Aware Diffusion: A Low Power and Reliable Delivery Routing Protocol for Wireless Sensor Networks

Kamil Samara¹, Hossein Hosseini²
Department of Computer Science¹, Illinois Wesleyan University¹, USA¹; Department of Electrical Engineering and Computer Science², University of Wisconsin - Milwaukee², USA²
ksamara@iwu.edu¹; hosseini@uwm.edu²

ABSTRACT
Directed diffusion is a well know routing algorithm for Wireless Sensor Networks (WSNs) that was designed in 2003. Directed diffusion saves energy by sending data packets hop by hop and by enforcing paths to avoid flooding. The main disadvantage of directed diffusion is that it is not aware of the network and nodes status and as a result cannot compare potential paths to find the best or healthier path. As an effort to try to overcome this disadvantage, we propose aware diffusion protocol. Aware diffusion collects data about the available paths and uses these data to enforce healthier paths using machine learning. The data collection is done by adding a new stage called data collection stage. In this stage the protocol designer can determine which parameters to collect then use these parameters in enforcing the best path according to a certain criteria. In our implementation of this paradigm we are collecting total energy on the path, lowest energy level on the path, and hop count. Again, the data collected is designer and application specific. The collected data will be used to compare available paths using non-incremental learning, and the outcome will be preferring paths that meet the designer criteria. In our case, healthier and shorter paths are preferred, which will result in less power consumption and higher delivery rate since healthier and less nodes will be doing the work.

KEYWORDS: DIRECTED DIFFUSION; SHORTEST PATH; WIRELESS SENSOR NETWORKS; HOP COUNT; RELIABLE DELIVERY.

CONCLUSION
Routing in WSNs is challenging task due to energy and hardware constrains. Directed diffusion is well known routing protocol but it is not aware of the network status and does not use any data about the network status in making routing decisions. We believe that directed diffusion could be improved by collecting data about the network status and use this data in enforcing routing paths. In this paper we proposed an aware diffusion routing protocol. Aware diffusion is an energy efficient and reliable routing protocol for wireless sensor networks which collects energy levels and hop count information about potential paths then through non-incremental machine learning gives preference to healthier and shorter paths.

The performance of aware diffusion was compared with that of directed diffusion through simulation experiments. Two experiments were conducted. The first experiment compared energy consumption and the second experiment compared data delivery. From the experiments results it is concluded that aware diffusion performed better in these two aspects. Also as the network size increased the difference in the performance in these two aspects increased.

Quick Response Code

Access this article online
Paper URL: http://ijrws.org/index.php/ijrws/issue/view/15
ISSN: 2320-3617

Citation
Aware Diffusion: A Low Power and Reliable Delivery Routing Protocol for Wireless Sensor Networks

Kamil Samara  
Department of Computer Science  
Illinois Wesleyan University  
Bloomington, Illinois 61701, USA  
ksamara@iwu.edu

Hossein Hosseini  
Department of Electrical Engineering and Computer Science  
University of Wisconsin - Milwaukee  
Milwaukee, Wisconsin 53211, USA  
hosseini@uwm.edu

Abstract - Directed diffusion is a well know routing algorithm for Wireless Sensor Networks (WSNs) that was designed in 2003. Directed diffusion saves energy by sending data packets hop by hop and by enforcing paths to avoid flooding. The main disadvantage of directed diffusion is that it is not aware of the network and nodes status and as a result cannot compare potential paths to find the best or healthier path. As an effort to try to overcome this disadvantage, we propose aware diffusion protocol. Aware diffusion collects data about the available paths and uses these data to enforce healthier paths using machine learning. The data collection is done by adding a new stage called data collection stage. In this stage the protocol designer can determine which parameters to collect then use these parameters in enforcing the best path according to a certain criteria. In our implementation of this paradigm we are collecting total energy on the path, lowest energy level on the path, and hop count. Again, the data collected is designer and application specific. The collected data will be used to compare available paths using non-incremental learning, and the outcome will be preferring paths that meet the designer criteria. In our case, healthier and shorter paths are preferred, which will result in less power consumption and higher delivery rate since healthier and less nodes will be doing the work.

Keywords— directed diffusion; shortest path; wireless sensor networks; hop count; power consumption; reliable delivery.

1. INTRODUCTION

Recent advances in technology especially in electronics and communications allowed the emergence of WSNs. A WSN is a collection of sensor nodes.

Sensor nodes have sensing capabilities which makes them a suitable solution for sensing and collecting data from different environments [1].

Sensor nodes are low cost nodes with limited computing power, scarce memory, low bandwidth, and most importantly limited energy source. Sensor nodes are usually operated by batteries, which in most cases are not rechargeable or easy to replace [2].

A WSN is a collection of hundreds or thousands of wireless sensor nodes that are often deployed in remote areas as shown in Fig. 1, whose job is to collect data wirelessly and deliver it to a base station. Each node contains a sensing component, processor, communication, and storage components [3].

WSNs have a wide range of applications in different fields, but the most common use for WSNs is monitoring. The WSN is usually deployed over a region or a structure to monitor one or more phenomenon then reporting the reading collected by its sensors to base station (sink node) which can convey the aggregated data to a human operator for further processing.

Some common examples of WSNs applications are military, where sensor nodes are used to detect enemy movement, traffic control, where sensor nodes collect data about car jams, health sector to monitor a patient’s condition, and many more [1].

Also WSNs are a backbone for the new emerging Internet of Things (IoT) technology. WSNs can cooperate with RFID systems to better track the status of things, i.e., their location, temperature, movements, etc. As such, they can augment the awareness of a certain environment and, thus, act as a further bridge between physical and digital world [4].

The special nature of WSNs, mainly the limited energy source in addition to low computation and memory capabilities, made traditional routing algorithms unsuitable for WSNs [5].
This motivated researchers to design routing algorithms that suit the needs and nature of WSNs. One of the early and well-known routing algorithms for WSNs is directed diffusion.

WSNs routing algorithms are classified into data-centric, hierarchical, location-based, and QoS routing. Directed diffusion falls into the data-centric category [6].

Directed diffusion tackles WSN limitations by introducing the naming scheme, traversing packets using multi-hop, and enforcing certain paths to avoid flooding [7].

Some of the major shortcomings of directed diffusion are it does not ensure shortest path, does not consider energy level, and does not avoid critical nodes [8].

In this paper, we will be introducing a modified directed diffusion that ensures data is being sent through the best path between the source and the sink according to path length and energy level metrics. This will result in less and healthier nodes being used to transmit the same data which in turn will result in less energy consumption and longer network life time.

The rest of the paper is organized as follows: section II is dedicated to Data-Centric protocols. In section III we will give a general overview of directed diffusion and how it works. Section IV will be briefly touch on some of the attempts that were made to improve on directed diffusion and their shortcomings. A short introduction to machine learning is presented in section V. Our proposed protocol is described in section VI. Section VII is dedicated for simulation results and finally conclusion is presented in section VIII.

II. DATA-CENTRIC AND FLAT-ARCHITECTURE PROTOCOLS

As mentioned earlier, the huge number of sensor nodes makes it very hard to assign IDs to nodes. Therefore, data-centric protocols treat all nodes equally, so the focus here is the data not the nodes. Data here is identified by attributes; the requesting data is done by the attributes of the phenomenon [9].

Data-centric networks have flat structure where all nodes play the same role in routing the data and all nodes collaborate to perform the routing task. This gives data-centric networks the advantage of simplicity where no topology management is required [5].

Examples of Data-centric protocols are: Flooding, Gossiping, SPIN, and Directed Diffusion [10].

III. Directed diffusion

Directed diffusion is a data-centric data dissemination protocol that is also application-aware in that data generated by sensor nodes is named by attribute-value pairs. The main idea of directed diffusion is that nodes request data by sending interests (also called queries) for named data. This interest dissemination sets up gradients within the network that are used to direct sensor data toward the recipient, and intermediate nodes along the data paths can combine data from different sources to eliminate redundancy and reduce the number of transmissions [11].

Directed diffusion does not rely on globally valid node identifiers, but instead uses attribute-value pairs to describe a sensing task and to steer the routing process. For example, a description for a simple vehicle-tracking application could be:

```
type = vehicle // detect vehicle location
interval = 20 ms // send data every 20 ms
duration = 10 s // perform task for 10 s
rect = [-100, -100, 200, 200] // from sensors within rectangle
```

That is, a task description expresses a node’s desire (or interest) to receive data matching the provided attributes. The data sent in response to such interests is also named in the same manner, that is, using attribute-value pairs [11].

Once an application has been described using this naming approach, the interest must be diffused through the sensor network. This process is shown in Fig. 2. A sink node periodically broadcasts an interest message to its neighbors, which continue to broadcast the message throughout the network. Each node establishes a gradient toward the sink node, where a gradient is a reply link toward the neighbor from which the interest was received. As a consequence, using interests and gradients, paths between event sources and sinks can be established. Once a source begins to transmit data, it can use multiple paths for transmission toward the sink. The sink can then reinforce one particular neighbor based on some data-driven local rule. For example, a sink could reinforce a neighbor from which the sink has received a previously unseen event. Toward this end, the sink resends the original interest message to the neighbor, which in turn reinforces one or more of its neighbors based on its own local rule [12].

![Fig. 2. (a)Interest propagation (b)gradients setup (c)data delivery](image)

IV. DIRECTED DIFFUSION IMPROVEMENTS

Several attempts were made to improve on directed diffusion after it had been initially developed. In [13] they proposed an energy-efficient diagonal-based directed diffusion. In [14] a passive clustering approach was used. In [15] they also used clustering to improve the performance of directed diffusion.

Most of these attempts focus on topology change to increase scalability and minimize the cost of flooding the interests. These approaches add topology management
overhead and rip directed diffusion from its simplicity. None of these attempts consider shortest paths, load balancing, or critical nodes.

V. MACHINE LEARNING

Machine learning is a field of Artificial Intelligence. Machine learning focus on improving a learned concept. The learning process is achieved by feeding the system with training examples form the environment. The system uses this data to form a general concept.

Machine learning could be categorized into incremental and non-incremental learning. In non-incremental learning the concept is learned from an initial set of data and not modified after that until a new process is initiated. In incremental learning on the other hand the learning process continues after the initial set of data by continuously collecting data about the environment and refining the learned concept [16].

Based on the fact that transmission is the main source of energy depletion in WSNs, in our work we decided to use non-incremental learning to avoid the continues data collection required in incremental learning which will result in more data collection packets being sent on regular bases, and as result more energy consumption that overweighs the benefit of improving the learned concept [17].

VI. PROPOSED PROTOCOL

This section will be dedicated for describing the design of our proposed modified directed diffusion routing protocol.

The design of the reinforcement scheme in directed diffusion is generally targeting minimum delay or maximum number of packets delivered. The problems with this design are:

a) The routing path is generally not the healthier path; path with higher energy levels.

b) The nodes on the path may have unbalanced remaining energy, i.e. some nodes may die sooner than the others.

c) The routing path is generally not the shortest path available.

We try to solve these problems by collecting data about all the available paths from the sensing node to the sink and then use these collected data to decide the best path using non-incremental machine learning according to our criteria.

The collected data are total energy on the path, number of nodes on the path (Hop Count), and the lowest energy level of a node on the path to be able to identify critical node.

Critical nodes are nodes which have been used more than others due to their location, which results in energy drainage.

The steps of our proposed routing algorithm are shown in Fig. 3 and are described below:

The steps of our proposed routing algorithm are shown in Fig. 3 and are described below:

![Flow chart of algorithm steps](image-url)

A. Interests Propagation

Interests are flooded through the sensor network. For each active task the sink will broadcast an interest message shown in Fig. 2 to all its neighbors. Each node that receives an interest message will also broadcast it to all its neighbors.

<table>
<thead>
<tr>
<th>Query ID</th>
<th>Source Address</th>
<th>Destination Address</th>
<th>Time to Live</th>
<th>Requested Data</th>
</tr>
</thead>
</table>

Fig. 4. Interest Packet

Every node maintains an interest cache. Each item in the cache corresponds to a distinct interest. Two interests are distinguished by the ID field.

Interest entries in the cache do not contain information about the sink, but just about the immediately previous hop. Also identical interests are aggregated into a single entry.

When a node receives an interest, it checks to see if the interest exists in the cache. If no matching entry exists the node creates an interest entry. This entry has a single gradient (a gradient specifies a direction in which to send events) pointing toward the neighbor from which the interest was received. If an interest entry does exist, but no gradient for the sender of the interest, the node adds a gradient with the specified value.

Finally, if both an entry and a gradient do exist, the node simply updates the attribute fields if they are different.

When an interest’s entry has expired, the interest’s entry is removed from the interests’ cache.

Not all received interests are resent. A node may suppress a received interest if it recently resent a matching interest.

B. Query Response

The query response stage is one of the major changes we introduced to directed diffusion. This stage does not exist in the original directed diffusion. This stage is which allows us to collect the needed data about each available path and then use these data in enforcing the best path according to our criteria.

In our case we are collecting total energy on the path, number of nodes on the path (Hop Count), and the lowest energy level of a node on the path to be able to identify critical node, but the set of collected data could be different from one application to another according to what is important to the application in use.

For example, an application that is concerned about fast delivery could collect data about nodes’ buffer size to avoid congested nodes.

When a node has at least one active interest, the node will switch on its sensors and start sensing for the requested data.

If the sensing node senses data that matches the requested data by the interest it will generate a Query Response Packet and send a copy of it to all the gradients associated with the interest.

The base station will receive the Query Response Packet through multiple paths. Then the base station can choose the shortest path and reinforce the source sensors to use the chosen path by using enforcing packets.

Forwarding nodes, on the other hand, could receive the same Query Response Packet flooded by the sensing node from multiple neighbors but it will only forward one of them.

<table>
<thead>
<tr>
<th>Query ID</th>
<th>Source Address</th>
<th>Destination Address</th>
<th>Hop Count</th>
<th>Total Energy</th>
<th>Lowest Energy</th>
</tr>
</thead>
</table>

Fig. 5. Query Response Packet

C. Reinforcing Paths

When the base station starts receiving Query Response Packets in the reply of an interest that was propagated earlier, it will receive the packets through multiple paths. This is due to the source node sending the Query Response Packet to all the nodes from which it received the interest propagation packet.

Each Query Response Packet received by the base station will hold hop count, total energy, and lowest energy fields about the path it took. Based on that information the base station will choose the best path.

Choosing the best path is done by calculating the promising factor (PF) for each path from the source (sensing node) to the destination (sink node) using the following formula:

$$PF_s \rightarrow d = \frac{\alpha(TE) \cdot \beta(LE)}{\gamma(HC)} \quad (1)$$

Where:
TE: Path Total Energy Ratio
LE: Path Lowest Energy Level Ratio
HC: Path Hop Count Ratio
\(\alpha, \beta, \gamma\): are the weights given to each factor in the equation (default value for all weight is 1)

The term TE in equation (1) will basically give preference to paths with higher energy, which is an indicator of the health level of a path. The LE will help us avoid critical nodes, which has been used more than others and their energy level dropped significantly. The term HC will give preference to shorter paths over longer ones which will result in faster delivery.

The Path Total Energy Ratio (TE) is calculated using the following formula:

$$TE \text{ for } \text{path } s \rightarrow d = \frac{\text{Paths } s \rightarrow d \text{ Total Energy}}{\sum \text{All Paths } s \rightarrow d \text{ Total Energy}} \times 100 \quad (2)$$

Where:

$$\text{Paths } s \rightarrow d \text{ Total Energy} = \sum_{\text{All nodes on path } s \rightarrow d} \text{Node Energy Level} \quad (3)$$
The Path Lowest Energy Level Ratio (LE) is calculated the following formula:

\[
LE_{\text{for } s \to d} = \frac{\text{Path } s \to d \text{ Lowest Energy Level}}{\sum \text{All Paths } s \to d \text{ Lowest Energy Level}} \times 100
\]

Where:

\[
Path_{s \to d} \text{ Lowest Energy Level} = \min_{\text{All nodes on Path } s \to d} \text{ (Node Energy Level)}
\]

The Path Hop Count Ratio (HC) is calculated the following formula:

\[
HC_{\text{for } s \to d} = \frac{\text{Path } s \to d \text{ Hop Count}}{\sum \text{All Paths } s \to d \text{ Hop Count}} \times 100
\]

Where Hop Count is the number of nodes on path \( s \rightarrow d \).

After choosing the best path, the base station will send an enforcing packet to the neighboring node that forwarded the Query Response Packet containing the best path information. In turn each forwarding node on the path of the enforcing packet will forward the enforcing packet to the node it received the Query Response Packet from. The forwarding node will know which node to forward the enforcing packet to from the interests table because the interest table of each node will cache the Query Response Packets forwarded in response to the interest we are currently working on. And the cached Query Response Packet contains the source node address.

D. Data Propagation

After the reinforcing phase is done the source nodes know which neighboring node to use to forward the data packets.

So every time it senses a matching data to interest requested data it will generate a data packet and forward the data packet towards the base station using the enforced node.

Every node along the path will do the same thing and forward the data packet through the enforced neighbor until the data packet reaches the base station.

This process will continue until the “time to live” field associated with the interest becomes zero. Then this interest will be removed from the table of interests in the source node and no more data packets will be generated in response to this interest.

VII. SIMULATION AND RESULTS

To compare the performance of aware diffusion with directed diffusion we have simulated both protocols using the same simulation parameters shown in table I.

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation Time</td>
<td>300 sec</td>
</tr>
<tr>
<td>X axis</td>
<td>40 – 180 meters</td>
</tr>
<tr>
<td>Y axis</td>
<td>40 – 180 meters</td>
</tr>
<tr>
<td>Number of Nodes</td>
<td>25, 256</td>
</tr>
<tr>
<td>Sink node</td>
<td>Node 0</td>
</tr>
<tr>
<td>Radio Type</td>
<td>CC2420</td>
</tr>
<tr>
<td>MAC Protocol</td>
<td>TMAC</td>
</tr>
</tbody>
</table>

Two simulation experiment were conducted. The first simulation experiment focus was to compare the total energy consumption between the two protocols in quest to prove that aware diffusion will result in less energy consumption.

The second simulation experiment focus was to compare the total number of packets delivered to the sink node between the two protocols in quest to prove that aware diffusion will provide higher reliable delivery of data packets to the sink node.

A. Experiment One: Energy Consumption

The simulation was performed with different number of nodes ranging from 25 to 256 to prove that energy conservation will occur regardless of networks size.

Fig. 6 shows the simulation results comparing the total energy consumption (in Joules) for both directed diffusion and aware diffusion.

As shown in Fig. 6 the total energy consumption was always less in the case of aware diffusion.
Actually as shown in Fig. 7 the difference in the total energy consumption was increasing as the number of nodes increased because in bigger networks the paths tend to be longer and assuring the healthier and shorter paths will decrease energy consumption significantly.

As shown in Fig. 8 the total number of packets was always more in the case of aware diffusion.

Actually as shown in Fig. 9 the difference in the total number of packets delivered was increasing as the number of nodes increased because in bigger networks the paths tend to be longer and assuring the healthier and shorter paths will increase packets delivery.

### B. Experiment Two: Reliable Delivery

The simulation was performed with different number of nodes ranging from 25 to 225 to prove that more packets will be delivered regardless of networks size.

Fig. 8 shows the simulation results comparing the total number of packets delivered to the sink for both directed diffusion and aware diffusion.

The performance of aware diffusion was compared with that of directed diffusion through simulation experiments. Two experiments were conducted. The first experiment compared energy consumption and the second experiment compared data delivery. From the experiments results it is concluded that aware diffusion performed better in these two aspects. Also as the network size increased the difference in the performance in these two aspects increased.

### REFERENCES

Aware Diffusion: A Low Power and Reliable Delivery Routing Protocol for Wireless Sensor Networks


[18] A. Boulis, “Castalia user’s manual”, NICTA.