Abstract—Protocols are one of the main aspect and essential feature of communication. In Wireless Sensor Networks (WSNs) different kinds of protocols and topologies are being used to support data transmission and data reliability issues. Hierarchical Cluster Base Routing Protocols (CBRPs) is most famous due to their lower energy consumption and flexible network scalability characteristics. In CBR, the sensor nodes are divided into two kinds of responsibilities such as Cluster Heads (CHs) and Cluster Members (CMs). CH’s perform data aggregation and data fusion related tasks while CMs only sense the environmental parameters and forward to their respective CHs. However, the data is only stored on the Base Station (BS). Data availability is one of the main concerns in the CBRP, if any CHs or CMs or BS is damaged due to energy depletion or hardware failure, the sensed data will be loose. In this research paper, we have performed a few experiments to evaluate the data availability related issues in CBR due to interference, network and BS failure. Therefore, we suggest some improvement to ensure data availability in WSNs due to network failure and hardware failure issues.

Index Terms—Wireless Sensor Network; Data Storage; Data Availability; Data Loss; Cluster Base Routing Protocols; LEACH.

I. INTRODUCTION

In any communication networks, the network topologies and protocols are responsible to control, transmit and manage the data transmission over wired or wireless medium. Topologies emphasis, how the communication devices or nodes should be deployed within a specific area and protocols control the data communication, data transmission process and ensure the data reliability within a network. Topologies and protocols are used to control various constraints such as energy, latency, computational resources and communication reliability within a network. The network topologies define, routing paths, communication pattern such as unicast or multicast, packet types/size, data aggregation and helps in reducing radio interference. It also helps in controlling the number of nodes within one cluster and constructs the communication network layout. Whereas, the routing protocols are classified into two categories such as flat routing protocols, and hierarchical routing protocols [1]. Both of them have different characteristics and capabilities, and used under various scenarios. Flat routing protocols are mostly used in small networks because all nodes perform identical tasks and contain equal capabilities. In flat routing data is transmitted hop by hop in the form of flooding [1]. Some of flat routing protocols example are Flooding and Gossiping, Sensor Protocols for Information via Negotiation (SPIN), Directed Diffusion (DD), Rumor, Greedy Perimeter Stateless Routing (GPSR), Trajectory Based Forwarding (TBF), Energy-Aware Routing (EAR), Gradient-Based Routing (GBR), Sequential Assignment Routing (SAR) and etc [1, 2]. However, hierarchical routing protocols have different characteristics because of limited energy and storage resources. In a hierarchical network the sensor nodes are divided into two types such as Cluster Heads (CHs) and Cluster Members (CMs) which perform different tasks and are organised into a large number of clusters. The following are the example of hierarchical protocols such as Low-energy Adaptive Clustering Hierarchy (LEACH), Hybrid Energy-Efficient Distributed clustering (HEED), Distributed Weight-based Energy-efficient Hierarchical Clustering protocol (DWEHC), Position-based Aggregator Node Election protocol (PANEL), Two-Level Hierarchy LEACH (TL-LEACH), Unequal Clustering Size (UCS) model, Energy Efficient Clustering Scheme (EECS), Energy-Efficient Uneven Clustering (EEUC) algorithm, Power-Efficient Gathering in Sensor Information Systems (PEGASIS), Threshold sensitive Sensors Energy Efficient sensor Network protocol (TEEN), The Adaptive Threshold sensitive Energy Efficient sensor Network protocol (APTEEN), Two-Tier Data Dissemination (TTDD), Hierarchical Geographic Multicast Routing (HGMR) and etc [1, 2].

A Cluster Base Wireless Sensor Network (CBWSN) is a combination of a large number of sensor nodes, which are connected to a one BS through multiple CHs [2]. CBWSN use Cluster Base Routing Protocols (CBRPs) for communication and data management. In CBWSN, sensor nodes are equipped with a transceiver that assists the in receiving and transmitting the sensed data between CMs, from CMs to CHs and from CHs to a BS. The sensed data is stored on a BS and available for the users’ interaction. The sensor nodes are autonomous small devices, which have several constraints such as low power, limited computation capacity, short communication range, prone to interference and small memory space due to their tiny size. These sensor nodes are deployed randomly within a specific area and left unattended for a long period; they are expected to perform their tasks independently and efficiently. As a result, the WSNs have usually varying degrees of sensor node density along its area size. Due to a large deployment of sensor nodes within the specific area; sensor nodes start to die due to limited energy. When nodes start to die, the network becomes less productive and data start to lose.

In CBRPs, each CH broadcast information to its CHs via Code Division Multiple Access (CDMA) and they join the nearest CH. The CHs used Time Division Multiple Access (TDMA) to offer data transmission time slot for every member to send their data which help in minimising and controlling network interference [2]. Similarly, a BS also transmits advertise message for CHs to connect and forward the

Evaluation and Improvement of Data Availability in WSNs Cluster Base Routing Protocol

Mujahid Tabassum and Dr. Kartinah Zen
Faculty of Computer Science & Information Technology
mujee1983@gmail.com
collected data. Once the cluster formations are completed, the CMs should forward sensed data or information to their respective CHs. After this CHs perform data aggregation and data fusion tasks and forward appropriate data to a BS. BS has sufficient data storage to save the data and keep it for unlimited time. In Cluster Base Networks, data is only stored on a BS, CHs and CMs have no capability to store the sensed data. Figure 1 shows the communication process of sensor nodes to a BS.

![Figure 1: Communication Process in WSNs](image)

CBWSN have two types of communication infrastructure such as multi-hop and single-hop. In single-hop networks, each CH directly communicates with a BS. The CHs that are far from the BS die faster as compared to the nearest CHs, because they need more energy to transmit the data over the long distance. Whereas in multi-hop cluster networks, data is transmitted between multiple CHs before reaching to the BS, which creates network congestion, increase data traffic and cause of packet loss at CHs near to the BS [1].

![Figure 2: Single Hop and Multi Hop Cluster Base Network](image)

The CBWSN mainly used two kinds of topologies such as Homogeneous Networks or Heterogeneous Networks [3].

i. Homogeneous Networks: A network in which all nodes consume the same level of energy.

ii. Heterogeneous Networks: A network in which some nodes are supported with more capabilities and assigned with more responsibilities such as data gathering and forwarding. Therefore, the energy consumption level is different among all nodes.

Data availability is to ensure that the data is always available at a required level of performance in both normal and disastrous situations [4]. In CBWSN, data is collected by CMs and transmitted to the BS to be stored. The CMs and CHs are responsible to sense, forward, perform data aggregation and data fusion tasks. Hence, when communication breaks between CMs and CHs or CHs and BS, there will be no availability of data as CMs and CHs are only responsible for transmitting the data, not to store the data. Hence, the CBRPs are not eligible to offer data availability service upon any node or network failure. Many researches and improvement have been made to enhance the performance of CBRPs [1, 2]; however, most of them are towards improving the network life and enhancing the cluster formation setup only.

In this research paper, we have performed some simulation base experiments using cluster base routing protocol LEACH to analyse the data loss due to interference, network and BS failure. Therefore, we are proposing an Artificial Intelligent (AI) WSN framework, which should have the ability of self-perceive the environment, adapt and learn according to the environment changes to offer significant network consistency and Data Backup on Demand (DBoD) service in a WSN to meet the user expectation.

II. LITERATURE REVIEW

A. Cluster Base Routing Protocol - LEACH

Low-Energy Adaptive Clustering Hierarchy (LEACH) protocol was initially proposed by Heinzelman in 2000 [5]. It is a distinctive hierarchical clustering routing protocol that supports distributed clustering algorithm methods such as cluster head rotation, data aggregation, and data fusion. In LEACH, a Time Division Multiple Access (TDMA) based Medium Access Control (MAC) protocol is integrated to form clusters with a simple routing protocol, which helps in reducing inter-cluster and intra-cluster collisions. LEACH used random and circularly method to choose the CHs, this helps in optimizing network energy by exchanging CHs in a network per round. Each time CHs energy finished or they die, the LEACH will perform new election to select new CHs. LEACH network has a fixed BS and sensor nodes are randomly deployed within communication radios. LEACH is a single hop protocol, which means each CH directly communicated with a BS, regardless of its distance far or near from the BS. Therefore, the CHs which exist far from the BS dies faster due to longer distance from a BS because they consume more energy to transmit data. LEACH network is made of two types of nodes such as CM and CHs. LEACH used hierarchical approach to arrange the network into a set of multiple clusters, where each cluster is managed by one cluster head. The CMs sensed the data and directly send to their respective CH. The CMs joined their nearest CH of high signal power value on the basis of principle of proximity. The CHs perform two tasks upon receiving of sensed data from the CM’s such as data aggregation to eliminate the data redundancy and data fusion before transmitting the data to the BS. Data aggregation and data fusion help LEACH to increase the network lifetime by minimizing the number of communication message transmission. When a CH dies due to battery depletion or node failure, LEACH will perform a new election to select new CHs, all nodes make own decision whether to become a CH or not, depending on their energy level in each round [5].

\[
T(n) = \begin{cases} 
\frac{p}{1 - p \ast \left(r \mod \left(\frac{1}{p}\right)\right)}, & \text{if } n \notin G \\
0, & \text{otherwise}
\end{cases}
\]

![Figure 3: LEACH Election Performing Equation](image)
Each round contains two phases to form clusters and perform data transmission, which are known as the Set-up Phase and Steady-state phase. In a Set-up Phase, LEACH performs election and randomly select CHs by randomly generating a number (n) between 0 and 1 for each node. If this randomly generated number is less than the threshold value calculated by threshold function T(n), the node would be selected as CH. Once the CH energy reached to certain threshold value, the LEACH will perform a new election to select a new CH. CMs who are previously selected as CHs will not become CH again in the next round. Upon a successful formation of CHs, each CH will broadcast information to its members via Code Division Multiple Access (CDMA), and they will join the nearest cluster head depending on Receiving Signal Strength Indicator (RSSI) value of advertising signal. CDMA is used to avoid a situation where a border node belonging to the cluster head ‘A’ distorts transmission directed to cluster head ‘B’. The CH use Time Division Multiple Access (TDMA) to offer data transmission time slot for every connected member; there is no peer to peer communication between CHs and nodes. Each CM is allocated a time slot for communication in order to minimize interference and collision among other CMs. The CMs go into sleeping mode with the exception of their time slots. Furthermore, each CH used different CDMA code to reduce likelihood collision inside and outside cluster sensor nodes and interference with other CHs, a TDMA based MAC protocol is used to manage balanced energy consumption in the whole network. Once the setup phase has been completed, the CHs will know their CMs and identifiers. CH construct a TDMA schedule, picks a CDMA code randomly and broadcasts this information in the broadcast schedule sub-phase. Whereas, in steady state phase CMs and CHs perform communication by transmitting or forwarding the data to the BS. CMs sensed the data and transmit to their respective CHs, whereas CHs perform data aggregation and data fusion activities, and then transmit the received data to the BS. The duration of setup phase is relatively shorter than the steady state phase, which minimizes the protocol overhead [5].

B. LEACH Enhancements

Later on, LEACH was improved and many developments were proposed to overcome its limitations. In LEACH Centralized (LEACH-C) network performance was improved 20% - 40% as it chooses the CHs according to their location information by adding GPS with individual sensor nodes [6]. LEACH-F was also proposed to enhance efficient clustering process by balancing the energy consumption between CMs and next CHs [6]. Solar aware LEACH (sLEACH) was proposed to extend network life by adding solar power to each node. Time-based Cluster-head selection algorithm for LEACH (TB-LEACH) was proposed to select a constant number of CHs autonomously without global information and increase the network life by 20% - 30% [7]. Clustered Diffusion with Dynamic Data Aggregation (CLUDDA) was developed to reduce redundant data transmission and enhance network lifetime. It also offers dynamic data aggregation nodes and improves network performance by even distribution of energy consumption within network nodes [8]. Furthermore, PEGASIS was proposed to increase the network lifetime by local coordination between closest sensor nodes. It also controls the data transmission to enhance network lifetime, PEGASIS protocols shows better performance as compared to LEACH and offers redundant data transmission. Later on, a better improvement was proposed in the form of TEEN a hierarchical protocol, which used the Hard Threshold (HT) and the Soft Threshold (ST) values to detect rapid changes occur in the environment and behave accordingly. HT and ST values are broadcast by CHs, when sensor nodes received HT values from CHs, they must transmit data to CHs on high priority; whereas, ST values represented low priority data. However, it also consumes more energy and shorten the network lifetime [9].

Many researches and improvement have been made to enhance the performance of LEACH. However, most of them are towards improving the network life and enhancing the cluster formation setup only, which shows a big research gap towards data availability and network consistency aspects WSNs.

C. Storage Node based Routing Protocols for Wireless Sensor Networks

In [10], a Storage Node (SN) hierarchical protocol for heterogeneous networks was proposed to offer data reliability. Each cluster consists of one CHs which has high storage and processing capability. CMs sensed the data and transmit to their nearest storage node or CHs according to their energy level and distance. CHs were integrated with a System-on-Chip (SoC) host platform and gigabyte range memory storage. This protocol was mainly developed for the healthcare application where all patient data is collected by SNs and SNs are continuous connected with power supply and will consider permanent CHs. SN hierarchical protocol shows less energy consumption and takes less communication time as compared to other hierarchical protocols. However, it increases the network cost by adding additional storage and battery power to SNs and if any storage node or CH is broken, the network does not have any data backup and all sensed data will be lost.

In [11], a high performance sensor storage and co-processing architecture was introduced to offer data backup and to minimise unnecessary data traffic over the network. They have developed a Co-S platform architecture that is integrated with a system on-chip (SoC) based host platform and a gigabyte scale energy efficient data storage low power system. He had used Sense-and-Send approach on sensor nodes for data transmission. Each sensor used a low power flash memory to store the sensed data along with key query results. The RISE storage board integrated with an SD Card and a NOR flash memory chip. Therefore, the SD card memory is used to store the sensed data; whereas, the NOR flash is used for random access of data to read or write as needed. The data is accessed via queries, in order to access specific data, the user need to write a manual query to extract the specific data from the SD Card. The system used a 16bit MCU (Renesas M16C/30280AFHP), USARTs (Universal Synchronous/Asynchronous Receiver Transmitter) and on-chip 8KB SRAM, 96KB Flash memory.

In [12], a secure and reliable data distributed storage scheme based on (m, n) Reed-Solomon (RS) codes was proposed to offer data backup and data reliability. This technique offers better data security and optimize data reliability, it also generates low storage heads and enhance the computational efficiency as compared to other storage node algorithm. However, adding storage nodes is the same issue as previous research.

In [13], the unlimited storage nodes idea was proposed. The proposed method was used to implement storage nodes to
control the heavy packet transmission to a BS by splitting the data across the network nodes. So even though one node dies or spoil, the duplicated data still can be accessed from the storage nodes. In this research paper the storage node placement problem was considered to minimize the total energy cost for gathering data to the storage nodes and replying queries.

In [14], a storage node base routing protocol was developed for healthcare applications. The developed protocol used hierarchical based network model in which proposed cluster heads was fixed and contained a large storage capacity along a large amount of power source. Therefore, each CHs contains large amount of memory and extra battery to support the CHs for longer time periods which increase the WSN deployment cost.

As a conclusion, there are a lot of improvements have been proposed for hierarchical protocols but each of them have their own pros and cons. However, adding extra storage nodes to support large amounts of memory also does not solve the current issue, instead it increases the WSN deployment cost.

Therefore, we are proposing an AI base framework solution, which should have the ability of self-perceive, adapt and learn according to the environment changes to offer significant network consistency and data backup on demand in a WSN to meet the user expectation. We are proposing an Artificial Intelligent base framework along Cognitive Radio (CR) at the physical layer to make the WSN more intelligent. The proposed framework will perceive the environment parameters as well as another node physical status before sending or receiving the appropriate data over the wireless medium, and adjust the communication parameters according to the network circumstances.

III. EXPERIMENTAL APPROACH

The main purpose of this research paper is to investigate and analyse the data storage and data unavailability research gap in the WSNs. This research could help in identifying data loss issues and analysing parameter which can help to improve network availability. Future solutions could be proposed to enhance and implement better data backup strategies for the data availability and consistent network performance in WSNs.

IV. RESULTS AND DISCUSSION

In the following experiments, we have simulated a hierarchical protocol LEACH in Matlab to analyse the data loss and data unavailability due to interference and BS failure.

A. LEACH Simulation Results

We have also performed several experiments to observe and understand the data availability at BS by changing two parameters such as number of sensor nodes (n) and probability of the cluster head (p). We have deployed two networks with 50 and 300 nodes and ‘p’ was kept at 10% and 20%. In all experiments the BS was static and fixed about 50 to 100 meters far away from the sensor nodes. Each experiment was run for 500 rounds. Table 1 shows the details of simulated parameters:

<table>
<thead>
<tr>
<th>Experiments</th>
<th>Simulation Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>n = 50, p = 0.1, Round = 500, Packets Size = 4000bit.</td>
</tr>
<tr>
<td>2</td>
<td>n = 300, p = 0.1, Round = 500, Packets Size = 4000bit.</td>
</tr>
<tr>
<td>3</td>
<td>n = 50, p = 0.2, Round = 500, Packets Size = 4000bit.</td>
</tr>
<tr>
<td>4</td>
<td>n = 300, p = 0.2, Round = 500, Packets Size = 4000bit</td>
</tr>
</tbody>
</table>

Figure 4 shows the overview of the cluster formation of four experiments. We can observe from the Figure 4 that the formation of cluster varies in each experiment. In first and third experiments the ‘p’ value were kept to be 0.1 (10%). However, in second and fourth experiments the ‘p’ value were kept to be 0.2 (20%). We can conclude from Figure 4 that in all experiments the formation of the cluster was unequal, different clusters have various numbers of nodes, which mean LEACH does not have knowledge of a number of cluster heads and number of nodes within each cluster. This unequal distribution of cluster heads and nodes within a cluster case of shortening network lifetime. Furthermore, the nodes and cluster heads far from the BS die faster as compared to the nearest nodes and cluster heads. This is because in single hop LEACH each cluster head directly communicate with the BS. Therefore, the CHs far from the BS dies faster as they need more power to transmit signal to cover longer distances.

We have also observed from the Figure 5 that when the value of ‘p’ was changed 10% to 20%, the number of the cluster heads was increased. In the first experiment the first node dies at 32nd rounds and almost all nodes die at 310th rounds. Whereas, in 3rd experiment the first nodes dies at 35th rounds.
rounds and almost all node dies at 290th rounds. Similarly, in 2nd experiment, the first node dies at 16th round and 300 nodes
dies in 500 rounds. Whereas, in 4th experiment, the first node
dies at 21st round and 300 nodes dies until 500 rounds. This
means, when the number of cluster heads increased, the
network lifetime decreased due to an unbalanced distribution
of the CHs in a network. Furthermore, when the number of
CHs increased, the network consumed more energy due to the
large number of CHs, which shorten the network lifetime.

Figure 6: Packets Transmission

Figure 6 shows the number of transmitted packets and the
number of packets received by CHs and BS. We can observe
that all experiments have a lower packet reception rate as
compared to the transmitted packets. Which means in a large
hierarchical network, large number of packets lost occur due
to network congestion, path lost or network interference.
Each time a packet is lost, it will be resent to BS or CH until
the source node received the acknowledgment. Therefore,
sending the same packet multiple time also creates a load on
the network, consume more energy and cause of shortening
the network lifetime and loss of data.

Data Receiving Percentage(%) at BS

Figure 7: Data Lost

Figure 7 shows the packets loss rate for all experiments.
We can observe that in scenario 1 the data lost rate was 20%,
in scenario 2 was 30%, in scenario 3 was 25% and in scenario
4 was 40%. So, whenever data is lost, it related to packets loss
which mean the packets will be retransmitted until
acknowledgement is received from the receiver. This means
if the packets loss rate is higher, more energy will be wasted
to retransmit same packets again and again which will shorten
the network lifetime.

Figure 8: Data availability at BS

Figure 8 shows the data availability at BS; when BS is active
(1 refer to data is received and 0 mean data is not received).
From the Figure 8, we can observe that when the number of
CHs increase, the more traffic is generated and more packets
are transmitted towards a BS. After a certain time, when the
number of CHs reduces, the traffic load on BS also deceased.
Therefore, the data availability at the BS increased while the
number of cluster heads reduced. However, when a number of
cluster heads increased, it also increases packets loss due to
high interference and network congestion.

The Figure 9 shows the data unavailability at BS, when BS
is inactive, so all transmitted data is lost. There is no
availability or data acknowledgement because data is only
saved on the BS. CHs have no knowledge of BS activity, when
CHs send the packets to BS, if there is no reply or
acknowledgment from the BS, the CHs will go into sleep
mode. In a normal process, the BS will send advertisement and
CHs will reply to join. Once connection is established the CHs
will start to send the sensed data. But if BS is down or inactive,
the CHs will only communicate with CMs but will not
communicate with BS which will cause of data loss. This is
because CHs are only responsible to forward data, perform
data aggregation and data fusion services.

Figure 9: Data Lost at BS

V. FUTURE WORK AND CONCLUSION

Current hierarchical protocols enhancement mostly focuses
on extending the network lifetime and equally distributing of
CHs. Storage nodes and distributed storage resource are also
proposed to offer data backup and data reliability. Most of the
solutions are designed to minimise interference and avoid
packet lost rather than making the sensor nodes more
intelligent and efficient upon node or network or hardware
failure. Increasing a sensor node memory is not an effective
solution to offer a backup or data reliability because the user
also needs to consider about cost and power consumption.
A larger size of memory consumes more power and increase
sensor cost. The main purpose of deploying a Wireless Sensor
Network is to be low cost as $1 to 10$ per node and efficient.
Normally, in paddy field hundreds sensor nodes are deployed
to cover a large area. Therefore, the cost of each node is very important for the user.

CBRPs are good and strong, but CHs are unable to make decisions upon network failure. CHs are not responsible to store or save the data, they are only responsible to perform sensing and forwarding services. This mean WSN performance could be easily compromised if the server or storage device or BS is down, which raised a big question on the reliability of WSN.

In order to enhance the WSN reliability and data consistency, we are proposing an Artificial Intelligent base solution, which should have the ability of self-perceive the environment, adapt and learn according to the environment changes to offer significant network consistency and data backup on demand in a WSN to meet the user expectation. The proposed system will have ability to detect dead nodes and BS status before the data transmission, and in a case when BS or CHs are unavailable the data will be saved at the sensor node for temporary purpose. Once the BS or CHs are active the stored data will be transmitted to the related CHs or BS. This process will enhance the data availability and performance of any WSN. Additionally, the system will have ability to automatically adapt the changes to compensate any node failures and data available for the users.

At the end, we can conclude that most of the research and development work only focusing to offer data availability by increasing the storage capacity. Industry hardware and protocols improvement do not relate to offering any backup service, except offering online backup which can be performed at BS only [15]. Therefore, the WSN field still required an intelligent, efficient and consistent architecture to support data availability and data backup upon any node or system failure. The proposed framework will help to enhance the network consistency and data availability in WSNs.

ACKNOWLEDGMENT

I would like to thank Universiti Malaysia Sarawak for their support and assistance.

REFERENCES