FSS: Fuzzy Supervised Learning for Optimal Path Selection in RPL

D. Gopika and P. Rukmani
Vellore Institute of Technology, Chennai, India
E-mail: gopikablue18@gmail.com

Abstract. IPv6 routing protocol for low power lossy network (RPL) is a standardized routing protocol developed by the IETF ROLL working group. RPL has attained a lot of attraction to the research community as it demands modification adding to its improvement. RPL constructs a Destination Oriented Directed Acyclic Graph (DODAG) considering the metrics and constraints through explicit Objective Functions (OFs). The OFs chooses the best parent for reliable data communication. The existing objective function due to a single metric lacks in satisfying the requirement of real-time IoT applications. Still, the mobility aspect of the node is not considered in the objective function for optimal path selection. In the proposed methodology a Fuzzy Supervised Learning algorithm for quality node prediction using the composite metric based OF from the candidate parent set. The testing phase shows 93% accuracy of prediction with Artificial Neural Network (ANN) classifier which outperforms Naïve Bayes and Decision Tree (DT).

1. Introduction

The Internet of things (IoT) is well-thought-out as an innovative paradigm that enables smart objects to connect over the internet. Various domains are covered by IoT today such as medical, industrial applications, smart cities, agriculture, and real-time monitoring, to make our day today life easier. The miniature structure of the smart devices utilized in the IoT environment is limited in terms of resources such as power, energy, and memory. WSN is a base of IoT. Smart objects of IoT are sensors deployed in a large area for sensing the information. WSNs are a collection of densely populated sensor nodes that perform event-based monitoring. The major concern of working with sensor nodes is its resource-constrained nature. A sensor node can be stationary or mobile, consisting of a processing unit, sensing unit, communication unit, and power source. The characteristics and requirements of WSNs include:

a) Energy: Nodes are energy deficient with the battery as a power source. They consume more energy in transmission and receiving activities. Energy is a primary metric for defining the lifetime of the node. It is a major factor that requires analysis.

b) Cost: They are generally cheap due to which we see around a hundred to thousand nodes being deployed in the network setup phase.

c) Communication: Connectivity between the nodes remains wireless which reduces the cost that is incurred in installing and maintaining the network.
d) **Distributed Sensing and Processing:** In general, the node deployment in WSNs is densely scattered which allows the sensing from multiple nodes as compared to a single node. This in turn creates a robust environment to overcome unexpected threats.

e) **Topology:** Multi-hop topology is supported in WSNs for long-range communication to lessen the energy utilization in nodes.

The standard routing protocol for the low power lossy nodes is IPv6 protocol for low power and lossy networks (RPL). Fig.1 shows the RPL topology. It has a lot of limitations as it deals with the existing single metric objective function (OF). The major advantage of RPL is its flexibility to adapt to the modifications and handle the frequent topology changes [1][2].

![RPL DODAG Structure](image)

**Figure 1.** RPL DODAG Structure

### 1.1. Open Challenges in IoT Routing

Considering the IoT trend these days worldwide, there is an increase in the connected devices, investments in technology, and also a large group of people who have shown interest in the IoT platform. However, there are various challenges involved in IoT that incurs to threats listed below.

a) **Mobility:** The mobility aspect of nodes creates various challenges in network management. More research is recommended in the movement detection of the nodes due to frequent location and topology changes. The existing routing protocols that support mobility are inefficient in managing IoT based devices due to their resource constraining behavior.

b) **Stability Problem:** The number of devices connected to the network in IoT applications is expanding tremendously. In the upcoming years, more than 50 billion devices will be linked to the internet which is more than the human population. It results in the enlargement of network size. There is an exponential increase in the network size resulting in difficulty in managing the network.

c) **Big Data:** IoT information is portrayed by heterogeneity which implies that data are created in a huge amount. The challenge of taking care of enormous information is critical as it affects the overall performance. Implementation of artificial intelligence-based IoT applications for handling data management will lead to better results.

d) **QoS Approaches:** Devices used in IoT applications have become independent and intelligent with advancements. According to the IoT application, the requirement varies. There are limitations concerning energy consumption that can be resolved by incorporating artificial neural networks. Energy management: Existing routing protocols are focused on communication with low power consumption mechanisms and are now in their initial stage of evolution. Hence, more sustainable technologies need to be evolved that support green computing.

e) **Security:** Challenges in IoT concerning security constitute a major area of research. With an increase in advancements in technology and new inventions that support mobility-based applications,
the threat of data confidentiality and access control exists with no trusted platforms. Due to various opinions and wrangle regarding data confidentiality, control, and sustainability factor in IoT.

The research community has a wide area with unsolved problems to be answered to bring advancement in IoT with a secure environment for improving technological development. Recent works focus on the security and privacy aspects rather than the underlying architecture. However, interoperability among technologies has shown better evolution in recent years. IoT, the future of the upcoming technological foundation faces vulnerability with its networking bases being unset and requires improvement and upgrade in the existing networking technology.

RPL provides support to Low power Lossy Networks (LLNs) in congestion avoidance, energy efficiency, and managing the network load. The main issue of RPL is to find the best patent node for best-path selection. The advancements in IoT applications lead to improve the quality of data delivery between the nodes. Thus, it is the need for the node to choose the optimal path for data transmission. The key aspect to understand is LLNs being lossy requires efficient node selection for data communication to reduce the packet loss and energy consumption of nodes. The optimized mechanism for the selection of quality nodes with the appropriate objective function is a necessity in RPL. LLNs with various issues in routing like mobility, topology changes, and stability affects the lifetime of the network and degrades the overall performance of the network. Sudden changes in the network topology and environment factors drastically affect the network quality parameters by increasing the end to end delay, packet loss, and node energy utilization. The main objective is to define an optimized objective function that is appropriate for IoT applications. The most suitable path is selected by combining both link and node metrics of the node such that the network performance is not affected and QoS is maintained in the network.

2. Literature Survey

IoT based applications face challenges in routing the sensed data from source to destination. The goal of routing protocols is to consider the packet delivery rate, network stability, convergence, reliability, energy consumption, etc. Currently, routing protocols are not just intended towards reliable delivery of data but are reformed to energy-aware routing algorithms to improve network longevity. The node energy gets drained during its communication phase. Thus energy-efficient protocols are designed considering less energy usage in the communication phase with efficiency and reliability as major considerations. Most of the routing protocols prefer the shortest route path to the destination that could lead to a network partition, as there is a high probability of the nodes to die soon. Table 1 shows the work related to energy-efficient routing in IoT. Important concepts related to energy efficiency used to estimate the performance of routing protocols are as follows:
a) **Lifetime of Node:** The lifetime of a node can be determined using the energy of the node. It is necessary to improve the lifespan of the node to avoid frequent topological changes due to the death of nodes.

b) **Distance:** The selection of relay nodes for the exchange of packets between nodes to reach the destination involves distance as a primary metric. Most of the existing protocols select a short route for packet transmission. But the repeated selection of such short routes may lead to disconnecting networks due to loss of nodes.

c) **Energy Consumption:** Routing protocols are modelled with less energy consumption design. Routes with high energy remaining nodes are considered.

d) **Active nodes:** This metric indicates the network lifetime.

e) **Idle Listening:** Nodes consume energy in sensing for the packet or traffic that is not being sent.

f) **Collision:** In a state when a large volume of data gets generated in a node, it leads to a collision that requires retransmission of the packet that consumes the node’s energy.

g) **Packet Size:** The Packet size indicates the transmission time. It can be reduced by compression or by sending multiple small packets as one large data packet.

h) **Residual Energy:** High residual energy is an important metric for a node to be selected as a relay node because it assures network connectivity and the survivability of the network.

| Methodology                                | Advantage                                      | Disadvantage                                      |
|--------------------------------------------|------------------------------------------------|--------------------------------------------------|
| Simulated using Contiki RPL [3]            | Increases lifetime of a node                   | Only the energy utilization metric is considered  |
| Energy component used as the main metric [10]| ETX and mobility of node is considered         | Results do not show many improvements when compared to RPL |
| Cost calculation using multiple metrics [11] | Suitable for industry-based application with increased lifetime | Mobility not supported                            |
| Collaborative strategy [9]                 | The use of optimized mechanisms shows good results in terms of network lifetime | Performance evaluation is not done in a real-world scenario |
| Residual energy used as a metric [12]      | The Lifespan of the network is increased       | Practical testing not done                        |
| Fuzzy logic-based approach [5]             | Increase throughput with better network performance | Mobility not supported                            |
| Fuzzy logic and corona approach [13]       | Supports mobility and enhances lifetime, latency, Packet delivery ratio | Scope of mobility is less                         |
| Data aggregation for minimum energy consumption [14] | Increases lifetime of nodes in the network | Mobility not supported                            |

In recent years advancements in RPL is happening to improve the parent selection by optimizing the objective function of RPL with the combination of node and link metrics. In [3], the authors proposed a new composite objective function OF-FUZZY with ETX, HC, and Energy consumption (EC) with the fuzzification approach. The proposed idea shows improvements in overall performance in terms of delay, energy, latency, and network stability. A novel OF formulated with delay, hop
count, ETX, and Link quality as a metric with the combination of corona mechanism to handle the node mobility in the network [4].

A single metric objective function mechanism was proposed in [5], using remaining energy for optimal path selection for data transmission. It improves network performance by managing the energy distribution equally to all nodes. In [6], the authors proposed a fuzzy- LEACH mechanism using battery level indication, distance, and congestion as the main metric as its objective function. Fuzzification with clustering approach shows improvements by reducing the energy consumption of nodes and increasing the network lifespan. Authors in [7], uses ETX and residual energy as a combination in the objective function. The simulation results of the proposed objective function using Contiki OS shows improvements in network lifespan when compared with MRHOF.

LLNs being low power in nature require an energy-efficient routing mechanism. In [8], the authors proposed a methodology with single and combination metrics using hop count, remaining energy, and ETX. It shows improvements in terms of power consumption and packet loss. But it fails to consider the link metrics.

With the advancements in IoT lot of smart applications are in development to make our day today life easier. IPv6 provides solutions to manage battery-based smart meters in RPL. In [9], it formulates a new objective function with ETX and energy consumption with additive methodology. It improves the network lifespan by 27% with equal energy distribution. It lacks a comparative study with other works algorithms.

3. Proposed Methodology

The proposed work focus on formulating an optimized objective function using the multi-metric approach with the combination of both link and node metrics. The novel mechanism of combining fuzzy logic with supervised learning for the best parent prediction to choose the optimal path for data transmission. DODAG topology of RPL requires the optimal parent selection from the candidate set for reliable data delivery.

3.1 Metric

LLNs most suitable routing protocol RPL with DODAG topology require to select the optimal path from the source node to the destination node considering the metrics of the objective function. RSSI (Received Signal Strength Indicator) is used to measure the signal strength for the selection of quality links based on the power consumption during data delivery. It indicates that the high value of RSSI is equal to a strong link with reliability. The RSSI value ranges from 10dBm to 92dBm. Due to the mobile nature of the nodes, the signal strength of the node is higher when it is closer to its associated node called the parent node. RE (Remaining Energy) of the node is measured using Eqn. (1) and Eqn. (2).

Energy metric helps to identify the best node in the network [15]. While choosing the optimal path with the quality parent in the DODAG queue utilization metric of the nodes helps in identifying the congestion state of the node. If queue utilization (QU) of a node is high, it is considered as a congested node with large data in the queue. A node when selecting a parent node checks for the less congested to achieve QoS. Eqn. (3) is used to calculate the node queue utilization.

\[
\text{RE}(i) = \text{TE}(i) - \text{EC}(i) \quad (1)
\]

\[
\text{EC}(i) = \text{ETX} + \text{EL} \quad (2)
\]

\[
\text{QU} = \frac{\text{No. of packets in the Queue}}{\text{Total Queue Size}} \quad (3)
\]

3.2. Fuzzification

The input parameters are taken as crisp values such as RSSI, RE and QU are considered as metrics with fuzzification to estimate the node of the quality. Linguistic variables for the given input parameters are calculated with the mamdani inference system [16].
3.3. Fuzzy logic for quality parent prediction

The need for improvement RPL based routing protocols due to the mobile nature of sensor nodes because of sudden topology changes. Our proposed methodology handles the mobile behavior of nodes by analyzing the energy, load, and signal strength of the sensor nodes. The optimized objective function uses the remaining energy to indicate the lifetime of the nodes. The quality node is predicted by training the model with the optimized objective function dataset. The accuracy of the trained dataset is tested using the multi class classifier algorithms. The classifiers involved in the testing phase of the trained dataset in the proposed model are as follows:

3.3.1. Artificial Neural Network (ANN)

A neuro fuzzy-based system is a layered approach. The first layer takes the input parameters of the objective function as input. The hidden layer evaluates the quality of the node based on the linguistic variables generated from the fuzzy membership functions. The output layer predicts the output target label. The weights of the layers are obtained from encoding the fuzzy sets randomly. Fig.3 shows the ANN multilayer operation.

![ANN Multilayer Operation](image)

Figure 3. ANN operation

3.3.2. Naïve Bayes

Naïve Bayes is a multi-class classifier that handles large datasets. It is the most efficient algorithm as it is quick in classification and prediction. Naïve Bayes uses a supervised learning approach. It is based on Bayes theorem with the naïve assumption that all predictors are independent. The Gaussian probability distribution is used to estimate the mean and standard deviation of the trained data set using Eqn. (4).

3.3.3. Decision Tree (DT)

The decision tree is an easy and commonly used multi-class classification as it is simple to interpret. The optimal path selection using a quality node towards the DODAG root from the sensing leaf nodes constitute classification rules. The decision tree is formed by choosing the root node and is split further based on the entropy and information gain. Eqn. (5) is used to calculate the entropy where \( p_i \) indicates the ratio of the target label from the set. Eqn. (6) is used to calculate the information gain. The selection at every node is based on the input feature of the node acquired from the training set. The tree node at each level estimates the quality of the target label and estimates the predictions based on the majority class among the target label.
\[ P(x_i | y) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right) \]  
\[ \text{Entropy} = -\sum_{i=1}^{n} p_i \log_2 p_i \]  
\[ \text{Gain} (T, X) = \text{Entropy} (T) - \text{Entropy} (T, X) \]  

3.4. Defuzzification

The last step of the methodology is to transform the linguistic variables of the given input to crisp values for analysis of the node quality. Centroid method is selected to formulate the values based on the degree of the membership function for defuzzification.

Algorithm: FSS
Input: Remaining Energy(α), RSSI (β), Queue Utilization(γ), Candidate set P∈{P_1, P_2, ..., P_n}.
Output: Optimal Parent (P_{best})

Step 1:/* Fuzzification */
Transforming the crisp input into linguistic values
\[ \{\alpha_{\text{crisp}}, \beta_{\text{crisp}}, \gamma_{\text{crisp}}\} \rightarrow \{\alpha_{\text{ling}}, \beta_{\text{ling}}, \gamma_{\text{ling}}\} \]

Step 2:/* Modeling the multi metric OF */
Model an OF as \( F(\alpha_{\text{ling}}, \beta_{\text{ling}}, \gamma_{\text{ling}}) \)
\( \alpha, \beta, \gamma : \{\text{low, average, high}\}, \{\text{strong, normal, weak}\}, \{\text{light, normal, heavy}\} \)

Step 3:/* Trained Model */
Training model with the target label Quality
\( \text{Quality} = \text{optimize}[F(\alpha_{\text{ling}}, \beta_{\text{ling}}, \gamma_{\text{ling}})] \)

Step 4:/* Testing the Model */
Estimating the trained model accuracy for quality prediction using the classifiers

Step 5:/* Defuzzification */
\[ \{\alpha_{\text{ling}}, \beta_{\text{ling}}, \gamma_{\text{ling}}\} \rightarrow \{\alpha_{\text{crisp}}, \beta_{\text{crisp}}, \gamma_{\text{crisp}}\} \]

Step 6:/* Selection */
Optimal Parent (P_{best}) will be chosen according to the best accuracy of learning from the candidate parent set.
\( F(\alpha_{\text{ling}}, \beta_{\text{ling}}, \gamma_{\text{ling}}) \rightarrow F(\alpha_{\text{crisp}}, \beta_{\text{crisp}}, \gamma_{\text{crisp}}) \)

4. System Model

In our proposed work of identifying the best parent for optimal path selection using a supervised learning mechanism with a fuzzy inference system for reliable data communication in DODAG. The approach of eliminating the fuzzy rule for quality node prediction when compared with intelligent learning technique improves the time complexity. While analyzing the node quality with optimized composite OF develops the QoS parameters. The seven classes to identify the quality node is categorized based on the target label with intelligence-based learning on the modeled mechanism. Fig.4 shows the process of the proposed model.

![Figure 4. Fuzzy Intelligent System](image-url)
Processing of data using supervised learning is an important process of machine learning. Training data set is given as input with output label to train in the learning phase. The learning phase predicts the dependency and relationship between the inputs and predicts the result of the output [17].

5. Results
The supervised learning in fuzzy logic uses a dataset in the testing phase to analyze the trained model in terms of accuracy with different classifiers. The prediction of the best parent for optimal path selection in RPL with a trained dataset predicts the best quality node from the available parent nodes. Fig.5 depicts the accuracy of the trained dataset in comparison with classifiers in the methodology. The selected multi-class classifiers artificial neural network (ANN), Naïve Bayes, and decision tree is used for measuring the accuracy level of trained dataset [18]. The dataset is obtained from vehicle communication having 3000 samples. The build dataset has been trained with 2400 samples. In the testing phase the results 600 samples are measured. The train-test 80:20 split ratio. The validation accuracy with 70: 30 is almost the same. All classifiers are tested with an equal number of samples. The optimized objective function with intelligence-based learning shows better accuracy using the ANN classifier.

![Accuracy Analysis](image.png)

Figure 5. Comparison of Classifiers

6. Conclusion
The supervised learning based fuzzy logic with an optimized objective function is used to predict the optimal path in DODAG. The objective function with RSSI, QU, and RE covers both node and link metrics such that it is suitable for real-time application. The selection of quality nodes from the proposed model with intelligence-based learning improves the QoS parameters. The proposed FSS mechanism estimates the quality of the node using different classifiers such as ANN, Naïve Bayes, and Decision Tree. The proposed methodology works better with ANN in comparison with Naïve Bayes and decision tree in terms of accuracy for predicting the parent quality node. The novel mechanism shows good result in terms of reliability, energy consumption, and latency. As a future enhancement deep learning approach can be applied to study the prediction and accuracy comparison of the model.

References
[1] Lamaazi, Hanane, and Nabil Benamar. 2020, Ad Hoc Networks 96, 2001.
[2] Lamaazi, Hanane, and Nabil Benamar. 2017, 13th international wireless communications and
mobile computing conference (IWCMC). IEEE.
[3] Lamaazi, Hanane, and Nabil Benamar. 2018, Journal of Network and Computer Applications 117: 42-58.
[4] Gaddour, Olfa, Anis Koubâa, and Mohamed Abid. 2015 , Ad Hoc Networks 33: 233-256.
[5] Kamgueu, Patrick Olivier, et al. 2013, Diss. INRIA.
[6] Agarkhed, Dr Jayashree, and Vijayalaxmi Kadrolli. 2017, Proceedings in International Journal of Control Theory and Applications (IJCTA) 10.9.
[7] Nurmio, J., Nigussie, E., & Poellabauer, C. (2015, October). In 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (pp. 83-91).
[8] Yunis, Juan Pablo, and Diego Dujovne. 2014, IEEE Biennial Congress of Argentina (ARGENCON).
[9] Shakya, Nikesh Man, Mehdi Mani, and Noel Crespi. 2017 ,Global Internet of Things Summit (GloTS). IEEE.
[10] Tsiftes, N., Eriksson, J., & Dunkels, A. 2010 ,International Conference on Information Processing in Sensor Networks (pp. 406-407).
[11] Lamaazi, H., Benamar, N., Imaduddin, M. I., & Jara, A. J. 2015, In 2015 International Conference on Wireless Networks and Mobile Communications (WINCOM) (pp. 1-8).
[12] Mohamed, B., & Mohamed, F. 2015, International Journal of Distributed Sensor Networks, 11.
[13] Kamgueu, P. O., Nataf, E., & Djotio, T. N. 2015 , IEEE (pp. 789-795).
[14] Jaiswal, S., & Yadav, A. 2013, (pp. 433-438). IEEE.
[15] Bouaziz, M., Rachedi, A., Belghith, A., Berbineau, M., & Al-Ahmadi, S. 2019 ,Future Generation Computer Systems, 97, 247-258.
[16] Shreyas, J., Singh, H., Bhutani, J., Pandit, S., Srinidhi, N. N., & SM, D. K. 2019 , (pp. 141-145). IEEE.
[17] Sliwa, B., Piatkowski, N., & Wietfeld, C. (2020, June). (pp. 1-7). IEEE.
[18] Akbari, Yalda, and Shayesteh Tabatabaei. 2020 , Wireless Personal Communications ,1-17.