, our detection scheme prevents their participation in cluster head selection. The performance of base station is highly precise and accurate because it elects only 5 cluster

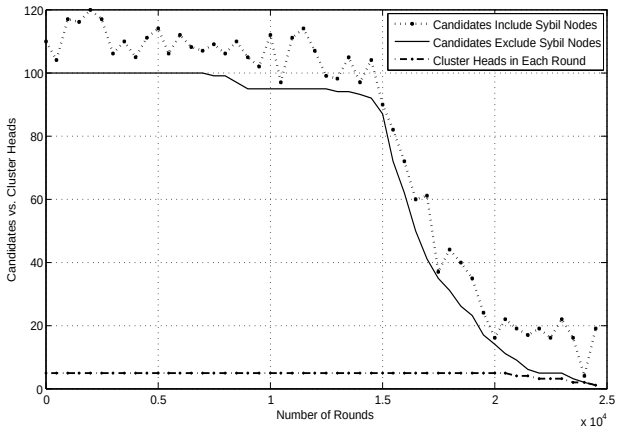

Fig. 7. Candidates vs. Cluster Heads

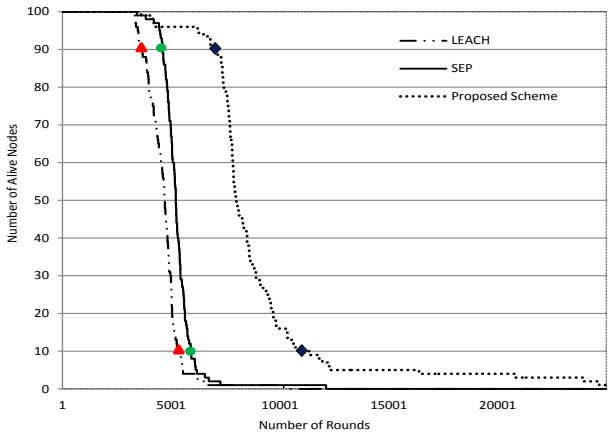

Fig. 9. Lifetime of the Network

heads in each round until the network has insufficient number of alive nodes toward the end. The optimal selection of cluster heads and prevention of Sybil nodes from participation in cluster head selection efficiently utilize energy of the nodes and enhance network lifetime.

C. Energy Consumption with Sybil Nodes

Total energy consumption of our scheme varies with the number of Sybil nodes and their forged identities in each round. In Figure 8, we calculate the amount of energy consumed in each round in presence and absence of Sybil nodes.

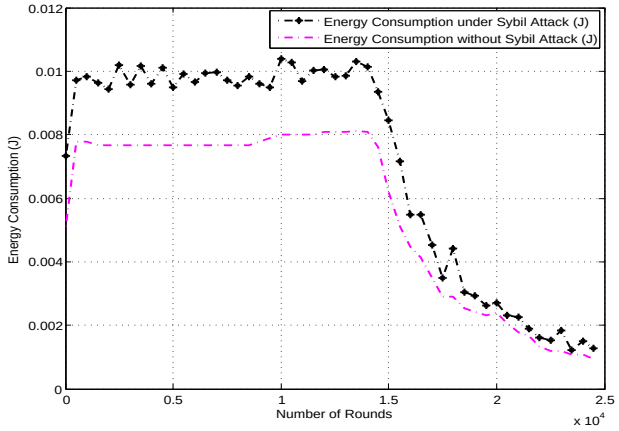

Fig. 8. Energy Consumption in Presence of Sybil Nodes

This increase in energy consumption is contributed much toward the control packets transmitted by Sybil nodes. Furthermore, locations of ordinary sensing nodes and Sybil nodes with respect to high energy nodes and base station have direct impact on energy consumption of the network.

D. Network Lifetime

The lifetime of a network is defined in terms of number of rounds it remains functional. In Figure 9, we compare the lifetime of our proposed network model with LEACH and Stable Election Protocol (SEP) [26]. Both LEACH and SEP randomly elect cluster heads using probabilistic threshold values and result in an excessive number of such nodes in various rounds.

In Figure 9, we define lifetime for clustering-based hierarchical networks in terms of rounds for two threshold values, i.e., 90 % and 10% of alive nodes. For 90% alive nodes, our network lifetime is 6989 rounds while LEACH and SEP have a lifetime of 3633 and 4599 rounds respectively. For 10% alive nodes, our network lifetime is 10822 rounds while LEACH and SEP have a liftime of 5363 and 5902 rounds respectively. Our simulation results show that our proposed scheme significantly improves network lifetime as compared to LEACH and SEP.

E. Packet Loss Rate

The traffic flow of our network is distributed in nature which allows us to use a random uniform model [27] to compute wireless transmission losses due to noise, interference and other channel impairments. The packet loss rate is the percentage of packets lost in the network over a specified duration such as the number of rounds.

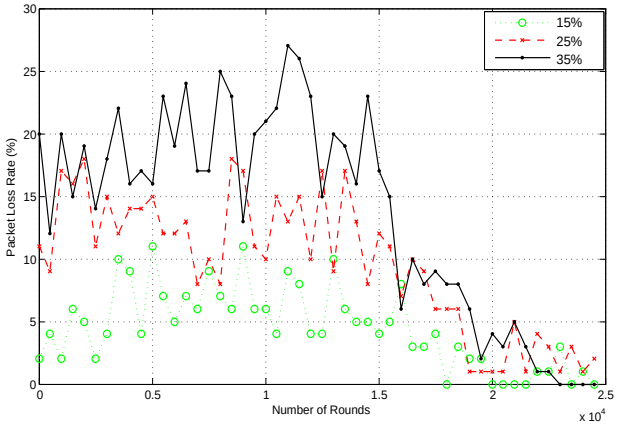

Fig. 10. Packet Loss Rate

The random uniform model shown in [27] calculates the probability of distributed packet losses with a mean value, p. Hence, we plot the packet loss rate for different values of p in Figure 10. From this figure, it is clear that the percentage of packet loss is higher at 35%. The mean value of p determines the quality of a network. In case of our network, the packet loss rate does not reach the threshold levels (15%, 25% and 35%) in most of the rounds which means that the network is sustainable and delivers most of the data required for decision-making at the base station.

F. Packet Acceptance Ratio

The packet acceptance ratio is defined as the number of packets successfully received at a base station to the number of transmitted packets. Packet acceptance ratio for our proposed network model is shown in Figure 11.

Packet acceptance ratio varies with the quality of communication links. The better the quality of links is, the higher the

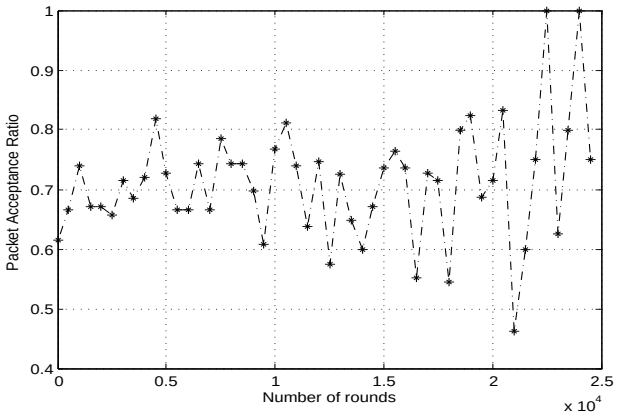

Fig. 11. Packet Acceptance Ratio

acceptance ratio is. Furthermore, it also depends on numerous other factors such as queuing capacity of cluster heads and high energy nodes, upstream traffic flow, data rate and interference are few to mention.

V. Conclusion

Wireless Sensor Networks (WSNs) have the ability of operating in remote, hazardous and hard-to-access locations. Tiny sensor nodes operating on small batteries are left unattended for an unspecified period of time for data acquisition and monitoring various applications. The absence of human intervention and remote monitoring of applications expose these networks to a wide range of security challenges at different layers. Among them, one such challenging threat is the presence of Sybil attack in which a single malicious node forges multiple identities to disrupt network operation. This type of attack is easily perpetrated in WSNs due to unstructured environment, distribute deployment of nodes and broadcast nature of data transmission. Furthermore, Sybil attacks do not require any specialized hardware but have the capability to wreak havoc by influencing the outcome of network operations such as data aggregation, voting and fair resource utilization.

In this paper, we have proposed a lightweight scheme for Sybil attack detection and its application to a centralized clustering-based hierarchical network. Our proposed scheme can detect Sybil nodes based on the signal strength of received packets. The collaboration of any two high energy nodes is required to determine the types of nodes in a network. These high energy nodes assist the base station in Sybil nodes detection and enable it to prevent such nodes from participation in cluster head selection. The candidate nodes for cluster heads are evaluated based on their residual energies, geographical locations and previous history of selection. In each round, an optimal number of cluster heads are selected using a balanced-clustering technique which enhances network lifetime and energy consumption. The proposed scheme can further be extended by incorporating mobility in it. The presence of mobile nodes in a network will require monitoring its coordinates on regular basis for calculating the Euclidean distance. Furthermore, mobile nodes in one cluster may belong to another cluster in the next round which may not result in balanced clusters. Moreover, a Sybil node will always try to sneak through the detection process. Currently, we are analyzing a scenario where an adversary somehow deceives high energy nodes and elects itself as cluster head.

References

[1] P. Rawat, K. D. Singh, H. Chaouchi, and J. M. Bonnin, "Wireless sensor networks: a survey on recent developments and potential synergies," *The Journal of supercomputing*, vol. 68, no. 1, pp. 1–48, 2014.

[2] J. A. Stankovic, A. D. Wood, and T. He, "Realistic applications for wireless sensor networks," in *Theoretical Aspects of Distributed Computing in Sensor Networks*. Springer, 2011, pp. 835–863.

[3] Y. Li and R. Bartos, "A survey of protocols for intermittently connected delay-tolerant wireless sensor networks," *Journal of Network and Computer Applications*, vol. 41, pp. 411–423, 2014.

[4] M. DI and G. Tsudik, "Security and privacy in emerging wireless networks," *IEEE Wireless Communications*, pp. 13–21, 2010.

[5] J. Newsome, E. Shi, D. Song, and A. Perrig, "The sybil attack in sensor networks: analysis & defenses," in *Proceedings of the 3rd international symposium on Information processing in sensor networks*. ACM, 2004, pp. 259–268.

[6] Y. Yu, K. Li, W. Zhou, and P. Li, "Trust mechanisms in wireless sensor networks: Attack analysis and countermeasures," *Journal of Network and Computer Applications*, vol. 35, no. 3, pp. 867–880, 2012.

[7] Y. Zhang, W. Liu, W. Lou, and Y. Fang, "Location-based compromise-tolerant security mechanisms for wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 2, pp. 247–260, 2006.

[8] K.-F. Ssu, W.-T. Wang, and W.-C. Chang, "Detecting sybil attacks in wireless sensor networks using neighboring information," *Computer Networks*, vol. 53, no. 18, pp. 3042–3056, 2009.

[9] M. Demirbas and Y. Song, "An rssi-based scheme for sybil attack detection in wireless sensor networks," in *Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks*. IEEE Computer Society, 2006, pp. 564–570.

[10] J. Zheng and A. Jamalipour, *Wireless sensor networks: a networking perspective*. John Wiley & Sons, 2009.

[11] N. A. Pantazis, S. A. Nikolidakis, and D. D. Vergados, "Energy-efficient routing protocols in wireless sensor networks: A survey," *Communications Surveys & Tutorials, IEEE*, vol. 15, no. 2, pp. 551–591, 2013.

[12] D. Gong, Y. Yang, and Z. Pan, "Energy-efficient clustering in lossy wireless sensor networks," *Journal of Parallel and Distributed Computing*, vol. 73, no. 9, pp. 1323–1336, 2013.

[13] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd annual Hawaii international conference on System sciences*. IEEE, 2000, pp. 1–10.

[14] M. A. Jan, P. Nanda, X. He, and R. P. Liu, "Enhancing lifetime and quality of data in cluster-based hierarchical routing protocol for wireless sensor network," in *IEEE 10th International Conference on High Performance Computing and Communications & IEEE International Conference on Embedded and Ubiquitous Computing (HPCC_EUC)*. IEEE, 2013, pp. 1400–1407.

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

[16] M. A. Jan, P. Nanda, and X. He, "Energy evaluation model for an improved centralized clustering hierarchical algorithm in wsn," in *Wired/Wireless Internet Communication*. Springer, 2013, pp. 154–167.

[17] Y. Chen, J. Yang, W. Trappe, and R. P. Martin, "Detecting and localizing identity-based attacks in wireless and sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 5, pp. 2418–2434, 2010.

[18] A. Hassanzadeh, R. Stoleru, and J. Chen, "Efficient flooding in wireless sensor networks secured with neighborhood keys," in *IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 2011, pp. 119–126.

[19] M. K. An, N. X. Lam, D. T. Huynh, and T. N. Nguyen, "Minimum latency gossiping in wireless sensor networks," in *Proceedings of the 2012 International Conference on Wireless Networks (ICWN)*, 2012.

[20] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 443–461, 2011.

[21] M. A. Jan, P. Nanda, X. He, and R. P. Liu, "Pasccc: Priority-based application-specific congestion control clustering protocol," *Computer Networks*, vol. 74, pp. 92–102, 2014.

[22] S. Zhong, L. Li, Y. G. Liu, and Y. R. Yang, "Privacy-preserving location-based services for mobile users in wireless networks," *Department of Computer Science, Yale University, Technical Report ALEU/DCS/TR-1297*, 2004.

[23] J. L. Burbank, W. Kasch, and J. Ward, *An introduction to network modeling and simulation for the practicing engineer*. John Wiley & Sons, 2011, vol. 5.

[24] J. Albath, M. Thakur, and S. Madria, "Energy constraint clustering algorithms for wireless sensor networks," *Ad Hoc Networks*, vol. 11, no. 8, pp. 2512–2525, 2013.

[25] S. A. Nikolidakis, D. Kandris, D. D. Vergados, and C. Douligeris, "Energy efficient routing in wireless sensor networks through balanced clustering," *Algorithms*, vol. 6, no. 1, pp. 29–42, 2013.

[26] G. Smaragdakis, I. Matta, and A. Bestavros, "Sep: A stable election protocol for clustered heterogeneous wireless sensor networks," Boston University Computer Science Department, Tech. Rep., 2004.

[27] C.-I. Kuo, C.-H. Shih, C.-K. Shieh, W.-S. Hwang, and C.-H. Ke, "Modeling and analysis of frame-level forward error correction for mpeg video over burst-loss channels," *Appl. Math*, vol. 8, no. 4, pp. 1845–1853, 2014.