Future Generation Computer Systems 80 (2018) 613–626

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

A Sybil attack detection scheme for a forest wildfire monitoring application

Mian Ahmad Jan [a,b], Priyadarsi Nanda [a,*], Xiangjian He [a,*], Ren Ping Liu [b]

[a] School of Computing and Communications, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

[b] Wireless and Networking Laboratory, CSIRO, Sydney, Australia

ARTICLE INFO

Article history:
Received 1 September 2015
Received in revised form
24 May 2016
Accepted 25 May 2016
Available online 2 June 2016

Keywords:
Wireless Sensor Network
Sybil attack detection
Wildfire monitoring
Queries
Cluster head
LEACH

ABSTRACT

Wireless Sensor Networks (WSNs) have experienced phenomenal growth over the past decade. They are typically deployed in human-inaccessible terrains to monitor and collect time-critical and delay-sensitive events. There have been several studies on the use of WSN in different applications. All such studies have mainly focused on Quality of Service (QoS) parameters such as delay, loss, jitter, etc. of the sensed data. Security provisioning is also an important and challenging task lacking in all previous studies. In this paper, we propose a Sybil attack detection scheme for a cluster-based hierarchical network mainly deployed to monitor forest wildfire. We propose a two-tier detection scheme. Initially, Sybil nodes and their forged identities are detected by high-energy nodes. However, if one or more identities of a Sybil node sneak through the detection process, they are ultimately detected by the two base stations. After Sybil attack detection, an optimal percentage of cluster heads are elected and each one is informed using nomination packets. Each nomination packet contains the identity of an elected cluster head and an end user's specific query for data collection within a cluster. These queries are user-centric, on-demand and adaptive to an end user requirement. The undetected identities of Sybil nodes reside in one or more clusters. Their goal is to transmit high false-negative alerts to an end user for diverting attention to those geographical regions which are less vulnerable to a wildfire. Our proposed approach has better network lifetime due to efficient sleep–awake scheduling, higher detection rate and low false-negative rate.

1. Introduction

Wireless Sensor Networks (WSNs) comprise of numerous miniature sensor nodes working together to monitor and collect data of interest [1]. These nodes are characterized by scarcity of their resources in terms of battery power, computation, storage and available bandwidth. They are typically deployed in a remote and hostile location to perform monitoring and data reporting tasks. Their limited resources need to be utilized efficiently to prolong network lifetime and throughput. These networks have found their applications in various domains such as automated irrigation system [2], telemonitoring system for healthcare [3], forest fire monitoring [4] and air pollution monitoring system [5].

Forest fires, also known as wild fires, are calamities which cause significant damages to human race and natural resources. These fires ignite because of environmental changes, human negligence or the combination of both. In countries like Australia, wildfires are frequently happening events because of its hot and dry climate. As a result of heat waves, extensive and fatal wildfires arise each year causing significant damages to infrastructure and human lives. On Saturday, 7th February 2009,[1] the state of Victoria saw its worst bushfires as a result of excessive high temperature, low relative humidity and high winds. These bushfires claimed 173 precious lives and caused significant damages to the infrastructure.

Wildfire terrains are hazardous and human-inaccessible, and require immediate reporting of time-critical and delay-sensitive events. WSNs are extensively used for wildfire monitoring to ensure that time-stamped events are instantly reported to a base station located outside a sensor field. Hefeeda and Bagheri [6] proposed the design of a WSN for an early wildfire detection system. They modelled the wildfire detection problem as a *k*-coverage problem [7] by employing an overpopulated static WSN. They used a distributed *k*-coverage algorithm for solving coverage issues within a forest to attain better accuracy. Sahin [8] proposed

* Corresponding authors.
E-mail addresses: Mian.A.Jan@student.uts.edu.au (M.A. Jan), Priyadarsi.Nanda@uts.edu.au (P. Nanda), Xiangjian.He@uts.edu.au (X. He), Ren.Liu@csiro.au (R.P. Liu).

http://dx.doi.org/10.1016/j.future.2016.05.034
0167-739X/

[1] (http://www.blacksaturdayfires.com/).

a wildfire detection scheme based on thermal data and animal behaviour. Sensor nodes were attached with animals to form a mobile biological WSN for detecting a possible wildfire on the basis of panic behaviour. García et al. [9] proposed a simulation environment for creating a model of a fire based on the analysis of captured data and geographical information. The main objective of this scheme was to automatically determine the geographical location and direction of a wildfire for transmission of alert signals to an end user. Dlamini [10] used Bayesian networks to determine abiotic, biotic and human factors which influence the occurrence of a wildfire. Furthermore, historical datasets were used to detect the possibility of a fire in remote geographical areas of Swaziland. Ha et al. [11] developed a simulation model for monitoring the possible outbreak of a wildfire. They designed an efficient cluster-based hierarchical protocol to reduce the overall energy consumption of sensor nodes. The protocol forms a multilayer cluster-based hierarchical network to reduce the number of transmission hops to a central base station. The proposed scheme reduces the delay incurred in transmission of sensitive alert packets from the nodes located in vicinity of a wildfire.

The use of WSNs for wildfire monitoring is a well-studied research topic and there exists lot of research in this context. However, all previous studies focused mainly on the improvement of QoS parameters of the collected data. Their main objective is to collect time-critical sensitive data and report them to a centralized base station without further delay. None of the previous studies focuses on the security aspects of the network in general and the data collected from the network in particular. Like any other application, security provisioning is a major challenging issue in wildfire monitoring application. The resource-constrained nature of WSNs coupled with the remote and intimidating terrains of a forest makes security provisioning become a daunting challenge. Furthermore, WSNs are left unattended after initial deployment. In the absence of human intervention, an adversary may capture critical alert packets, maliciously manipulates them and transmits to a base station. It may transmit false-negative alerts to a base station in order to mislead it about a particular geographical region. In doing so, attention of an end user is diverted to those regions which are less vulnerable to a possible wildfire. The absence of human intervention along with remote and in-accessible terrains of a forest may allow an adversary to forge multiple illicit identities at a given time, i.e., a Sybil attack, to influence the outcome of any decision taken by an end user.

In this paper, we propose a Sybil attack detection scheme for a wildfire monitoring application using a cluster-based hierarchical architecture. The major contributions of this paper are summarized below.

1. A two-tier detection scheme is proposed which operates at two levels. Initially, Sybil identities are detected by the high-energy nodes at a lower level. However, due to the error-prone communication links within a forest, one or more identities of the Sybil nodes may sneak through the detection mechanism employed at high-energy nodes. To detect such identities, we employ a higher level detection mechanism at the two base stations. The ultimate objective of a two-tier detection scheme is to prevent the participation of Sybil nodes in cluster head selection. Although the sneaked Sybil nodes are prevented by the base stations from participation in cluster head selection, these nodes can still participate in network communication as non-cluster heads. They associate themselves with one or more cluster heads as member nodes in order to transmit their data to the base stations. These sneaked Sybil nodes provide high false-negative readings to their respective cluster heads in order to mislead an end user. The purpose of any false-negative reading is to detract the attention of the base stations or an end user from highly sensitive regions to those which are less vulnerable to a wildfire within a forest. The two base stations ensure that the false-negative readings of sneaked Sybil nodes are detected and discarded before reaching an end user.
2. An energy consumption-based Sybil attack detection scheme is proposed in which one or more forged identities of a Sybil node are detected based on their energy consumption.
3. After Sybil nodes detection, the base stations elect an optimal percentage of cluster heads to collect time-critical and delay-sensitive alert packets within a forest. The elected cluster heads are advertised through nomination packets. Each nomination packet contains the identity of a cluster head and an end user's specific query. Query is a simple and declarative programming logic enabling an end user to collect data according to predefined conditions [12]. Once these conditions are met, each node wakes up, senses the environment, collects data and transmits to a nearest base station. Queries efficiently schedule the duty cycling of each node and as a result, reduce their energy consumption.
4. Two different types of queries, i.e., spatial and on-demand, are used to collect data. The assignment of these queries depends on the requirements of end user and the nature of collected data.

The remainder of the paper is organized as follows. In Section 2, related works from literature for Sybil attack detection, wildfire monitoring and cluster-based hierarchical networks are provided. In Section 3, we present the design considerations for a wildfire monitoring application. In Section 4, we provide a detailed explanation of our proposed scheme of Sybil attack detection in a forest wildfire monitoring environment. Our experimental work is provided in Section 5. Finally, the paper is concluded and directions for future research are provided in Section 6.

2. Related work

In this section, we provide the related research works on wildfire monitoring, Sybil attack detection and cluster-based hierarchical networks. An adversary carries out Sybil attacks on the routing layer of a WSN, hence, we choose cluster-based hierarchical architecture as the underlying platform because cluster head selection and cluster formation are part of network layer. Another reason for choosing these protocols is their energy-efficient nature in data transmission and communication [13]. In this section, we provide existing literature on cluster-based hierarchical networks followed by wildfire monitoring schemes and Sybil attack detection techniques.

In cluster-based hierarchical routing protocols, such as *Low-Energy Adaptive Clustering Hierarchy* (LEACH) [14] and *Stable Election Protocol* (SEP) [15], the nodes organize themselves into clusters. A cluster head gathers data from member nodes within each cluster, aggregates it and relays it back to a remote base station. Within a cluster, the neighbouring member nodes may detect similar data. In order to reduce data redundancy, each cluster head performs local data aggregation to eliminate similar data patterns which enhances the quality of data required for decision-making [16]. These protocols operate in rounds, and, in each round, a new set of cluster heads are elected for a uniform distribution of energy load within a network.

In literature, there exist various works on wildfire monitoring using an underlying cluster-based hierarchical platform. Ha et al. [11] proposed *Energy-efficient Fire Monitoring Protocol* (EFMP), which operates in three states namely, watch, slave and master. In a watch state, the nodes observe the detection of a possible wildfire. Among all the cluster heads, the one which first detects a fire, known as master head, transforms itself into master state. The master head informs all other cluster heads which transform themselves to slave state in order to act as slave heads within the

network. In this fashion, a layered hierarchical architectural model is formed. Irrespective of master or slave head, all the cluster heads collect data within their clusters. However, only a master head is eligible to transmit the data to a base station. All slave heads transmit their data to a master head which aggregates the data and transmits to a base station. EFMP reduces energy consumption because of master and slave head concepts.

Aslan et al. [4] proposed a simulation framework to monitor and detect a possible forest fire. The operation of the nodes depends on terrain, current weather forecasts and season of the year. Upon network deployment, the sensor nodes associate themselves with their nearest cluster heads. Each cluster head assigns transmission slots to member nodes to avoid contention for transmission on a wireless link. Furthermore, each cluster head has the ability to transform the member nodes into sleep mode in a Round Robin fashion to minimize their energy consumption. In the event of a fire detection, the nodes in close vicinity of a fire alter their normal transmission patterns and react more aggressively. If a fire ignites near a cluster head, it needs to immediately elect the most suitable member node as a replacement cluster head.

Zhang et al. [17] proposed a cluster-based hierarchical WSN to detect a possible wildfire. Sensor nodes were deployed to measure relative humidity and temperature readings within a forest. The cluster heads collect alarm packets from member nodes and transmit to a gateway node which ultimately delivers it to a centralized monitoring computer.

Yoon et al. [18] proposed a reliable wildfire monitoring system for a sparsely deployed WSN. The proposed scheme is reactive [19] in nature because the nodes remain in sleep mode and wake up only when an event is detected. While awake, each node remains in normal or in an alert mode for transmission of captured data without further delay. Moreover, the authors proposed separate routing paths for normal data and delay-sensitive data.

Ballari et al. [20] studied the behaviour of nodes within a forest. The authors proposed a mobility constraint model for providing adequate coverage to such events. They argued that the risk factors of a possible wildfire ignition characterize the coverage density of the nodes. Those areas which are more vulnerable to a possible wildfire occurrence require higher coverage densities. To provide accurate readings of the happening events, the nodes need to move towards hotspots to ensure a complete network coverage.

In WSNs, communication over an error-prone wireless channel exposes nodes to various types of malicious activities. Among them is Sybil attack where an adversary forges multiple identities at a given time to mislead legitimate nodes into believing that they are having many neighbours [21].

Newsome et al. [22] 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 the random key pre-distribution. However, it requires excessive resources on part of each node and is computationally complex requiring ample amount of memory space.

Ssu et al. [23] 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 a high node density. They argued that forged identities typically had the same set of neighbours because they were all associated with the same physical device. Therefore, the presence of a malicious node can easily be detected by checking the neighbourhood of a suspected victim of a Sybil attack.

Demirbas and Song [24] proposed a *Received Signal Strength Indicator* (RSSI) based solution for a Sybil attack detection. They argued that although an RSSI was a time-varying parameter and unreliable in nature, using RSSI ratio from multiple receivers might be used for a Sybil attack detection.

Chen et al. [25] proposed an identity-based detection scheme for Sybil and spoofing attacks in WSNs. Their proposed approach uses a detector to identity malicious activities of the malevolent entities capable of adjusting their transmission power. The detector locates the positions of such entities and prevent them from network participation.

All these existing schemes focus either on wildfire monitoring with an underlying cluster-based hierarchical platform or a Sybil attack detection scheme used with flat network topology. The existing wildfire monitoring techniques do not address any security challenges incurred within a forest while a flat topology suffers from flooding and implosion within a network [19]. These shortcomings motivate us to fill the research gap by proposing a Sybil attack detection scheme for a wildfire monitoring application. We propose two different detection techniques for a possible Sybil attack followed by a centralized cluster head selection approach. Similar to *Distributed Energy-Efficient Clustering with Improved Coverage* (DEECIC) algorithm [26], the residual energy of each node plays a crucial role in cluster head selection. Apart from the residual energy, our cluster head selection technique relies on the previous history of selection and geographical location of each node. We use different types of threshold-based queries to collect data within each cluster. Mobility is provided by the base stations to enhance network coverage and avoid any hotspot issues.

3. Design considerations for a wildfire monitoring application

First, we provide a brief overview of various environmental parameters which influence the behaviour and characteristics of a wildfire within a forest. Next, we explain various network parameters and design consideration which govern the data collection capabilities of the sensor nodes.

3.1. Characteristics of burning wildfire scenario

The behaviour of a wildfire is characterized by the following factors [27]: *Fuels*, *Weather Conditions* and *Topography*. These three factors determine how quickly a wildfire can vanish or turn into a raging blaze which may scorch thousands of acres of land in a short period of time.

The fuels include grasses, dry leaves, twigs, shrubs and branches of the trees. Small pieces of fuels burn quickly, particularly when they are larger in quantity, dry and loosely arranged.

The weather also plays a crucial role in igniting and spreading a wildfire. On a hot and windy day, fuels are at their driest which increases the risk of a wildfire. The three weather ingredients, i.e. temperature, relative humidity and wind speed have the ability to ignite and spread a wildfire to engulf a vast region. Temperature has a direct impact on fire ignition because the fuels are closer to their ignition points at a very high temperature. Each fuel has a *flash point*, a temperature reading at which it bursts into flames. The typical flash point for various types of dried fuels is[2] 300 °C. At flash point, the fuels release hydrocarbon gases that mix with oxygen in the air and cause wildfire due to combustion. Relative humidity is the percentage of moisture in the air. It has an adverse impact on the intensity and flammability of a wildfire. During hot summer days and dry conditions, humidity is relatively low and it adds to the possibility of a wildfire as the fuels do not have sufficient moisture in the air to absorb. The possibility of a wildfire

[2] http://science.howstuffworks.com/nature/natural-disasters/wildfire.htm.

increases when the relative humidity drops below 30% (*critical point*) in the air. Wind supplies oxygen which further dries the fuels and spreads the fire across a wide geographical region. The stronger the wind blows, the faster a wildfire will spread. The *threshold* wind speed of 12–15 km/h has a significant impact on the behaviour of a wildfire. Low relative humidity coupled with strong winds and high temperature readings rapidly spread a wildfire.

Topography or the slope of a land also influences the behaviour of a wildfire. Topography can either aid or hinder the progression of a wildfire because a fire spreads quickly and much faster up a slope and slows down as it goes down a slope.

Among these three factors, weather condition is highly crucial in igniting and spreading a wildfire. Furthermore, the behaviour of a fuel and topography are characterized by weather forecast as well. It is for this reason that we have formulated our network model based on the flash point, critical point and threshold values of the three weather ingredients. Temperature and relative humidity ignite a wildfire while the speed of wind facilitates in spreading it.

3.2. Network parameters and design consideration

In WSNs, the source nodes located in close vicinity of each other capture somewhat identical data packets [28]. If such data is transmitted to a base station, it may flood the whole network with multiple copies of the same data. To avoid data redundancy, an end user specifies various conditions for data transmission. In a wildfire monitoring application, an end user is mainly interested in threshold values of certain environmental parameters. Threshold values enable an end user in taking swift actions according to the environmental conditions.

In our proposed scheme, each normal node is equipped with three sensors for monitoring temperature, relative humidity and wind speed within a forest. Each node remains in sleep mode and wakes up only when the threshold conditions are satisfied by the captured data. The sleep–awake scheduling of our approach enhances network lifetime by reducing the energy consumption. Upon capturing the events of interest, they are locally processed and relayed back to a nearest base station. Similar to *Threshold-sensitive Energy Efficient-sensor Network* (TEEN) [29] protocol, the operation of these nodes is governed by hard and soft threshold values.

Hard threshold (H_T) is the minimum value of a sensed event which triggers a sensor to operate. A sensor node remains idle or in sleep mode until H_T is reached. For our proposed network model, H_T is set to 100 °C for temperature reading, 40% for relative humidity and 8 km/h for wind speed. These values of H_T are stored locally in each node as a reference for future readings. Recall that the temperature's flash point is 300 °C, relative humidity's critical point is below 30% and wind speed's threshold is 12–15 km/h. These values are the maximum threshold readings for environmental parameters at which a wildfire ignites and spreads across the forest. The end user needs to be informed well before the maximum threshold readings. It is for this reason that we have set the values of H_T for temperature, relative humidity and wind speed to inform an end user well before any emergency situation. Each node sends an alert packet to a nearest base station when H_T is reached for these parameters. The alert packets are constantly transmitted to keep an end user up-to-date about the current status of a forest. In a wildfire monitoring application, an end user is not interested in an incoming alert packet for which the H_T reading is similar to the previous one.

To ensure that the incoming alert packet has a different reading than the previous ones, we set a soft threshold (S_T) for each parameter. S_T is an incremental change in the value of a sensed event which triggers the transmitter of a sensor to capture packets and relay them to a nearby base station. In our proposed approach, S_T is set to 40 °C for temperature, 2% for relative humidity and 1 km/h for wind speed.

Therefore, the transmitter of each sensor will be triggered for the first time when H_T is 100 °C for temperature, 40% for relative humidity and 8 km/h for wind speed. In the second time, it will be triggered when H_T is 140 °C for temperature, 38% for relative humidity and 9 km/h for wind speed. In the third time, it will be triggered when H_T is 180 °C for temperature, 36% for relative humidity and 10 km/h for wind speed. This process of adjusting the values of H_T continues and alert packets are transmitted to a nearest base station on regular intervals. The value of S_T is added to the previous H_T value in order to obtain a new H_T value for the current round. The threshold values of H_T and S_T provide sufficient time for an end user to take precautionary measures in an emergency situation within a forest.

4. Detection of Sybil attack in a forest wildfire monitoring application

In this section, we provide a detailed explanation of our proposed scheme. First, the architectural model of our network is presented followed by its deployment model. Next, two different approaches for Sybil attack detection are presented. Sybil nodes are detected prior to cluster formation and cluster head selection. This ensures that only legitimate nodes are elected as cluster heads. Once Sybil nodes are barred from cluster head selection, a cluster-based hierarchical network is formed to obtain environmental data based on the conditions specified by an end user.

4.1. Network architectural model

The network architectural model of our proposed scheme is depicted in Fig. 1. In this figure, Sybil attack detection and cluster-based algorithm are supported at the network layer. During routing, the forged identities of Sybil nodes are detected to prevent their participation in cluster head selection. The three environmental parameters of temperature, relative humidity and wind speed are supported at the application layer. The goal of Sybil nodes is to sneak into the network and provide false-negative alert readings of these environmental parameters to an end user. Furthermore, the presence of lossy links and a hostile environment pose a potential threat that some of these Sybil nodes may sneak through the detection process due to varying signal strength at different time intervals. The escaped Sybil nodes are ultimately detected by the base stations to prevent their participation in cluster head selection. However, the escaped Sybil nodes are still eligible to participate in the network as *legitimate* non-cluster head nodes. This is because the normal nodes in the network have no idea about the identities of other normal nodes or Sybil nodes. However, the malicious data of Sybil nodes are ultimately discarded by the base stations to prevent its usage during critical decision-making.

4.2. Network deployment model

Our proposed network model consists of 200 normal nodes and 20 high-energy nodes deployed in a wide geographical region of 100×100 square metre area. The normal nodes are equipped with 2 joules while high-energy nodes are having 5 joules of residual energy. High-energy nodes assist the base stations in Sybil attack detection and relay back vital information. The number of nodes in the network affects the quality of delivered data and security provisioning. With the increase of the number of nodes, the number of transmitted packets increases which ultimately leads to higher end-to-end delay, retransmission attempts, packet loss,

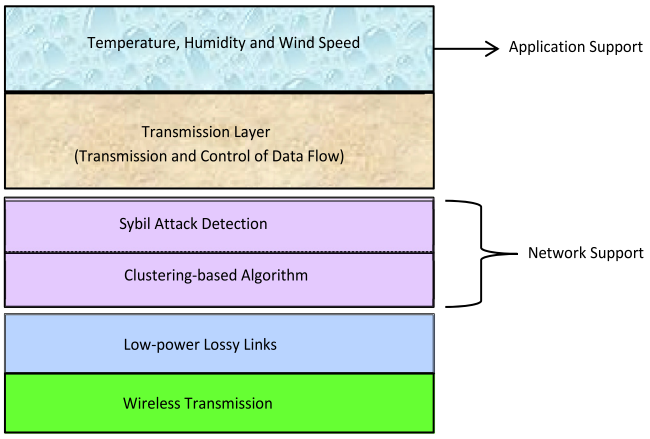


Fig. 1. Network architectural model.

in-network processing, network congestion and QoS degradation. These metrics increase the exploitation of network resources and energy consumption of the nodes. Moreover, securing a network of 200 normal nodes from various attacks incurs less overhead as compared to a network of 1000 normal nodes.

We use joint-sink mobility [30], in which two base stations are used to avoid hotspot problems within the geographical region. Both of these base stations support *to-and-fro* motion to cover a subset of the nodes. Base station 1 moves horizontally between the coordinates (25, 25) and (75, 25), while base station 2 moves horizontally between the coordinates (25, 75) and (75, 75) as shown in Fig. 2. Unlike a random waypoint mobility model [31] that suits mobile nodes, we are more interested in the sink mobility, also known as base station mobility. Our model does not support mobile nodes as it will bring too much fluctuation which ultimately affects the ratio of RSSI values. The fluctuation in RSSI enables Sybil nodes to sneak through the detection process and as a result, such nodes may be reported as normal, also known as legitimate nodes, to the base station. Both base stations provide sufficient coverage to the nodes by partitioning the geographical region equally between themselves. To-and-fro motion ensures that disjoint regions are covered and the source nodes no longer require long-haul transmissions to the two base stations.

In a wildfire monitoring and many other delay-sensitive applications, a single base station may not be sufficient to provide complete network coverage in a wide geographical region. In these applications, the source nodes generate critical events which need to be reported immediately to a base station [32]. Missing one or more such events may result in a wrong interpretation of the forest environment which may lead to catastrophic circumstances. A single base station may require constant movement with a predefined velocity to ensure that critical events are not lost [33]. The nodes located at extreme ends will require long-haul transmissions to reach the nearest base station. As a result, a considerable amount of energy on part of these nodes will be consumed. Such long-haul transmissions ultimately lead to hotspot problems [34], in which network connectivity and geographical coverage are seriously jeopardized. The consequences get worsen if the transmitted data is highly-prioritized. The transmission of such highly-prioritized data over long-haul communication channels will make the data useless if it does not reach within its time-stamp. In our proposed scheme, a series of critical events are detected based on the hard and soft thresholds of the environmental parameters. The two base stations constantly move around the sensor field to ensure that time-critical and delay sensitive events generated due to on-demand queries are not lost. Our mobility model avoids long-haul transmissions and ensures that the two base stations do not end up in the same geographical region.

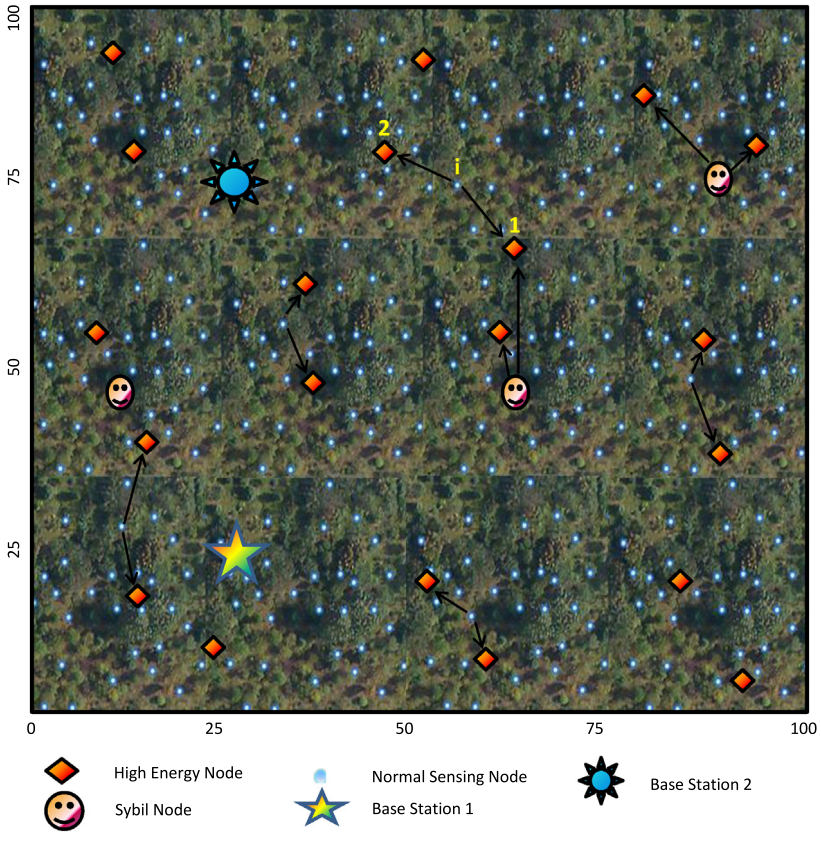


Fig. 2. RSSI-based Sybil attack detection.

4.3. Detection of Sybil attack

Before environmental monitoring to proceed, each node needs to authenticate itself to ensure that only legitimate nodes are 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. In WSNs, each node has the ability to adjust its transmission power to reach a far distant node [35]. Using this feature of WSNs, one or more Sybil nodes may cover a wide geographical region by forming multiple clusters with each one of their forged identities as a separate cluster head. In a wildfire monitoring application, Sybil identities may result in catastrophic circumstances by constantly providing vague results to an end user. If such identities are elected as cluster heads, they may discard time-critical and decision-making data of member nodes within each cluster. Instead, they may either fabricate their own data or transmit data from those member nodes which is less critical and may not be useful for an end user to take precautionary measures within a forest. If there is a high probability of wildfire in a particular geographical region, they may mislead an end user into believing that there is no such possibility of wildfire in that specific region. Furthermore, they may divert its attention to those areas which are less vulnerable to a wildfire.

In view of the above discussion, we propose two different techniques for Sybil attack detection in a forest wildfire monitoring application. Our first approach is based on RSSI of the transmitter nodes that enables the nearest two high-energy nodes to detect Sybil nodes and their forged identities. The proposed approach calculates the ratio of signal strength at two different time intervals and identifies Sybil nodes if the ratio is same for multiple identities. Our second approach is based on the residual energy of the nodes and detects Sybil nodes if the residual energy field of two or more control packets matches. The objective of both these approaches is to prevent Sybil nodes from electing themselves as cluster heads as they may wreak havoc in the network. High-energy nodes are assigned the task to detect such nodes and report them to a nearest base station to ensure that only legitimate nodes are elected as cluster heads in each round.

4.3.1. RSSI-based Sybil attack detection

In our proposed scheme, a variable number of Sybil nodes with multiple forged identities are injected before the start of each round. Both normal nodes and Sybil nodes are isomorphic in nature, i.e., having similar capabilities in terms of sensing, processing, communication and broadcasting. Each node broadcasts control packets to its two nearest high-energy nodes as shown in Fig. 2. This packet contains the identity and residual energy of each node.

Suppose that high-energy nodes, *hen*1 and *hen*2, receive control packets from node i at time t_1. At this point, high-energy nodes do not know whether the transmitter is a normal node or a Sybil node. They determine the type of a node based on the received signal strength. If identity of node i in control packets is x, then the received power (RSSI), R^x_{hen1}, is calculated by *hen*1 using Eq. (1).

$$R^x_{hen1} = \frac{P_t \cdot k}{d^{\alpha}_{hen1}}. \tag{1}$$

Here, P_t is the transmitted power, k is constant, d_{hen1} is the Euclidean distance between node i and *hen*1, and α is the path-loss exponent. The value of α depends on the deployed environment. Its value is 2 for free-space, is between 1.6 and 1.8 for buildings with line-of-sight connection, and is between 2.7 and 3.5 for urban area, respectively [36]. The value of α represents the signal level attenuation caused by free space propagation, reflection, diffraction and scattering. Furthermore, α varies with the distance between a receiver and a transmitter node.

The value of α for a free-space environment is computed using Eq. (2) [37]. Let λ_c be the wavelength of a radio signal and be equal to 3×10^8 m/s. Then,

$$\alpha = \left(\frac{4\pi d_{hen1}}{\lambda_c}\right)^2. \tag{2}$$

The transmitted power (P_t) is related to received power (P_r) as shown in Eq. (3).

$$P_t = \frac{P_r}{(1/d_{hen1})^{\alpha}}. \tag{3}$$

The location of node i with respect to *hen*1 can be determined by computing the Euclidean distance between i and *hen*1 using Eq. (4).

$$d_{hen1} = \sqrt{(x_{hen1} - x_i)^2 - (y_{hen1} - y_i)^2}. \tag{4}$$

Solving Eqs. (2)–(4) and substituting their values into Eq. (1) enable *hen*1 to calculate R^x_{hen1}. At this point, *hen*1 creates its own control packet and appends the value of signal strength, R^x_{hen1}, in it and transmits the packet to its nearest high-energy node, *hen*2. Recall that *hen*2 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 *hen*1. Next, *hen*2 calculates the radio signal strength ratio as shown in Eq. (5).

$$\frac{R^x_{hen2}}{R^x_{hen1}} = \left(\frac{P_t \cdot k}{d^{\alpha}_{hen2}}\right) \Big/ \left(\frac{P_t \cdot k}{d^{\alpha}_{hen1}}\right). \tag{5}$$

Further evaluation results in

$$\frac{R^x_{hen2}}{R^x_{hen1}} = \left(\frac{d_{hen1}}{d_{hen2}}\right)^{\alpha}, \quad \text{when } t = t_1. \tag{6}$$

At time $t_1 + t_0$, node i again broadcasts control packets with a different identity, y. High-energy nodes, *hen*1 and *hen*2, perform similar operations as before and coordinate with each other to calculate the radio signal strength ratio at *hen*2 as shown in Eq. (7).

$$\frac{R^y_{hen2}}{R^y_{hen1}} = \left(\frac{d_{hen1}}{d_{hen2}}\right)^{\alpha}, \quad \text{when } t = t_1 + t_0. \tag{7}$$

At this point of time, *hen*2 compares the ratios obtained at time t_1 and $t_1 + t_0$. If the difference between these two ratios is very close to zero, as shown in Eq. (8), *hen*2 concludes that a Sybil attack has occurred.

$$\frac{R^x_{hen2}}{R^x_{hen1}} - \frac{R^y_{hen2}}{R^y_{hen1}} \approx 0. \tag{8}$$

A single physical node, i, has forged multiple identities, x and y, to its nearest high-energy nodes at different time intervals. As the radio signal strength ratios are equal, it means that the locations are also the same for alleged multiple identities. The procedure for RSSI-based Sybil attack detection is shown in Algorithm 1.

The RSSI-based scheme detects Sybil nodes based on the received signal strength at high-energy nodes. The value of a signal strength is influenced by various factors such as reflection, refraction, physical obstacles, channel impairment, transmitter power, antenna type and distance between a transmitter and a receiver node. Furthermore, the strength of a signal also depends on Line-of-Sight (LoS) and Non-Line-of-Sight (NLoS) radio transmissions. In a typical forest environment, NLoS transmission along with channel impairment causes significant reflection and refraction of radio signals through various obstacles that result in highly fluctuating RSSI values. As a result, the quality of a signal strength is quite poor so that one or more Sybil nodes may

Algorithm 1 RSSI-based Sybil Attack Detection

Input: E_i, n, s, m, α, k ▷ E_i represents the residual energy of node i, n and s represent the number of normal nodes and number of Sybil nodes, respectively, m represents the maximum number of forged identities of each Sybil node, and α and k are those appeared in Equation 5.
Output: {Sybil *or* non-Sybil}
Initialization:

1. Node i is assigned E_i ▷ $\forall\, i \in \mathbb{N}$, where $\mathbb{N}$=n+s
2. Normal node i is assigned an identity, ID_i ▷ $\forall\, i \in n, ID_i \in \{ID_1, ID_2, \cdots, ID_n\}$
3. Generate s Sybil nodes: syb_j, where $j \in \{1, 2, \cdots, s\}$, and $s = \text{round}(\text{rand}(1) * s) + 1$; ▷ Generate a random number of s Sybil nodes in each round. At least, one Sybil node is generated in each round using rand and round functions.
4. Generate m forged identities: id_l, where $l \in \{1, 2, \cdots, m\}$, and $m = \text{round}(\text{rand}(1) * m) + 2$; ▷ Each Sybil node i has a random number of m identities. The value of 2 is added to obtain at least two identities for each Sybil node in each round.

for $i = 1$ `to` $\mathbb{N}$ **do**
  for $b = 1$ `to` c **do** ▷ c represents the number of high-energy nodes, c=20
    Calculate Euclidean distance, d_{ib}, between nodes i and the b-th high-energy node as shown in Equation 4.
    Sort d_{ib} in ascending order to get two nearest high-energy nodes, i.e., the b'-th and b''-th high-energy nodes, where $b', b'' \in \{1, 2, \cdots, c\}$.
    At time, t_1
    `SEND` $(E_i,\ ID_i)$ ▷ Node i sends its control packets to the b'-th and b''-th high-energy nodes, where, node i may be normal or Sybil.
    `Calculate` $R_{b'}$ ▷ Calculate the received power as shown in Equation 1
    `Calculate` $R_{b''}$ ▷ Calculate the received power as shown in Equation 1
    `Next, the` b'`-th high-energy node transmits` $R_{b'}$ `to the` b''`-th high-energy node`
    `Calculate` $R_{b''}/R_{b'}$ ▷ Calculate the ratio using Equation 5.
    At time, $t_1 + t_0$, Repeat Step 8–12
    `Compare both ratios` ▷ Obtained at times t_1 and $t_1 + t_0$
    if *Ratios are equal* ***and*** *having similar identities for node i* **then** Node i is **Sybil**
    else Node i is **non-Sybil**
    end if
  end for
end for

sneak through the detection process. The fluctuating values of the received signal strengths at time, t_1, and time, $t_1 + t_0$, may result in varying RSSI ratios. The differences in these ratios are sufficient to convince nearby high-energy nodes into believing that a transmitter is not a Sybil node. In this fashion, one or more identities of Sybil nodes go undetected and are reported to a nearest base station as *normal* nodes.

High-energy nodes have a fair detection policy that is equally applicable to both normal nodes and Sybil nodes. High-energy nodes have no prior knowledge if a node is normal or Sybil. For high-energy nodes, undetected Sybil nodes are also normal nodes. Furthermore, undetected Sybil nodes try to participate in cluster head selection as well. In fact, it is the ultimate goal of these nodes to be elected as cluster heads because, in that way, they may wreak havoc and disrupt the whole network operation. To prevent such catastrophic situation, we implement a two-tier detection process: one at high-energy nodes and the other at the two base stations. The Sybil nodes may sneak through the detection process at high-energy nodes due to the fluctuating RSSI values within a forest. However, it is highly improbable that such nodes may sneak through the detection process at the base stations. Unlike high-energy nodes, each base station maintains the identities of the normal nodes locally within a database. In case of Sybil nodes, there will be a mismatch of identities and they will be barred from participation in cluster head selection.

At first glance, it seems that the detection mechanism deployed at the high-energy nodes may not be necessarily required because the two base stations guarantee to detect each and every Sybil node. However, apart from Sybil nodes, there are 200 normal nodes in the network as well. It will put a lot of burden on these nodes in terms of resource consumption if they transmit their control packets directly to the two base stations. To avoid long-haul transmissions, we deploy the detection mechanism at high-energy nodes to reduce the resource consumption (i.e., energy consumption) of the nodes and network, end-to-end delay, packet loss and congestion.

4.3.2. Residual energy-based Sybil attack detection

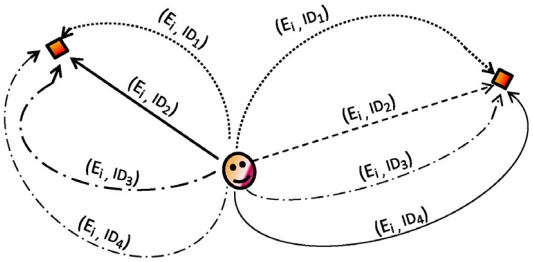
Here, we propose another scheme for Sybil attack detection based on the residual energy of each node. Unlike a normal node, each Sybil node requires multiple transmissions of control packets to validate its forged identities. A Sybil node forges four different identities to its nearest high-energy nodes as shown in Fig. 3(a). Recall that all the forged identities of a Sybil node reside in a single physical location. To authenticate its illegitimate identities, a Sybil node transmits two control packets for each one of them. It requires four different such transmissions and in each transmission, it appends one of its illicit identity. Each control packet contains residual energy, E_i, and forged identity of a Sybil node.

Upon reception of control packets, each high-energy node retrieves the identity (node ID) and residual energy from them and stores them locally within a queue as shown in Fig. 3(b). The Sybil node has launched an attack by transmitting four different pairs of control packets for its forged identities. High-energy node can detect such an attack by examining the residual energy field of each control packet. If there is a match between residual energy fields of two or more control packets and a mismatch between their identities, it means that a Sybil attack has occurred. In that case, the forged identities are reported to a nearest base station. It is possible that there may be a match between residual energy field of a normal node and that of a forged identity of a Sybil node. To rule out such possibility, we calculate the precision of the residual energy field up to 5 decimal digits. This value ensures that it is highly unlikely that there will be a match between residual energy of a normal node and forged identities of a Sybil node.

In this scheme, we implement our logic using two high-energy nodes in order to detect and compare the information on the forged identities in the network. Furthermore, we calculate the residual energy whose values after five decimal digits are truncated as any values after five decimal digits are too negligible to consider.

4.4. Cluster-based hierarchical network

After Sybil attack detection, a cluster-based hierarchical network is formed. This network consists of two phases, a set-up phase and a steady-state phase. During set-up phase, cluster heads selection, spatial queries distribution, schedule creation and cluster formation take place. The creation of schedule, i.e., allocation of *time division multiple access* (TDMA) slots, within a cluster enables the member nodes to share the transmission medium. This concept of slot allocation enables the nodes within a cluster to remain inactive for most of their lifetime and at the same time avoid contention for transmission over a wireless link.



(a) Energy consumption in transmission.

Node ID	Node Residual Energy
1	1.43252
2	1.74536
3	1.94345
4	1.74536
5	1.33991
⋮	⋮
n-1	1.54254
n	1.88972

(b) Queue of a high-energy node.

Fig. 3. Residual energy-based Sybil attack detection.

The completion of set-up phase is initiated by steady-state phase during which each cluster head collects data as specified in each query that contains H_T and S_T values for the three environmental parameters. Each base station has the ability to modify or discard H_T or S_T at any time and transmits new queries according to end user requirements. Furthermore, the base station may request data from specific nodes using on-demand queries. First, we explain the set-up phase followed by the steady-state phase.

4.4.1. Set-up phase

During Sybil attack detection, both normal nodes and Sybil nodes communicate with nearby high-energy nodes to authenticate themselves for participation in network communication. Once Sybil nodes are detected, they are reported to a nearest base station. Each high-energy node creates a control packet and transmits to a nearest base station. This packet contains the residual energy fields of normal nodes, their identities and forged identities of detected Sybil nodes. Those Sybil nodes that sneak through the detection process are reported as *normal* nodes because high-energy nodes consider only those nodes, which satisfy the detection criteria, as normal. From their perspective, normal nodes along with undetected Sybil nodes are eligible to participate in cluster head selection. However, this is not the case at the two base stations because they know the identity of each normal node within the network. Based on the identity verification at base stations, undetected Sybil nodes are barred from participation in cluster head selection. The base stations maintain two queues: a blacklisted queue and a cluster head eligible queue. Sybil nodes detected by high-energy nodes and those detected by the base stations are stored in the blacklisted queue while normal nodes become part of cluster head eligible queue.

The Sybil nodes that sneak through the detection process at high-energy nodes are ultimately detected by the base stations and barred from participation in cluster head selection. However, it does not mean that they cannot communicate with the elected cluster heads. The base stations prevent them only from participation in cluster head selection but such nodes can still communicate as member nodes with their respective cluster heads. Only those Sybil nodes that were detected earlier by high-energy nodes are permanently disabled. They can neither participate in cluster head selection nor as member nodes within the network. In our proposed scheme, high-energy nodes act as *gateway* nodes to a nearest base station. If a Sybil node that was detected earlier by high-energy nodes tries to transmit sensed data, it will be discarded straight away before reaching a base station. However, undetected Sybil nodes can still communicate within their clusters because high-energy nodes and cluster heads consider them as normal nodes.

Both base stations coordinate with each other on regular intervals to elect an optimal percentage of cluster heads. The two base stations are resource-rich entities and are fully synchronized with each other to elect a single set of cluster heads in each round. Each base station evaluates the residual energy of normal nodes to derive an average energy threshold, E_{avg}, as shown in Eq. (9).

$$E_{avg} = \sum_{i=1}^{n} \frac{E_i}{n}, \tag{9}$$

where, n is the total number of normal nodes in the network and E_i is the residual energy of normal node i. n is equal to $\mathbb{N}$-s, where $\mathbb{N}$ is the number of all nodes and s the total number of Sybil nodes. Recall that the Sybil nodes detected by high-energy nodes and the sneaked Sybil nodes detected by the two base stations are already barred from participation in cluster head selection.

A normal 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. It is the job of the two base stations to elect a desired percentage of cluster heads among candidate nodes. In our proposed scheme, the following criteria are used for the election of a candidate to become a cluster head:

- Is E_i of the candidate node greater than or equal to E_{avg}?
- Has the candidate been elected as a cluster head during the past $\frac{1}{p}$ rounds? Here, p is the optimal percentage of cluster heads and it is equal to 5% of n.
- Is the candidate's residual energy higher than those of other candidates located in the same geographical region?

Each base station elects an optimal percentage of cluster heads for a particular round and broadcasts nomination packets containing their identities and end user spatial queries. Within each spatial query, an end user specifies certain conditions for collecting critical events from the source nodes within a geographical region [38]. Initially, each cluster head is assigned a very simple query containing H_T values for temperature, relative humidity and wind speed. Next, each cluster head advertises itself to the nearest neighbouring nodes in order to form clusters. Each advertisement message contains the identity of a specific cluster head and the end user's spatial query. Each neighbouring node receives advertisement messages from multiple cluster heads but it associates itself with the one having the strongest signal strength to form a cluster.

After the cluster formation, neighbouring nodes become the non-cluster heads or member nodes of a cluster and they must abide by the conditions associated with the advertised query. Each node remains in sleep mode and wakes up only when the conditions specified in the assigned query are met. An end user may modify or discard a transmitted spatial query at any point of time. This is because an end user only wants the data of interest in making critical decisions. These queries are useful to collect user-specific data from the nodes within a geographical region.

4.4.2. Steady-state phase

In our proposed scheme, one or more member nodes may be reporting time-critical, delay-sensitive alert packets within each cluster. These member nodes are either normal nodes or undetected Sybil nodes. The normal nodes transmit genuine alert

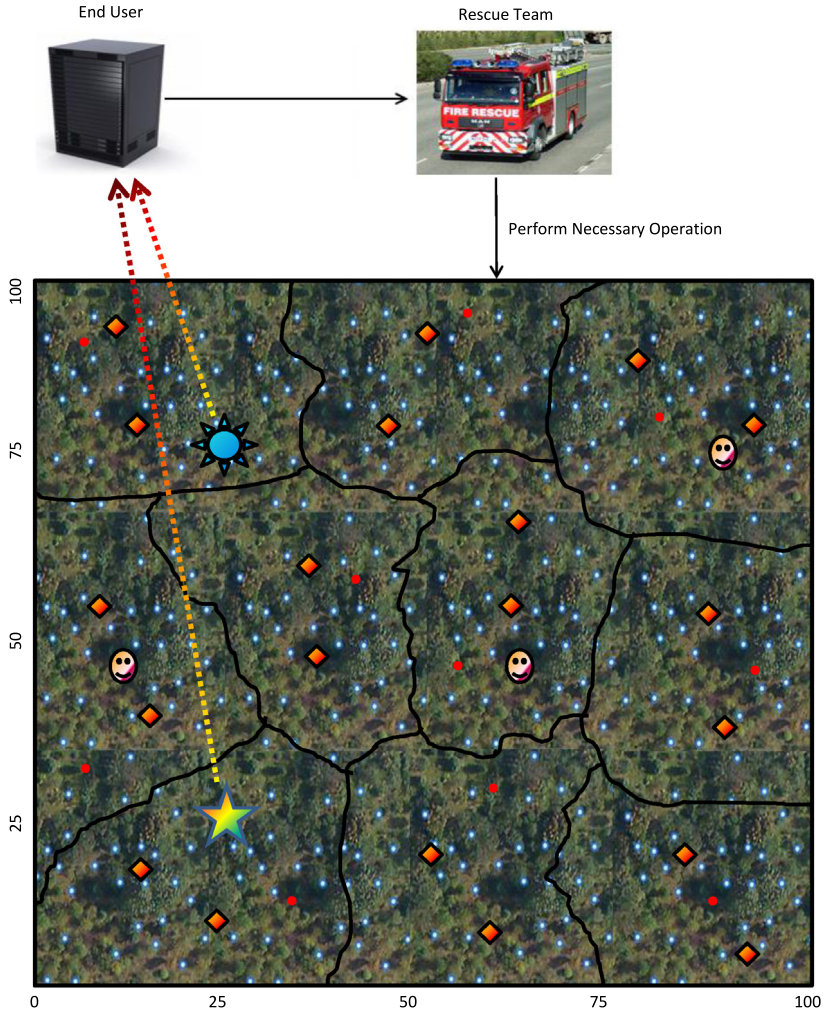


Fig. 4. Clustering-based hierarchical network.

packets while the undetected Sybil nodes transmit fake, i.e., false-negative, alert packets. No matter whether the packets are genuine or fake alerts, they are generated as a result of spatial queries distributed by the cluster heads among the member nodes within their respective clusters. For a cluster head, it is impossible to determine the nature of these alerts. Each cluster head collects and aggregates alerts from member nodes and transmits to the nearest base station using a two-hop transmission link. Each cluster head transmits the aggregated data to a nearest high-energy node, which in turn transmits the data to the base station. The only exception is the presence of a nearby base station. In that case, the cluster head directly delivers data to it.

Upon reception at a base station, the identity, geographical location and sensitivity of an alert packet are examined. A mismatch of identities between an incoming alert packet and those stored in a base station enables the base station to discard such packet. This comparison enables each base station to weed out malicious and false-negative alert packets of Sybil nodes. At this point, the nearest base station assigns on-demand queries to those member nodes, i.e., normal nodes, which are reporting time-critical, delay-sensitive alert packets. The normal nodes store these queries and report alert packets on regular intervals based on the specified conditions, i.e., the hard and soft thresholds, of the three environmental parameters. Based on these conditions, the member nodes generate alerts packets that need to be reported immediately to the nearest base station.

Each cluster head consumes a considerable amount of time in data aggregation, in-network processing and relaying back the data to a base station. To avoid delay and maintain the integrity of time-stamp of each alert packet, the nearest base station collects such alerts and relays them back to an end user. These alerts are generated as a result of the conditions specified in on-demand queries, assigned to normal nodes. Based on the gathered data, an end user decides what to do in the next step. If a subset of neighbouring nodes is reporting sensitive alert packets on a regular interval, the end user informs a rescue team in order to take precautionary measures within a forest. An end user evaluates the aggregated data from each perspective before declaring an emergency situation within a forest. The two phases of our cluster-based hierarchical network are shown in Fig. 4. The completion of set-up and steady-state phases is coined as one complete round. Our proposed cluster-based hierarchical network is iterative in nature and operates in rounds to collect data within a forest.

In Algorithm 2, the generation, transmission and outcome of a simple spatial query are shown. The base stations elect an optimal set of cluster heads, opt_{CH}, and append H_T values of temperature, relative humidity and wind speed in each control packet, ctr_i for each cluster head, CH_i. Here, $opt_{CH} = \{CH_1, CH_2, \ldots, CH_{opt}\}$, $opt \in \{1, 2, \ldots, n\}$. The base stations also append the identity, latitude (LatT) and longitude (LongT), of CH_i, in ctr_i. Next, ctr_i is sent to a neighboring high-energy node, hen_i, for ultimate transmission to CH_i. Each CH_i gathers alert packets from member nodes based on the specifications in ctr_i and transmits them to a nearby base station that evaluates the nature of these incoming alert packets. If the identity of a member node, N-CH_x, does not match with identities in $Database_{BS}$ stored within a base station, the incoming alert

packets of CH_x are discarded and it is considered as a Sybil node. Each base station prevents data from CH_x to reach an end user. A match of identity $N\text{-}CH_x$ means that CH_x is a legitimate node and its alert packets are further evaluated for the possibility of assigning an on-demand query to this node.

Algorithm 2 Generation, Transmission and Outcome of Spatial Query

```
procedure SPATIAL QUERY
    for each identity of a cluster head, CH_i ∈ opt_CH do
Create a control packet, ctr_i
Append H_T along with LatT, LongT of CH_i in ctr_i
SEND ctr_i to hen_i
    end for
    if N-CH_x ∉ Database_BS then
Discard the incoming alert packets from CH_x
    else
Further Evaluation of alert packets
    end if
end procedure
```

Once data are collected from each cluster head, the two base stations further analyse this data. The dynamic nature of happening events (wildfire in this case) is not restricted to a particular geographical location. Moreover, the data collected from various geographical regions may have different temperature, relative humidity and wind speed readings. Both base stations carefully examine the data before delivering them to an end user. During this process, the nodes that have highly sensitive data are analysed and assigned on-demand queries. In Algorithm 3, a subset of nodes are assigned an on-demand query. This query probes a total of 6 nodes to obtain specific values for temperature, relative humidity and wind speed. Furthermore, the incoming alert packets must have a time-stamp otherwise they will be discarded. The allocation of on-demand queries is restricted to legitimate nodes only because the base stations have already discarded alert packets of undetected (sneaked) Sybil nodes in Algorithm 2.

Algorithm 3 Generation and Outcome of an On-demand Query

```
procedure ON-DEMAND QUERY
SELECT Temp, R.H, W.S, Timestamp FROM Nodes
WHERE (Node ID = 2 TO 7) ∧ (Temp>= 219, R.H< 0.34, W.S> 10)
    if Collected data is Time stamped then
End user checks data pattern collected from nodes in the above range
    else
Discard data of the nodes without Time stamps
    end if
end procedure
```

5. Experimental results and analysis

In this section, we provide a series of simulation results for our proposed scheme. We use a 100 × 100 square metre geographical area for our network deployment. Normal nodes and Sybil nodes are isomorphic in terms of battery power, storage and processing capabilities. We use First-order radio model [39] to minimize the energy consumption of nodes by efficiently scheduling their duty-cycles. Next, we evaluate our scheme in terms of detection rate, accuracy of the application, network lifetime and average size of the clusters. These parameters are calculated for a period of up to 30 000 rounds.

5.1. Detection of Sybil attack

Fig. 5(a) illustrates the effect of s Sybil nodes and their $|m|$ forged identities on the detection rate for a network of $s = 20$ and $n = 300$. The value of m varies between 10 and 26. The detection rate increases with the increase in the value of m. In comparison with Ssu et al. [23], the detection rate of our approach is slightly lower at high-energy nodes. However, we have a better detection rate at the two base stations. Unlike their detection scheme, our proposed scheme operates in a hostile environment within a forest that causes high fluctuation in RSSI values. As a result, one or more identities of Sybil nodes sneak through the detection process at high-energy nodes in various rounds resulting in a lower detection rate. However, our proposed scheme has a 100% detection rate at the two base stations and all previously sneaked identities are detected at this stage. It is the ultimate goal of our proposed approach that all Sybil nodes and their forged identities are prevented from cluster head participation and a 100% detection rate achieves that objective.

In Fig. 5(b), Sybil nodes are detected based on their residual energy consumptions. The comparison is made for a network of 200 normal nodes and up to 15 Sybil nodes. In most of the rounds, Sybil nodes consume more energy as compared to normal nodes because each Sybil node forges $|m|$ identities. The energy consumption varies with m identities forged by each Sybil node over the course of network lifetime. According to Section 4.3.2, the energy consumption of Sybil nodes is contributed towards the number of transmitted control packets that are used to validate their forged identities.

5.2. Accuracy of wildfire monitoring application

The accuracy of the proposed wildfire application, $\phi_{wildfire}$, is a percentage value that is calculated as the number of genuine alerts denoted by N to the number of total alerts denoted by T as shown in Eq. (10).

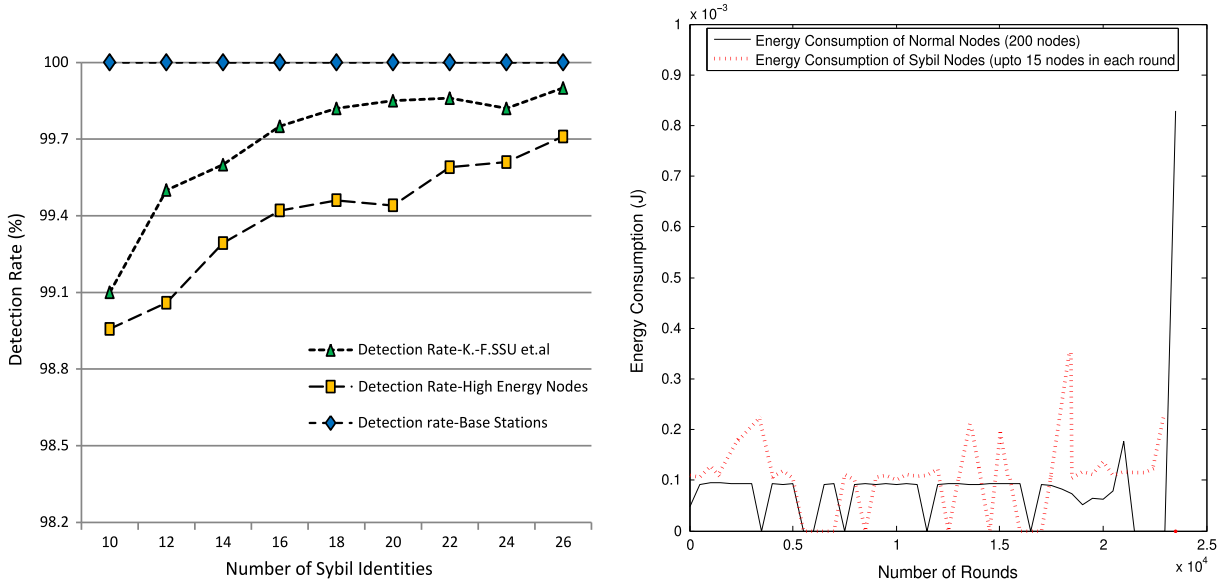
$$\phi_{wildfire} = \frac{N}{T} \times 100. \tag{10}$$

In Eq. (10), we calculate the accuracy of our wildfire monitoring application in presence of up to 20 Sybil nodes. Each Sybil node is capable of forging up to 10 identities.

In Fig. 6(a), the data packets of the three environmental parameters captured over the span of network lifetime are shown. These packets are generated by the normal nodes and Sybil nodes. Furthermore, these packets may or may not be alert packets. Each data packet is an alert packet if it satisfies the H_T and S_T conditions specified for an environmental parameter. Let Temp represent the temperature reading, RH represent the relative humidity reading and WS represent the wind speed reading sensed by a member node. This node will transmit an alert packet to its respective cluster head only when Eq. (11) is satisfied.

$$(Temp \geq H_{Temp}) \wedge ((RH \geq H_{RH}) \vee (WS \geq H_{WS})) = True, \tag{11}$$

where, H_{Temp}, H_{RH} and H_{WS} are the hard thresholds for temperature, relative humidity and wind speed. An alert is generated and transmitted only when H_{Temp} is reached **and** either H_{RH} **or** H_{WS} is reached. Eq. (11) is user-specific and may be modified according to the demands of an end user. This is because an on-demand query has different conditions for an alert transmission as compared to a spatial query. As previously discussed, an on-demand query is highly generalized and assigned to a member node based on the outcome of a spatial query. For example, an end user may require only temperature alerts of above 200 °C from a specific geographical region and may not be interested in relative humidity and wind speed alerts within the same region. In that case, Eq. (11) will change accordingly. The alerts include both genuine and fake readings as shown in Fig. 6(b). Fake alerts belong to the forged identities of one or more Sybil nodes while the genuine alerts belong to the normal node. In Fig. 6(b), there are as many as 147 genuine alerts and up to 20 fake alerts in various rounds. In Fig. 6(c), the percentage accuracy of our wildfire monitoring application is obtained using genuine and fake alerts of Fig. 6(b).



(a) RSSI-based detection.

(b) Energy consumption-based detection.

Fig. 5. Detection of Sybil attack.

5.3. Lifetime of the network

The lifetime of a network is defined in terms of stability period and instability region. Stability period is the point of time when the first node dies while instability region is the point of time when there are not sufficient nodes to form balanced clusters. In [40], the authors argued that a network is unable to sustain balanced clusters when 97% of its nodes die. Using their argument, we have compared our proposed approach with LEACH, SEP and PASCCC protocols in Fig. 7(a).

Unlike LEACH and SEP protocols, the sensor nodes in our approach wake up only when H_T is reached for each environmental parameter. The use of H_T and S_T within the spatial and on-demand queries efficiently manages the sleep–awake scheduling of the nodes. On the other hand, PASCCC protocol uses H_T and S_T parameters. However, the selection of cluster heads is similar to LEACH and SEP that enable each node to elect itself as cluster head irrespective of its residual energy. The threshold-based event detection along with a centralized cluster head selection significantly improves the stability period and instability region of our scheme. In our proposed scheme, the instability region reaches in round 28 322 and has a 62% improvement over the nearest reading, i.e., PASCCC protocol. In terms of stability period, our proposed scheme has almost 3 times better performance as compared to PASCCC protocol. In Fig. 7(b), we calculate the network lifetime of our proposed approach in the presence and absence of high-energy nodes. The presence of only 20 high-energy nodes may have less impact on the lifetime of the network, however, they perform vital tasks of Sybil attack detection to the base stations.

Apart from Sybil attack detection, high-energy nodes minimize network congestion, retransmission attempts, packet loss, and energy consumption. In absence of high-energy nodes, each normal node will use a single-hop long-haul communication channel to reach the base station. Such transmissions will lead to higher packet loss, delay, congestion and degradation in Quality of Service (QoS). In our proposed scheme, the presence of high-energy nodes reduces long-haul transmissions to the base station. For every n packet received, a higher energy node transmits only a single packet to the base station as discussed in Section 4.4.1. As a result, the probability of congestion, packet loss and packet retransmission attempts is reduced. Moreover, the amount of time required in retransmitting a lost/missing packet to a high-energy node is much smaller compared to the time required in a retransmission attempt intended for a base station. Next, we present the energy consumption analysis of any normal node in the presence and absence of high-energy nodes.

5.3.1. Analysis of energy consumption

The energy consumption, E_n, of any normal node in absence of high-energy nodes is calculated as shown in Eq. (12).

$$E_n = kE_{elec} + kE_{amp}d_{BS}^4, \quad \text{where } d \geq d_c. \tag{12}$$

Here, E_{elec} is the energy consumption of the processing unit of a normal node in executing a k-bit packet, E_{amp} is the energy consumption of the amplifier component in transmitting a k-bit packet and d_{BS} is the distance between a normal node and a base station and it is always greater than or equal to crossover distance, d_c [14].

The energy consumption of any normal node in presence of high-energy nodes is calculated as shown in Eq. (13).

$$E_n = kE_{elec} + kE_{amp}d_{HEN}^2, \quad \text{where } d < d_c. \tag{13}$$

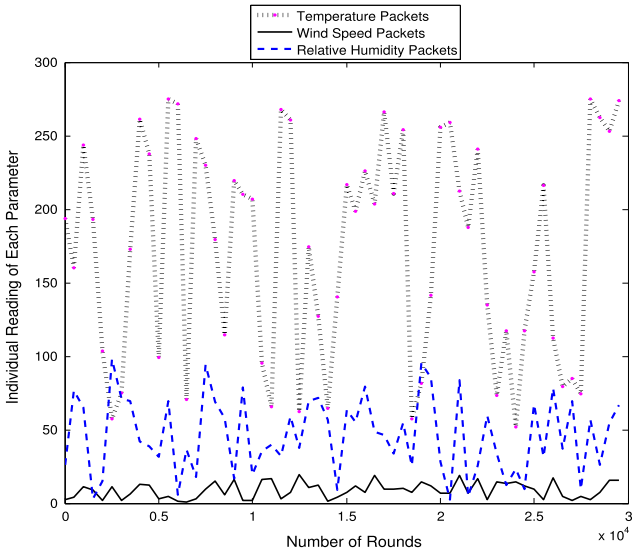
Here, d_{HEN} is the distance between a normal node and its nearest high-energy node and is always less than d_c [14].

Comparing Eqs. (12) and (13), the energy consumption of a normal node is much smaller in presence of high-energy node because $d_{HEN} < d_{BS}$. Recall that the data transmission via high-energy nodes involves two-hop communication. The decrease in energy consumption is contributed much towards the second hop because it is the high-energy nodes that perform data transmission on behalf of their nearest neighbouring normal nodes.

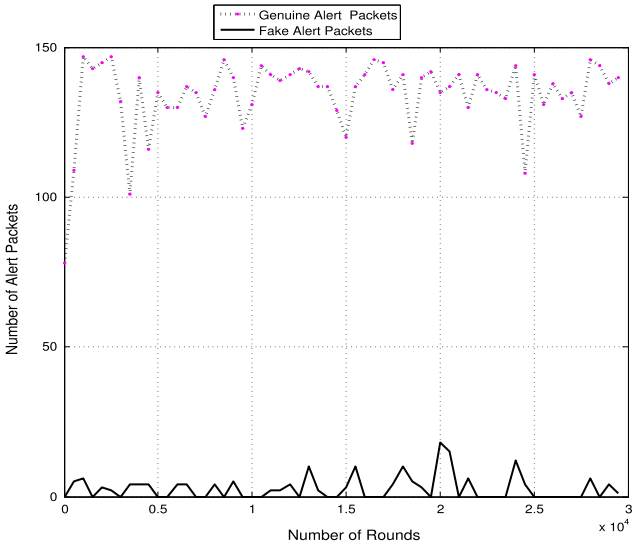
5.4. Average size of the clusters

This parameter determines the efficiency of an algorithm in terms of data aggregation and geographical coverage [26]. The average size of a cluster, $C_{average}$, is defined as the ratio of number of alive nodes to the number of clusters within a network and is calculated using Eq. (14).

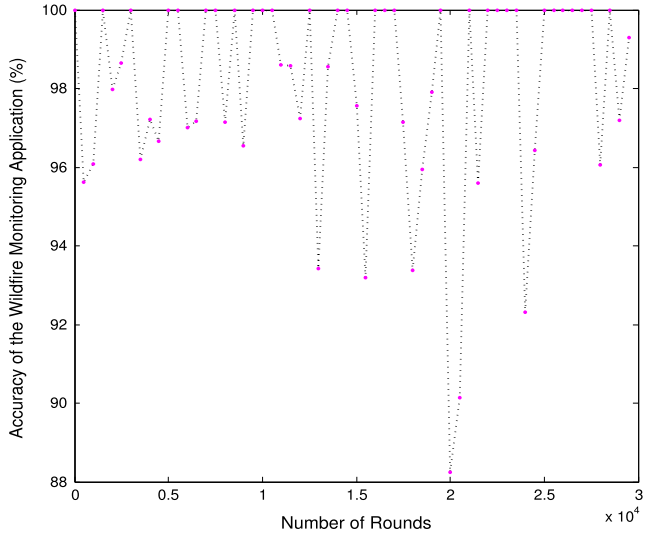
$$C_{average} = \frac{count\,(n_{alive} \mid E_i(n_{alive}) > 0,\ \forall\, n_{alive} \in n)}{count\,(CH)}, \tag{14}$$



(a) Readings of environmental parameters.



(b) Genuine alerts vs. fake alerts.



(c) Percentage accuracy.

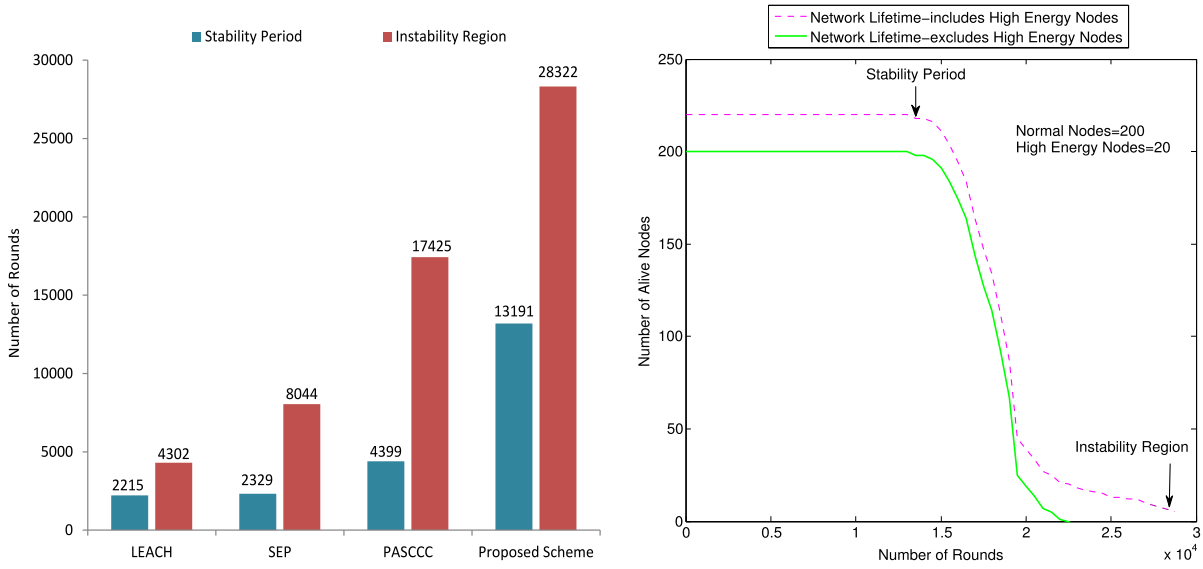
Fig. 6. Accuracy of the wildfire monitoring application.

where, n_{alive} denotes the number of alive nodes participating in cluster formation and CH denotes the number of clusters. In cluster-based hierarchical routing protocols, there is one cluster head per cluster and both these parameters are always equal. The average size of a cluster expresses the efficiency of an algorithm in terms of data aggregation, data fusion and the minimum number of cluster heads required to cluster a large geographical network. In Fig. 8(a), the average cluster size of our proposed scheme is compared with LEACH and DEECIC protocols.

Our proposed scheme has a significantly higher $C_{average}$ as compared to DEECIC and LEACH protocols. Our algorithm is centralized in nature and elects an optimal percentage of cluster heads in each round. On the other hand, DEECIC and LEACH are randomly distributed in nature and cannot guarantee an optimal percentage of cluster heads in each round. The selection of an optimal percentage of cluster heads means a uniform distribution of network load, better coverage, high throughput, better data aggregation and low delay within each cluster [41]. In LEACH and DEECIC protocols, each node elects itself as cluster head based on a generated random number [14]. As a result, either too many or very few cluster heads are elected in each round so that unbalanced clusters occur. Too many member nodes within a cluster increase the load on a cluster head and very few member nodes make the concept of clustering rather inefficient and ineffective. In Fig. 8(b), the total number of cluster heads elected in each round is shown for a network of 100 normal nodes. Irrespective of the candidate nodes, each round results in 4 or 5 cluster heads as long as the network is stable. As we are using a centralized approach, the base station elects an optimal percentage of cluster heads among the candidate nodes. Fig. 8(a) is derived based on the statistics of Fig. 8(b). As discussed earlier, $C_{average}$ is the ratio of number of alive nodes (candidates nodes of Fig. 8(b)) to the total number of clusters in each round. The number of clusters is always equal to the number of cluster heads.

6. Conclusion

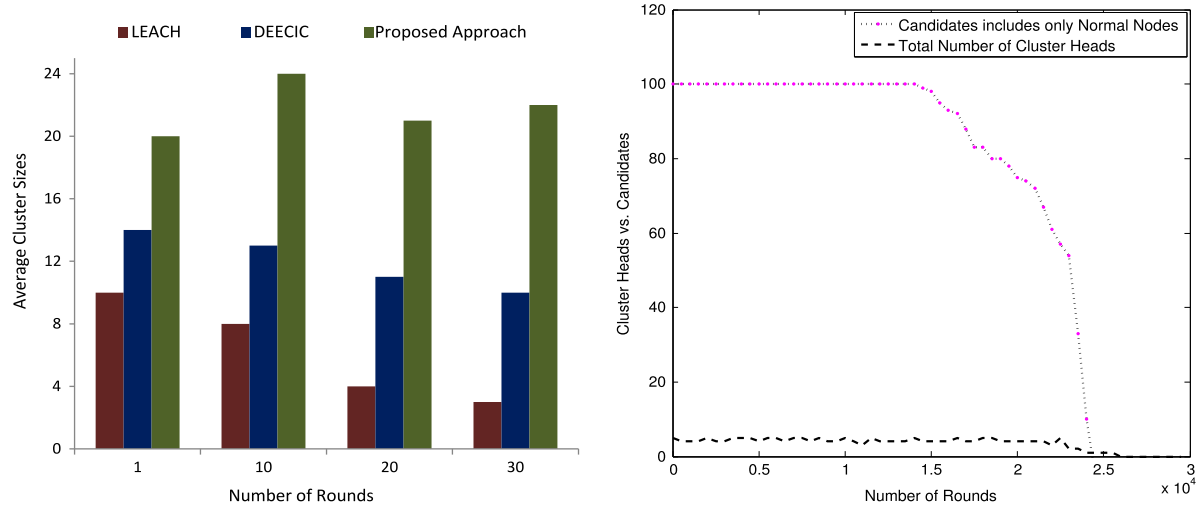
This paper proposes two different techniques for Sybil attack detection for a forest wildfire monitoring application. A two-tier detection technique uses high-energy nodes operating at a lower level to detect forged identities of Sybil nodes. Due to the hostile environment within in a forest, one or more identities may sneak



(a) Stability period and instability region.

(b) Prolonged network lifetime.

Fig. 7. Lifetime of the network.



(a) Average cluster size.

(b) Cluster heads in each round.

Fig. 8. Coverage of a geographical region.

through the detection process. These identities are ultimately detected by the two base stations operating at a higher level. A residual energy-based detection technique uses the residual energy of each node to detect a possible Sybil attack at the high-energy nodes. If two or more incoming control packets have the same residual energy but different identities, it means that a Sybil attack has been launched by an adversary. The main objective of these techniques is to prevent Sybil identities from participation in cluster head selection. After a Sybil attack detection, an optimal percentage of cluster heads are selected by the base stations using a centralized approach. Each cluster head is assigned a spatial query to collect data about the environmental parameters within a forest. Each cluster head advertises itself to its neighbouring nodes in order to form clusters and collect data from member nodes. The data collected from member nodes may belong to either normal nodes or sneaked Sybil identities of Sybil nodes. To deceive an end user, the sneaked Sybil identities may broadcast high volume of false-negative alerts. The two base stations remain vigilant to prevent any such alerts from reaching an end user. Only incoming genuine alerts from normal nodes are analysed, and, if any of them is of significant importance, the base stations assign on-demand queries to the origin of those alerts, i.e. the source nodes. Our proposed approach has fully implemented spatial queries, however on-demand queries are yet to be implemented. Currently, we are implementing high level advanced on-demand queries that will probe all those member nodes who claim to have highly critical alert packets. Moreover, the selection of cluster heads is computationally complex and large overhead is involved. This problem can be improved further in our future work using a rather simple logic.

Acknowledgements

This research is supported by the University of Technology, Sydney (UTS) International Research Scholarship (IRS) and Commonwealth Scientific and Industrial Research Organisation (CSIRO) Information and Communication Technologies (ICT) Centre Top-up Scholarship, Grant ID: 2013000855.

References

[1] J. Yick, B. Mukherjee, D. Ghosal, Wireless sensor network survey, Comput. Netw. 52 (2008) 2292–2330.
[2] J. Gutiérrez, J.F. Villa-Medina, A. Nieto-Garibay, M.Á. Porta-Gándara, Automated irrigation system using a wireless sensor network and gprs module, IEEE Trans. Instrum. Meas. 63 (2014) 166–176.
[3] J.M. Corchado, J. Bajo, D.I. Tapia, A. Abraham, Using heterogeneous wireless sensor networks in a telemonitoring system for healthcare, IEEE Trans. Inf. Technol. Biomed. 14 (2010) 234–240.
[4] Y.E. Aslan, I. Korpeoglu, Ö. Ulusoy, A framework for use of wireless sensor networks in forest fire detection and monitoring, Comput. Environ. Urban Syst. 36 (2012) 614–625.
[5] K.K. Khedo, R. Perseedoss, A. Mungur, et al. A wireless sensor network air pollution monitoring system, 2010. ArXiv Preprint arXiv:1005.1737.
[6] M. Hefeeda, M. Bagheri, Wireless sensor networks for early detection of forest fires, in: IEEE International Conference on Mobile Adhoc and Sensor Systems, 2007, MASS 2007, IEEE, 2007, pp. 1–6.
[7] Z. Zhou, S. Das, H. Gupta, Connected k-coverage problem in sensor networks, in: Proceedings of 13th International Conference on Computer Communications and Networks, 2004, ICCCN 2004, IEEE, 2004, pp. 373–378.
[8] Y.G. Sahin, Animals as mobile biological sensors for forest fire detection, Sensors 7 (2007) 3084–3099.
[9] E.M. García, M.Á. Serna, A. Bermúdez, R. Casado, Simulating a wsn-based wildfire fighting support system, in: ISPA, Vol. 8, 2008, p. 896.
[10] W.M. Dlamini, A Bayesian belief network analysis of factors influencing wildfire occurrence in swaziland, Environ. Modell. Softw. 25 (2010) 199–208.
[11] Y.-g. Ha, H. Kim, Y.-c. Byun, Energy-efficient fire monitoring over cluster-based wireless sensor networks, Int. J. Distrib. Sens. Netw. 2012 (2012).
[12] Y. Yao, J. Gehrke, Query processing in sensor networks, in: CIDR, 2003, pp. 233–244.
[13] X. Liu, A survey on clustering routing protocols in wireless sensor networks, Sensors 12 (2012) 11113–11153.
[14] W.R. Heinzelman, A. Chandrakasan, 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.
[15] G. Smaragdakis, I. Matta, A. Bestavros, SEP: A stable election protocol for clustered heterogeneous wireless sensor networks, Technical Report, Boston University Computer Science Department, 2004.
[16] M.A. Jan, P. Nanda, X. He, R.P. Liu, Enhancing lifetime and quality of data in cluster-based hierarchical routing protocol for wireless sensor network, in: 2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing (HPCC_EUC), IEEE, 2013, pp. 1400–1407.
[17] J. Zhang, W. Li, N. Han, J. Kan, Forest fire detection system based on a zigbee wireless sensor network, Front. For. China 3 (2008) 369–374.
[18] I. Yoon, D.K. Noh, D. Lee, R. Teguh, T. Honma, H. Shin, Reliable wildfire monitoring with sparsely deployed wireless sensor networks, in: 2012 IEEE 26th International Conference on Advanced Information Networking and Applications (AINA), IEEE, 2012, pp. 460–466.
[19] N.A. Pantazis, S.A. Nikolidakis, D.D. Vergados, Energy-efficient routing protocols in wireless sensor networks: A survey, in: Communications Surveys & Tutorials, Vol. 15, IEEE, 2013, pp. 551–591.
[20] D. Ballari, M. Wachowicz, A.K. Bregt, M. Manso-Callejo, A mobility constraint model to infer sensor behaviour in forest fire risk monitoring, Comput. Environ. Urban Syst. 36 (2012) 81–95.
[21] M.A. Jan, P. Nanda, X. He, R.P. Liu, A sybil attack detection scheme for a centralized clustering-based hierarchical network, in: Trustcom/BigDataSE/ISPA, 2015 IEEE, Vol. 1, IEEE, 2015, pp. 318–325.
[22] J. Newsome, E. Shi, D. Song, 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.
[23] K.-F. Ssu, W.-T. Wang, W.-C. Chang, Detecting sybil attacks in wireless sensor networks using neighboring information, Comput. Netw. 53 (2009) 3042–3056.
[24] M. Demirbas, 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.
[25] Y. Chen, J. Yang, W. Trappe, R.P. Martin, Detecting and localizing identity-based attacks in wireless and sensor networks, IEEE Trans. Veh. Technol. 59 (2010) 2418–2434.
[26] Z. Liu, Q. Zheng, L. Xue, X. Guan, A distributed energy-efficient clustering algorithm with improved coverage in wireless sensor networks, Future Gener. Comput. Syst. 28 (2012) 780–790.
[27] P. Cheney, A. Sullivan, Grassfires: Fuel, Weather and Fire Behaviour, CSIRO PUBLISHING, 2008.
[28] J.M. Bahi, A. Makhoul, M. Medlej, An optimized in-network aggregation scheme for data collection in periodic sensor networks, in: Ad-hoc, Mobile, and Wireless Networks, Springer, 2012, pp. 153–166.
[29] A. Manjeshwar, D.P. Agrawal, Teen: a routing protocol for enhanced efficiency in wireless sensor networks, in: Proceedings of the 15th International Parallel and Distributed Processing Symposium, IPDPS, April-2001, San Francisco, CA, USA, pp. 23–27.
[30] J. Luo, J.-P. Hubaux, Joint sink mobility and routing to maximize the lifetime of wireless sensor networks: the case of constrained mobility, IEEE/ACM Trans. Netw. 18 (2010) 871–884.
[31] P. Wang, I.F. Akyildiz, Effects of different mobility models on traffic patterns in wireless sensor networks, in: Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE, IEEE, 2010, pp. 1–5.
[32] M. Dong, K. Ota, H. Li, S. Du, H. Zhu, S. Guo, Rendezvous: towards fast event detecting in wireless sensor and actor networks, Computing 96 (2014) 995–1010.
[33] M. Dong, K. Ota, L.T. Yang, S. Chang, H. Zhu, Z. Zhou, Mobile agent-based energy-aware and user-centric data collection in wireless sensor networks, Comput. Netw. 74 (2014) 58–70.
[34] O. Cayirpunar, E.K. Urtis, B. Tavli, The impact of base station mobility patterns on wireless sensor network lifetime, in: 2013 IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), IEEE, 2013, pp. 2701–2706.
[35] S. Sudevalayam, P. Kulkarni, Energy harvesting sensor nodes: Survey and implications, IEEE Communications Surveys & Tutorials 13 (2011) 443–461.
[36] J.L. Burbank, W. Kasch, J. Ward, An Introduction to Network Modeling and Simulation for the Practicing Engineer, Vol. 5, John Wiley & Sons, 2011.
[37] T. Singal, Wireless Communications, Tata McGraw-Hill Education, 2010.
[38] R.I. Da Silva, D.F. Macedo, J.M.S. Nogueira, Spatial query processing in wireless sensor networks–a survey, Inf. Fusion 15 (2014) 32–43.
[39] M.A. Jan, P. Nanda, X. He, Energy evaluation model for an improved centralized clustering hierarchical algorithm in WSN, in: Wired/Wireless Internet Communication, Springer, 2013, pp. 154–167.
[40] M.A. Jan, P. Nanda, X. He, R.P. Liu, Pascc: Priority-based application-specific congestion control clustering protocol, Comput. Netw. 74 (2014) 92–102.
[41] H. Yang, B. Sikdar, Optimal cluster head selection in the leach architecture, in: Performance, Computing, and Communications Conference, 2007. IPCCC 2007. IEEE International, IEEE, 2007, pp. 93–100.

Mian Ahmad Jan recently completed Ph.D. in the Faculty of Engineering and Information Technology (FEIT) of the University of Technology Sydney (UTS) Australia. His research interests are Cluster-based Hierarchical routing protocols in Wireless Sensor Networks, Congestion detection and mitigation, Internet and Web of Things and efficient Intrusion and malicious attack detection in Wireless Sensor Networks.

Priyadarsi Nanda is a Senior Lecturer, in the School of Computing and Communications and a Core Research Member of Centre for Real-time Information Networks (CRIN). He is currently the leader of Network Security Research group. His research interests include network QoS, network securities, Internet of Thing (IOT) securities and wireless sensor networks. He has more than 25 years of experience in teaching and research and has published more than 60 research publications. He is a senior member of the IEEE.

Xiangjian He is a Professor of Computer Science, School of Computing and Communications. He is also Director of Computer Vision and Pattern Recognition Laboratory of the Global Big Data Technologies Centre (GBDTC), and the Director of UTS-NPU International Joint Laboratory on Digital Media and Intelligent Networks at the University of Technology, Sydney (UTS). He is an IEEE Senior Member. His research interests are image processing, pattern recognition, computer vision and Network security. He has received many research grants including four national Research Grants awarded by Australian Research Council (ARC).

Ren Ping Liu is a principal scientist of networking technology in CSIRO ICT Centre, Australia and Adjunct Professor, Macquarie University and University of Technology, Sydney. His interests include scheduling, QoS modeling and performance analysis of wireless networks, including IEEE 802.11, Wireless Mesh Networks, Wireless Sensor Networks, LTE, and Cognitive Radio Networks. He delivered networking solutions to government and industrial customers, including Optus, AARNet, Nortel, Queensland Health, CityRail, Rio Tinto, and DBCDE.