Penguin search with Harris-Hawk optimization algorithm to improve clustering performance in wireless network

Chitra Sabapathy Ranganathan¹, Rajeshkumar Sampathrajan²
¹Information Technology, Mphasis Corporation, Chandler, United States
²Cloud Architect, McKinsey and Company, Fort Worth, United States

ABSTRACT

Integrating optimal search algorithm concepts across the wireless core and cluster structure enables next-generation wireless networks to effectively provide reliable low-delay communications and connectivity for internet of things (IoT) devices. This article describes penguin search with the Harris-Hawk optimization algorithm (PHHO) to improve clustering performance in wireless networks. The penguin search optimization algorithm (PSO) algorithm computes the fitness value for feature selection from the database. Harris-Hawk optimization (HHO) algorithm to reduce the time and energy required for network transmission. This mechanism builds the clusters based on node communication range. The node direction, node mobility, node bandwidth availability, and energy parameters to decide the cluster head (CH) by applying the HHO algorithm. This approach uses a PSO algorithm fitness function to select the feature subset to minimize error and overhead in the network. Using a network simulator (NS)-3, this method assesses and chooses the most efficient way for data transmission, and the result is compared to a baseline mechanism.

This is an open access article under the CC BY-SA license.

1. INTRODUCTION

The internet of things (IoT) necessitates a new computing paradigm that accedes to the scalability of the wireless network [1]. Nowadays, IoT is an attractive research field to enhance the environment of smart homes and smart cities [2]. A lot of data traffic is generated due to the intelligent wireless communication system’s recent quick development and the emergence of new kinds of applications [3]. Users’ expectations for service quality and wireless resources for various applications will rise along with improvements in quality of life. The intelligent wireless communication system may provide varied services for various application situations [4]. However, wireless networks’ energy use and quality of service (QoS) performance are the main issues. We use effective clustering to tackle the issue, as mentioned earlier. Although gathering environmental data is the main objective of clustering, wireless nodes are more effective at gathering this data owing to resource limitations. Metaheuristic algorithms have local optima that may be resolved utilizing hybrid methods connected to cluster head (CH) selection [5].

Batteries are required to power the wireless nodes. However, they are usually challenging to replace. Furthermore, sustaining QoS performance depends on latency, energy use, and network coverage. The difficult energy conservation issues inside the network must thus be considered when implementing effective QoS solutions. However, more studies must be done to solve the constraints that exist in this area. Energy usage is necessary since an object’s productivity rises with age. Network lifespan is a key feature.
since it shows network performance decline [6]. The next node also perishes when a node fails. How to increase network lifespan and save node power is thus an essential issue that has to be resolved. Harris-Hawk optimization (HHO) imitates the surprise pounce style of hunting used by Harris Hawk s in the wild. Harris Hawk s, one of the most intelligent birds in nature, can mimic various pursuit methods depending on various conditions and prey evasion actions [7].

Problem statement: in a wireless network, efficient clustering and routing is a major difficulty. In this mechanism, improved artificial bee colonies (IABC)-based energy-efficient clustering-based HHO routing is used to assess and choose the most efficient way [8]. However, this mechanism does not use the feature selection algorithm; as a result, increases the overhead and delay in the network. To solve these issues, penguin search with HHO algorithm to improve clustering performance in wireless network is proposed.

Work contribution: the contribution of the penguin search with the Harris-Hawk optimization algorithm (PHHO) mechanism is specified below. The goal of the penguin search optimization algorithm (PSO) is to replicate how penguins seek to locate the optimum food sources. Starting with a population of individual penguins, the algorithm runs. Each penguin stands for a possible answer to the optimization issue. These answers are assessed using an objective function that gauges their performance. The PSO algorithm computes the fitness value for feature selection from the database and it is used to select the feature subset to minimize error and overhead.

This mechanism builds the clusters based on node communication range. The node direction, node mobility, node bandwidth availability, and energy parameters to decide the CH by applying the HHO algorithm. The HHO algorithm to reduce the time and energy required for network transmission. Simulation results demonstrate that PHHO method assesses and chooses the most efficient way for data transmission, minimizes both the delay and routing overhead.

The residuals of this article are prepared as follows. The PHHO mechanism is introduced in section 2. In section 3 explains the simulation results, and section 4 concludes the article with a summary and future research of PHHO mechanism.

A metaheuristic method called emperor penguin optimizer (EPO) imitates the emperor penguins’ huddling behavior [9]. EPO often provides superior optimization results than the other well-known metaheuristics. When it comes to intensity and diversification, EPO outperforms its rival algorithms. Nevertheless, EPO cannot provide the optimal answer for discrete issues with discrete search spaces. Two factors make up EPO’s major strength. First, EPO has a minimal learning curve based on the straightforward parallel of emperor penguins’ natural huddling behavior. Second, EPO provides a simple implementation. In the EPO, the emperor penguins stand in for the candidate solution, the huddle for the search area, which is represented by a two-dimensional L-shaped polygon plane, and the viable solution is represented by the emperor penguins placed randomly [10]. The goal is to identify a powerful mover representing the ideal solution among all emperor penguins. The EPO’s better global search capability is supported by superior diversity brought about by the huddling process. The EPO’s effective movement provides better intensification, which improves the EPO’s outstanding local search capability. It allows for a smooth transition from intensification (local search) to diversification (global search).

The EPO algorithm and M-tree-based multicast mechanism are presented for the best choice of multicast routes to extend lifespan [11]. With the estimate of many factors, route inclusion, and destination, this mechanism uses the advantages of exploitation and exploration inherited from the EPO algorithm. It emphasizes latency, shortest distance, connection stability, and energy for the best tree selection. Using network information, a PSO algorithm with a multi-agent reinforcement learning prediction algorithm can anticipate illnesses [12]. It is used to determine if a disease is present and to suggest a course of action for the patient. A hybrid EPO/particle search algorithm is used in the node redeployment approach for improved communication [13]. By correctly placing the sensor nodes, this hybridization lowers the node failure rate and network energy consumption ratio. This method improves the node redeployment strategy by computing the fitness function while guaranteeing the stability of the network architecture.

A method for feature selection based on EPO and bacterial foraging optimization. The feature selection method reduces the number of features while increasing classification precision [14]. Eight performance measures with statistical bases are computed in addition to the execution time. The maximum accuracy is attained using the hybrid optimization strategy in conjunction with random forest. The penguins’ spiral-shaped movement in the huddle and heat emission control the EPO algorithm to identify an energy-efficient routing that complies with QoS standards [15].

A mobile sink (MS) energy-efficient network is suggested utilizing the hybrid Harris Hawk and salp swarm (Hybrid HH-SS) optimization method [16]. Finding the best route for MS is an essential challenge for energy efficiency. Using an adaptive ant colony optimization (ACO) method, the MS finds the best way to link to the CHs. As a result, the suggested hybrid algorithm reduces energy usage and lengthens the network’s lifespan. With a recursive feature elimination method, stochastic gradient boosting is used for
feature selection [17]. The characteristics are grouped for classification using the adaptive HHO clustering method. The categorization uses an improved deep genetic algorithm based on clustered characteristics. This technique enhances the deep neural network’s performance by supplementing the deep neural network’s initial weights with an improved genetic algorithm that suggests better weights for the neural network.

A clustering formula is built on the HHO formula [18]. The suggested structure mimics Hawk s’ coordinated foraging behavior. Optimize the energy utilization and an optimal CH selection using the HHO algorithm. An effective data distribution incorporating a fuzzy hierarchical network model with HHO is a safe data diffusion technique. It reports assaults and monitors how node information exchange procedures behave. To deliver reliable and optimal routing, this technique combines fuzzy clustering, geographic and energy-aware data circulation, and routing capabilities [19]. The approach associates Harris eagle people with traffic samples, determines the ideal location via several algorithm iterations, and then utilizes this as the initial clustering algorithm center to direct data traffic categorization.

The ACO method enhances the routing efficiency and balances the network load. The ACO method is used to solve link failures [20]. A greedy ACO routing method selects the single best route from sender to receiver to reach the highest packet delivery and throughput [21]. The hybrid PSO is established by arranging the PSO and harmony search algorithms to solve the traveling salesman issue [22]. The adaptive neuro-fuzzy inference system uses heuristics and meta-heuristics algorithms to discover the greatest solution [23]. The emperor penguins colony is an inspired meta-heuristic algorithm with fewer errors and better performance. A multi-objective emperor penguin optimizer is employed to alter the QoS [24]. This method uses the QoS parameter, particularly on throughput, and applies cloud rank 3 with the help of the multi-objective penguin optimizer algorithm. Bio-inspired algorithms can optimize different issues in several fields, like artificial intelligence [25]. Joint particle swarm optimization and PSO algorithm concentrate on discovering the rescue targets by optimizing their movement [26]. Intelligent power control mechanism is a superior approach that tracking maximum power point [27], [28].

2. PROPOSED METHOD

This approach uses the HHO and PSO algorithms to build a hybrid technique, the PHHO approach, to enhance wireless network performance. This approach consists of two phases: the HHO algorithm selects the CH, and the feature selection uses the PSO algorithm. The node communication range parameters separate several clusters. Next, the HHO algorithm selects the CH by applying node energy, mobility, direction, and bandwidth availability. The feature selection in which selects the feature subset using the PSO algorithm. The fitness function of the PSO algorithm removes the errors.

2.1. Fuzzy-based cluster formation

Centres for the cluster and the connected cluster member nodes for every cluster centre is distinguished use fuzzy algorithm by the sender. Every node recognizes its cluster centre via data disseminated by the sender regarding the centre of cluster and the associated member nodes. Initially, the sink randomly selects a node from each connected member nodes and disseminates them as the CH. Nodes which are chose as CHs, announce themselves as CHs. Nodes choose the CH whose promoted message is obtained with the highest received signal strength indication (RSSI). Nodes then transmit the joining request to their CH. CHs synch their members of cluster for data communication.

Every node awakes up during its communication time and forwards the data to the particular CH. This is extended awaiting the end of the round. Every node independently calculates its fitness to become the CH for the next round utilizing the fuzzy system with them RE and distance to equivalent cluster centre. The node fitness value is forward to the sink via particular CH at the end of every round. The sink equates the fitness of the centre of cluster connected member nodes to decide a node with the utmost fitness value as the CH for the next round. Next, the sink disseminates the CHs for the next round. This procedure extends till all nodes of the network expire.

2.2. HHO-based CH selection

Initially, launches all solutions as the preliminary Hawk population. The population contains several clusters, the cluster denotes the Hawk s, and the CH represents the prey. The cluster members compute the fitness function. Fitness function computation based on Hawk s direction, Hawk s mobility, bandwidth availability and energy. This fitness function computation is specified (1).

\[ F = (D, BA, M, E) \] (1)

Here, the Hawk’s direction is indicated as a [0,1]. If the direction of the Hawk s and pray is present similarly, that represents 1 or 0, and the node mobility is lesser, indicating 1 or 0, the availability of Hawk s bandwidth is greater than, depicts 1 or 0. In addition, the Hawk’s present energy is greater than the average energy, which is 1; otherwise 0. A value of 1 increases the fitness function.

Phases of exploration, transformation, and exploitation make up the HHO algorithm. Throughout the exploring module, the Hawk’s will appear randomly to watch and capture the prey. In order to take advantage of the predicted presence of cluster members during the exploitation phase, the Hawk s make surprise pounces or teams of swift dives. The locations of Hawk s are seen to be potential answers. The intended posture for praying is the ideal position.

The exploration module performs the procedure of waiting, seeking, and finding prey. Transformation module changes from the exploration module to the exploitation module. The energy, mobility, bandwidth availability, and direction of prey are decomposed during escaping behavior. In the exploitation module, the Hawk s attempt the picked-out prey forecast in the previous module. In this mechanism, the HHO algorithm decides the prey is represented as a CH among cluster members. After selecting the CH, then the CH selects the features from the database using the using PSO algorithm.

2.3. Feature selection using PSO algorithm

Feature selection is selecting set of features that offer the optimal reaction categorization outcomes to improve their fineness and minimize the computational time. The feature selection method is important because of the availability of redundant and irrelevant features in the datasets that affect the effectiveness of the learning algorithm. The optimal choice is to example the feature selection issue as an NP-hard optimization and determine whether it utilizes optimization algorithms. The PSO method chooses the best ranked features by applying the information gain matrix. The PSO algorithm is established through the fish-hunting behavior of penguin groups. The penguins build themselves into definite group counts, and every group arbitrarily searches for fish until their oxygen reserves are depleted. Next, they re-establish the oxygen and search until they discover adequate fish. Next, they distribute the food positions with other groups and choose the best position for hunting. This procedure is modified for the PSO algorithm and applied for the best selection of features. This approach uses 50,000 medical features, and the PSO algorithm selects 82%, 41,000 features.

Initially, the population of penguins is initialized, and the early oxygen reserves and other parameters are set. Next, the penguins are gathered into little groups, and every group travels toward one of the accessible food positions. The mapping of the PSO algorithm to the feature selection issue should be executed sufficiently. The arbitrary population of penguin resolutions is set as features, and the groups are calculated as feature subsets. The fitness is calculated for every feature, and an optimal values determine on the best solution. The other solutions progress toward the best solutions. This measurement is conveyed as:

\[
O_j(t + 1) = O_j(t) + f(Y_{new} - Y_{old}) \times |Y_{new} - Y_{old}|
\]

here, \(O_j(t + 1)\) represents the new oxygen reserve \(O_j(t)\) represents the last purpose function formulated and established on the error rate function. Correspondingly, the capacity of eaten fish (CEF) and the group membership of penguins are also simplified. The CEF is conveyed as the food energy content.

\[
CEF_j(t + 1) = CEF_j(t) + \sum_{j=1}^{n} O_j(t + 1) = O_j(t)
\]

\[
P_j(t + 1) = \frac{CEF_j(t)}{\sum CEF_j(t)}
\]

The affiliation update of penguins is done through evaluating the probability of connecting to group i. Once the revised operations are fulfilled, the overall best features are received. The PSO procedure for feature selection is summarized as follows.

3. SIMULATION ANALYSIS

This paper uses the network simulator (NS)-3 to measure the network performance of the IABC and PHHO approaches. The wireless network distributes 100 nodes, and these nodes are moving arbitrarily.
This approach uses 802.11 medium access control (MAC) for execution [29]. This approach makes several datasets accessible, including audio, video, text, and images taken from the medical database. Totally 50,000 datasets are used, and every data class has varying file sizes. The wireless nodes transmission range is 150 metres, and nodes are formed the several clusters. The wireless nodes speed from 1 m/s to 5 m/s, correspondingly. The function of the PHHO is measured by delay, packet loss, throughput, and energy utilization of routing [30]. The amount of time it takes for a data packet to go from the sender to the recipient is referred to as delay. The delay is calculated as a function of the transmission, waiting, and propagation times. Figure 1 compares the delay according to the number of nodes.

![Figure 1. Delay of IABC and PHHO based on wireless nodes count](image)

Delay is a crucial performance indicator that demonstrates the effectiveness of the PHHO routing and data transmission strategy. Although there are more nodes in the proposed work, the delay is kept to a minimum since a PSO algorithm effectively uses feature selection. In the PHHO mechanism, the HHO algorithm chooses the best CH by considering several variables like direction, energy, bandwidth availability, and mobility. Also, the IABC mechanism does not use the feature selection algorithm; as a result, increases the delay. The rate of lost packets to the total number of packets sent across the network is known as the packet loss rate. Figure 2 compares the packet loss ratio depending on the count of wireless nodes.

![Figure 2. Packet loss of IABC and PHHO based on wireless nodes count](image)

The HHO method is used for CH selection; as a result, it minimizes the packet loss ratio. The wireless node may be dynamic, and the route may be unstable. As a result, even when the number of nodes increases, the PLR also rises. On the other hand, in the IABC mechanism, more packets are lost during data transmission. According to this figure, the PHHO mechanism minimizes the PLR since it selects the CH efficiently and avoids unwanted delay. But, the IABC mechanism raises the delay during the data transmission, hence; several packet losses in the network. The quantity of control messages delivered to build and maintain the routes using routing technologies is known as communication overhead. A communication overhead for many wireless nodes is shown in Figure 3. As there are more wireless nodes, all routing systems have higher communication overheads.

![Figure 3. Communication overhead of IABC and PHHO based on wireless nodes count](image)
Because the IABC mechanism does not use the feature selection algorithms, there is an increased communication cost since all data is sent to the CH. However, the proposed solution uses the PSO algorithm to choose important data features. Therefore, compared to the IABC mechanism, the PHHO mechanism communication overhead is lower. But, the IABC mechanism overhead is high since it does not use the feature selection algorithm; as a result; increases the redundant data, unrelated data and noisy. The volume of data transferred across the network within the allotted time period is known as throughput. Figure 4 compares throughput based on the number of wireless nodes.

As the number of nodes grows, throughput decreases. This is due to the fact that clustering requires careful feature selection when there are more nodes in the wireless network. In the PHHO mechanism, the PSO algorithm allows the CH to forward an important feature to the receiver. Additionally, the HHO algorithm chooses the best CH that improves the routing. Due to these factors, our throughput is higher than the IABC mechanism’s. The IABC mechanism increases the network delay and routing overhead, thus, minimizes the network throughput in the network.

4. CONCLUSION

The clustering in wireless networks can enhanced the throughput with network efficiency. The this article presents penguin search with HHO algorithm to improve clustering performance. This mechanism combining the HHO algorithm with the PSO to enhanced the network efficiency. The HHO algorithm picked out the CH based on Hawk’s direction, Hawk’s mobility, bandwidth availability and energy. Feature selection is selecting one or a set of features that offer the best reaction classification outcomes to improve their fineness and minimize the computational time. This mechanism utilizes the PSO algorithm to extract the redundant and irrelevant features in the datasets. Simulation results displays that the proposed method improved the network throughput and minimizes both the routing overhead and delay. However, the PHHO
mechanism lacking the load balancing. In the future, we will solve load-balancing issues by applying artificial intelligence, machine learning, and crow search optimization algorithm to improve the efficiency of the network. Furthermore, this approach using cognitive network.

REFERENCES


**BIOGRAPHIES OF AUTHORS**

**Chitra Sabapathy Ranganathan** is a (client partner, account management, IT transformation strategy, digital engineering solutions and advisory, agile delivery adoption, sales, CoE and CoP setup. Results-oriented, accomplished business technology leader with 23+ years of experience in software engineering and design. Proven track record of conceptualizing, architecting, and delivering reliable and scalable systems in a variety of areas comprising multi-technologies including Cloud, big data, AI, ML, advance analytics, blockchain, mainframe, and business intelligence. Executed complex engagements across multiple verticals, manage sales, IT delivery and operations, established vision, strategy, and journey maps that align with business priorities. Enterprise leader in digital engineering solutions and advisory, agile delivery adoption and management, pre-sales, CoE and CoP setup, IT transformation strategy, enterprise quality and digital assurance. He can be contacted at email: Chitrasarabpathyaranganathan@gmail.com.

**Rajeshkumar Sampathrajan** is a principal cloud architect at Mckinsey, where he leads a team of engineers and architects in designing and building highly scalable, resilient, and distributed systems using the latest cloud native technology in google cloud platform (GCP). He has over 17 years of experience in the IT industry, spanning various domains such as banking, retail, healthcare, and consulting. With multiple GCP certifications, as well as credentials in Azure, Snowflake, HashiCorp, Teradata, Cloudera, and ITIL. Rajesh is an expert in cloud computing, big data, machine learning, and security. He has successfully delivered solutions for complex and large-scale data analytics, data engineering, and data science projects, leveraging GCP BigQuery, Vertex AI, Dataiku, and other tools. He is passionate about helping clients transform their businesses with data-driven insights and innovative solutions. He can be contacted at email: rajasampathrajan@gmail.com.