

PAPER

Multi-Hop Distributed Clustering Algorithm Based on Link Duration

Laiwei JIANG^{†a)}, Member, Zheng CHEN^{††b)}, and Hongyu YANG^{†c)}, Nonmembers

SUMMARY As a hierarchical network framework, clustering aims to divide nodes with similar mobility characteristics into the same cluster to form a more structured hierarchical network, which can effectively solve the problem of high dynamics of the network topology caused by the high-speed movement of nodes in aeronautical ad hoc networks. Based on this goal, we propose a multi-hop distributed clustering algorithm based on link duration. The algorithm is based on the idea of multi-hop clustering, which ensures the coverage and stability of clustering. In the clustering phase, the link duration is used to accurately measure the degree of stability between nodes. At the same time, we also use the link duration threshold to filter out relatively stable links and use the gravity factor to let nodes set conditions for actively creating links based on neighbor distribution. When selecting the cluster head, we select the most stable node as the cluster head node based on the defined node stability weight. The node stability weight comprehensively considers the connectivity degree of nodes and the link duration between nodes. In order to verify the effectiveness of the proposed method, we compare them with the N-hop and K-means algorithms from four indicators: average cluster head duration, average cluster member duration, number of cluster head changes, and average number of intra-cluster link changes. Experiments show that the proposed method can effectively improve the stability of the topology.

key words: link duration, distributed, multi-hop cluster, aeronautical ad hoc network

1. Introduction

With the rapid development of the aviation industry, the demand for aeronautical communication services is increasing [1]. However, the satellite networks and air-to-ground networks of current aeronautical communication solutions have problems such as high cost, low coverage, limited capacity, and high transmission delay. Aeronautical ad-hoc network (AANET) is a self-organizing network composed of civil aviation aircraft as network nodes. It aims to enable data transmission by establishing an air-to-air (A2A) link between network nodes [2]. In this process, the nodes act as routers, and the data generated by the source node is continuously routed through these A2A to the destination node. At the same time, AANET can also transmit data to the ground through direct nodes with ground stations or satellites. AANET can enhance wireless communication applica-

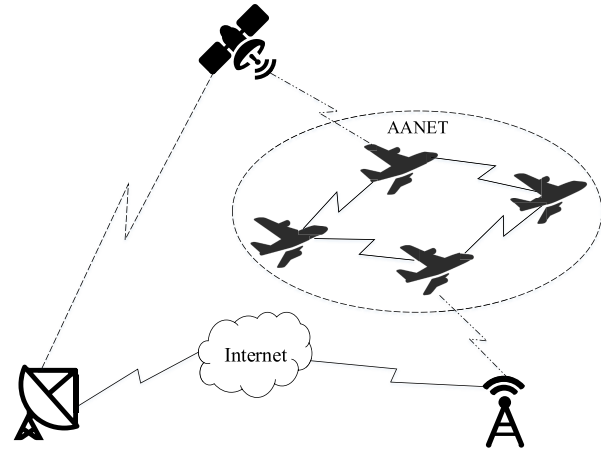


Fig. 1 Architecture diagram of aeronautical ad hoc network.

tions in aviation, such as aeronautical data transmission, air traffic control, aircraft tracking, and in-flight entertainment [3]. Compared with satellite networks and air-to-ground networks, AANET has the advantages of low cost, low latency, and high coverage. The architecture diagram of AANET is shown in Fig. 1.

However, the unstructured nature of AANET and the hyperdynamic nature of aircraft make links between nodes in AANET environments frequently disconnected, which makes it difficult to maintain the sustainability of the network topology [4]. Therefore, in this work, our task is to improve the stability of the AANET network topology. In recent years, in order to improve the stability of ad hoc networks, a multi-hop clustered network framework is proposed [5], [6]. The purpose of this network framework is to cluster network nodes, so that nodes with similar mobility characteristics are grouped into the same cluster to form a more stable network structure [7]. Multi-hop network clustering methods have been extensively studied in mobile ad hoc network [8]–[10], vehicle ad hoc network [11]–[14] and flying ad hoc network [15], [16]. Compared with single-hop clusters, multi-hop clustering are more suitable for very large AANET. The multi-hop clustering approach can support larger networks. In the single-hop clustering method, network nodes need to communicate directly with the cluster head node, so the number of nodes is limited. The multi-hop clustering method allows nodes to forward through intermediate nodes, thereby expanding the size and capacity of the network. The multi-hop clustering method has better re-

Manuscript received August 11, 2023.

Manuscript revised November 22, 2023.

Manuscript publicized January 30, 2024.

[†]The authors are with School of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China.

^{††}The author is with School of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China.

a) E-mail: lwjiang@cauc.edu.cn (Corresponding author)

b) E-mail: 2021051004@cauc.edu.cn

c) E-mail: yhyxlx@hotmail.com

DOI: 10.23919/transcom.2023EBP3137

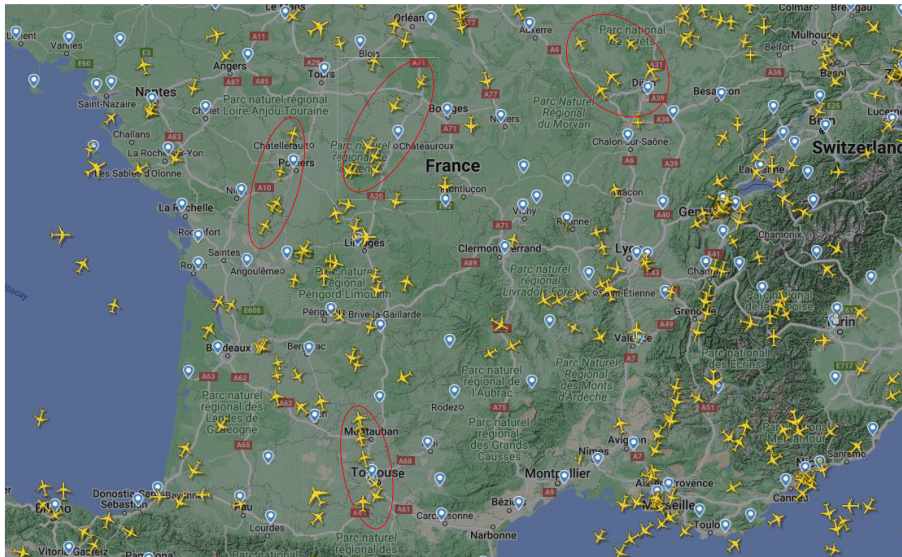


Fig. 2 Sample of aircraft distribution in French airspace.

siliency and fault tolerance. When a node in the network fails or goes offline, surrounding nodes can communicate through other paths. The presence of such redundant paths increases the stability and robustness of the network. The multi-hop clustering method can extend the coverage of the network. With multi-hop forwarding, signals can be transmitted to nodes that are far away, improving the coverage and accessibility of the network. Therefore, we hope to apply the clustering framework in AANET to improve its topological stability. Figure 2 shows a sample of aircraft distribution in French airspace. In the figure, we marked and grouped some of the aircraft. From this figure, we can see that these nodes in the same group have closer distances and similar flight directions.

Aiming at the problem of AANET clustering, a multi-dimensional clustering method is proposed in [1], [3] to cluster AANET nodes. Here, each clustering dimension is one of the attributes of the aircraft, such as node location, flight direction, flight speed, and so on. In [17], [18], the network is clustered by using the location information of nodes. They divide the network into domains by region to generate network clustering. However, they do not account for mobility between nodes, so the topological stability within the cluster is still not very good. In [19]–[21], the network is clustered by considering the relative movement speed or aggregate movement speed between nodes, aiming to effectively improve the stability of the topology in the cluster by reducing the relative movement speed between nodes in the cluster. However, these methods do not take into account the location of the node, which causes nodes at the edge of the cluster to leave the cluster frequently. [22]–[24] uses machine learning algorithms for AANET clustering, which also takes into account information such as the position, direction, heading and altitude of nodes. However, the limitation of machine learning algorithms is the need to fetch the properties of all nodes. Therefore, these methods are difficult to

apply directly to real AANET scenarios.

Based on the analysis of the above literature, we propose a multi-hop distributed clustering algorithm based on link duration. The algorithm clusters AANET in a distributed manner and uses link duration as an indicator of inter-node stability. This method aims to group nodes with similar mobility characteristics into the same cluster to obtain a more structured and stable network topology. The main contributions of the method are as follows:

- We use link duration as a measure of inter-node stability. The link duration can fully consider the location, moving speed, and direction of the node and can accurately express the stability between nodes.
- During the clustering process, we use a threshold to determine whether the link between nodes is stable. At the same time, we also set a specific gravity factor k , so that nodes can automatically set the conditions for actively establishing links according to the distribution of neighbors.
- In order to select the most stable node as the cluster head node, in the cluster head selection process, we define the node stability weight, which takes into account the connectivity of the nodes and the link duration between the nodes. This makes the cluster head selected as the stable node in the cluster.
- We verify the effectiveness of the proposed method by conducting simulation experiments on real aviation historical flight data.

The rest of this article is organized as follows: the second section describes our system model. The third section describes the proposed algorithm in detail. The experimental results are presented in the fourth section. The fifth section summarizes the papers.

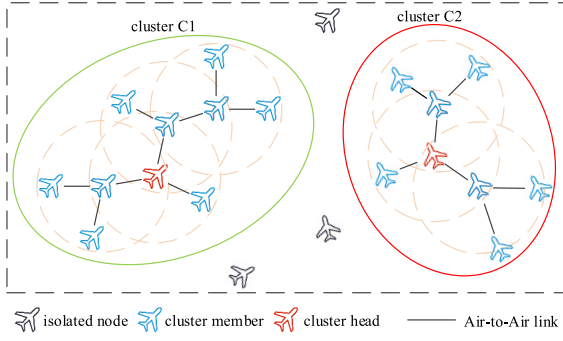


Fig. 3 Example diagram of high-dynamic aeronautical ad hoc network clustering scenario.

2. System Model

In this paper, our goal is to solve the problem of network topology instability caused by the high-speed movement of nodes in AANET through network clustering strategy. Therefore, we only consider the aviation layer consisting of aircraft. These aircrafts fly through the air at a specified route and speed. At the same time, we also assume that all nodes have the same maximum communication radius and can transmit data to all aircraft within their maximum communication radius. The considered aeronautical ad hoc network clustering model is shown in Fig. 3. All nodes in the diagram are divided into two clusters, and the movement properties of the nodes in each cluster are similar.

We use $N^{[t]}$ to represent the set of all nodes at time t and i to represent the index of the node. At time t , we use $S_i^{[t]} = (\theta_i^{[t]}, \phi_i^{[t]}, h_i^{[t]})$ to represent the position coordinates of aircraft i in the WGS-84 coordinate system, where $\theta_i^{[t]}$, $\phi_i^{[t]}$, $h_i^{[t]}$ represent the latitude, longitude, and altitude of node i at time slice t , respectively. Using Eq. (1), we can convert the position coordinates $S_i^{[t]} = (\theta_i^{[t]}, \phi_i^{[t]}, h_i^{[t]})$ to the coordinates $L_i^{[t]}$ in the Cartesian coordinate system [25].

$$\begin{aligned} L_i^{[t]} = & [(R_e + h_i^{[t]})\cos\phi_i^{[t]}\cos\theta_i^{[t]}, \\ & (R_e + h_i^{[t]})\sin\phi_i^{[t]}\cos\theta_i^{[t]}, \\ & (R_e + h_i^{[t]})\sin\phi_i^{[t]}] \end{aligned} \quad (1)$$

where $R_e = 6371$ km, expressed as the radius of the Earth. Therefore, we can express the distance between two nodes as $d_{ij} = \|L_i^{[t]} - L_j^{[t]}\|$.

2.1 Mobility Model

In AANET, aircraft generally fly on a pre-set route and speed. In a short time, the aircraft moves in a relatively linear path without changing the direction of movement and the parameters of movement. Therefore, the trajectory of the aircraft can be considered pseudolinear [1]. Therefore, the flight trajectory of an aircraft can be regarded as a process consisting of several straight lines moving at a uniform speed. The

position of the node can be determined by the initial position and movement speed of each time period, which can be represented by Eq. (2).

$$L_i^{[t]} = \begin{cases} L_i^{[t_s]} + V_i^{[t_s]}(t - t_s), & t_s < t < t_1 \\ L_i^{[t_1]} + V_i^{[t_1]}(t - t_1), & t_1 < t < t_2 \\ \dots \\ L_i^{[t_k]} + V_i^{[t_k]}(t - t_k), & t_k < t < t_e \end{cases} \quad (2)$$

where the vector $L_i^{[t]}$ represents the position of node i at time t . The vector $V_i^{[t]}$ represents the velocity of node i at time t . t_s and t_e represent the start time and end time of the aircraft node flight, respectively. $t_n (n = 1, 2, 3 \dots)$ represents different time slice indices.

2.2 Communication Radius

VHF has characteristics of long-distance omnidirectional communication, and it is also the main mode of communication between civil aviation aircraft. So in this article, we use VHF to calculate the maximum communication radius of nodes. VHF operates at a frequency of 119–137 MHz and adopts line-of-sight communication. However, line-of-sight communication is the ideal transmission method. In reality, the maximum communication radius available is much smaller than the ideal communication radius due to factors such as channel fading. Aviation VHF channels generally use Rayleigh fading channels [26], and their link bit error rate can be expressed by Eq. (3).

$$BER_L = \frac{1}{2} \left(1 - \frac{\sqrt{\frac{2\sigma_f^2 p_t G_t G_r c^2}{(4\pi f_c R)^2}}}{\sqrt{P_{therm} + \frac{2\sigma_f^2 p_t G_t G_r c^2}{(4\pi f_c R)^2}}} \right) \quad (3)$$

where R is the available communication radius; p_t is the transmit power; G_t and G_r represent the gain of the transmitting antenna and the receiving antenna, respectively; f_c represents the carrier frequency; c represents the speed of light; σ_f^2 represents the mean square value of the signal envelope in the Rayleigh distribution; P_{therm} represents thermal noise power, expressed in Eq. (4).

$$P_{therm} = FkT_0R_b \quad (4)$$

where F represents the noise value; $k = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant; T_0 represents the normal temperature value; R_b stands for data transfer rate. According to the actual situation of aeronautical communication, the values of each parameter are shown in Table 1. From Eq. (3) and Eq. (4), the communication radius of the aircraft node can be calculated to be about 250 km.

Link establishment between nodes depends on the distance between nodes and the maximum communication distance of nodes. If the distance between the two nodes is less than the communication radius of the node, a link can be created between the nodes. Therefore, when $d_{ij} < R$, it means

Table 1 The values of the parameters of the physical layer.

parameter	value	parameter	value
P_t/w	50	f_c/Mhz	131.45
G_t/dbi	1	G_r/dbi	1
F/dB	6	T_0/K	300
σ_f^2	1	$R_b/(\text{kb/s})$	31.5

that a communication link can be created between node i and node j . At the same time, we use $B(i) = \{j | d_{ij} < R\}$ to represent the set of all neighboring nodes within the communication range of node i .

3. The Proposed Clustering Method

In the multi-hop distributed clustering algorithm based on link duration, we use link duration as a measure of stability between nodes. Therefore, in this section, we first give the calculation method of link duration; Secondly, the conditions for nodes to actively create links are given. Then, the strategy of cluster head selection is elaborated. Finally, we describe the algorithm flow in detail.

3.1 Link Duration

In AANET, frequent broken links between nodes are the root cause of high dynamic changes in network topology. Link duration refers to the length of time that communication links are maintained between nodes, which is an important evaluation index to measure node stability. The higher the link duration, the higher the stability between nodes. We use $S_{time}(i, j)$ to represent the link duration between two nodes. The calculation method of $S_{time}(i, j)$ is shown in Eq. (5).

$$S_{time}(i, j) = \frac{-\vec{D}_{ij} \vec{V}_{ij} + \sqrt{(\vec{D}_{ij} \vec{V}_{ij})^2 + V_{ij}^2 (R^2 - D_{ij}^2)}}{V_{ij}^2} \quad (5)$$

where R represents the maximum communication radius of the node, vector D_{ij} represents the relative position between node i and node j , and vector V_{ij} represents the relative moving speed between node i and node j . In this paper, we try to form stable network clustering by creating stable links between nodes. Therefore, we set a link duration threshold T_{Stime} , and we consider links with a link duration greater than T_{Stime} as stable.

3.2 Cluster Topology Formation

In AANET, aircraft nodes are distributed along the routes. This will result in a high density of nodes on or near route crossings, while areas with no route distribution or fewer routes will have a small density. This results in very different regional node densities in AANET.

Therefore, we set a gravity factor k to make the node determine whether it meets the condition of actively creating links with neighbors based on the distribution of neighbors.

When the ratio of all neighbors of a node to the number of stable links is greater than k , the node satisfies the condition for actively creating links. At this point, the node creates links with nodes that can create stable links to form network clustering.

In this method, we do not limit the size of the cluster by setting the maximum hop from the cluster member to the cluster head. Our goal is to assign relatively stable nodes to the same cluster as much as possible. If we set the maximum hop number n , nodes with a distance greater than n from the cluster head cannot join the cluster, resulting in a large number of orphaned nodes. As the number of hops increases and the cluster size increases, the time to configure the cluster also increases. However, the increase in configuration time is negligible compared to the duration between nodes. According to the above, the maximum communication radius between nodes is 250 km, so it can be calculated that the time for data exchange between two nodes is generally a few milliseconds. Although there is no limit on the maximum number of hops, AANET is not infinite, so n is a constant less than 10 in general. Therefore, the time of a clustering configuration is generally tens of milliseconds. In our method, we set T_{Stime} to establish a stable link between nodes, and T_{Stime} is usually a few minutes, which is thousands of times the clustering configuration time. Therefore, in this article, we will ignore the timing of clustering configuration.

3.3 Cluster Head Selection

After the clustering topology is formed, the entire network can be divided into subnets. To obtain the final cluster structure, a cluster header (CH) node should be chosen. CH is used to manage data transmission between other cluster member nodes (CM) within a cluster. Choosing stable CH is conducive to improving the success rate of data transmission in the network. In a network, the greater the link duration of the link created by a node, the more stable the node is. The more links are established, the more centrality of the node is higher. Therefore, we should choose a node with more stable links as the cluster head node. In this article, we define node stability weights to measure node stability. Its definition is shown in Eq. (6) where $f(x)$ is a coefficient function and defined as in Eq. (7). $B(i)$ is represented as a collection of neighbors for node i .

$$W_i = \sum_{j \in B(i)} f(S_{time}(i, j)) * S_{time}(i, j) \quad (6)$$

$$f(x) = \begin{cases} 1, & x \geq T_{Stime} \\ 0, & \text{other} \end{cases} \quad (7)$$

$$CH_C = \arg \max_{i \in C} W_i \quad (8)$$

According to the defined node stability weight, the most stable node is selected as CH by Eq. (8) in the cluster, where CH_C is represented as CH of cluster C .

Table 2 INFO_TABLE table entry information.

ID	location	velocity	heading	CHID	$S_{time}(i, j)$	IS_connected
----	----------	----------	---------	------	------------------	--------------

Algorithm 1 The pseudocode of the proposed multi-hop clustering algorithm.

Input: $T_{Stime}, k, N^{[t]}$;
Output: clustered network topology;

```

1: for  $i$  in  $N^{[t]}$  do
2:   Broadcast and receive HELLO;
3:   Parse HELLO and modify INFO_TABLE
4:   Use Eq.(5) to calculate  $S_{time}(i, j)$ 
5:   Mark links whose link duration is greater than  $T_{Stime}(U_{SL})$ 
6:   if  $k * |U_{SL}| \geq |\text{INFO\_TABLE}|$  then
7:     for  $i$  in  $U_{SL}$  do
8:       Send JOIN_REQ to build link
9:     end for
10:    Use Eq.(6) to calculate  $W_i$ 
11:    Broadcast  $W_i$  through established links
12:    if  $W_i$  is maximal then
13:      Node  $i$  is CH
14:      Broadcast CH_BEACON
15:    else
16:      if receive CH_BEACON then
17:        Set the CH information according to CH_BEACON
18:      end if
19:    end if
20:  end if
21:  if  $k * |U_{SL}| < |\text{INFO\_TABLE}|$  then
22:    if receive JOIN_REQ then
23:      Reply JOIN_RESP and build link
24:    else
25:      Establish a link with the node with the largest link duration to
        join the cluster
26:    end if
27:  end if
28:  if node  $i$  leaves its cluster then
29:    Establish a link with the node with the largest link duration to
        join a new cluster
30:  end if
31: end for

```

3.4 Algorithm Detailed Process

In this section, we divide the entire clustering algorithm into three stages: the information table generation stage, the cluster formation stage, and the cluster maintenance stage. The detailed procedure of the clustering algorithm is shown in Algorithm 1. In the following, we will elaborate on these three stages.

3.4.1 Information Table Generation

In AANET, each node maintains a INFO_TABLE neighbor table. Each node maintains neighbor information within its communication range in the INFO_TABLE. The information table contains the ID of the neighbor nodes, location of the neighbor nodes, moving speed of the neighbor nodes, heading of the neighbor nodes, the ID of the cluster head of the neighbor nodes (CHID), link duration to neighbors, and whether to create a link with it (IS_connected). The

entries in the table are shown in Table 2. The link duration is calculated from Eq. (5) after obtaining information about the neighbor's location, speed, and heading. Each node is aware of the presence of neighbor nodes by broadcasting and receiving HELLO packets. When a node receives a HELLO packet from a neighbor node, it modifies the INFO_TABLE table entry. If no HELLO packets are received from the neighbor node within the specified time interval, the table entry is dropped.

3.4.2 Cluster Formation

During the information table generation phase, each node sends a HELLO beacon packet to its single-hop neighbor node, which updates its own INFO_TABLE by parsing the received HELLO packet. Each node calculates the link duration according to Eq. (5) and modifies the corresponding table entry for the INFO_TABLE. In the INFO_TABLE, the link duration is compared with T_{Stime} , and the stable links are marked to form the stable link collection U_{SL} . Here, we set a specific gravity factor k . When the state of the node satisfies $k * |U_{SL}| \geq |\text{INFO_TABLE}|$, the node actively sends a JOIN_REQ information request to the neighbor node to establish a link with it. After the node receives the JOIN_REQ, it replies with JOIN_RESP packet to create the link. After the link is established, all nodes are divided into several clusters. At this point, the nodes calculate their respective stability weights according to Eq. (6) and broadcast in the cluster through the established link. When each node receives the stability weight of other nodes in the cluster, it compares its own stability weight with them, and when its own stability weight is the largest, it makes the node the cluster head node. Once identified as a cluster head node, the cluster head node broadcasts CH_BEACON packets in the cluster to notify all cluster member nodes. After the cluster member node receives the broadcast, the cluster head node information is set. This procedure corresponds to lines 6–20 in Algorithm 1. If a node fails to establish links with other nodes and there are neighbor nodes, the node selects from the INFO_TABLE to establish a connection with the node with the largest link duration to join the cluster. This procedure corresponds to lines 21–26 of Algorithm 1. At this point, the network topology of the clustered architecture is formed.

3.4.3 Cluster Maintenance

In the process of cluster formation, the cluster has higher stability by establishing a stable link between nodes in the cluster. However, during node movement, nodes at the edge of the cluster lose their connection to the cluster due to node movement. Therefore, at the beginning of each time slice, each node needs to obtain neighbor information at

the beginning of each time slice and detect whether it is connected to other members in the cluster. If this connection is lost, that is, the node leaves the cluster, the node will reselect the most stable node from all one-hop neighbors based on the INFO_TABLE information to create a link and rejoin the cluster. The process is as 28–30 lines in Algorithm 1. After a number of time slices, the clustering structure will change dramatically, which means that we need to perform the network clustering process again. We call this process here the reclustering process, and this time period is called the reclustering interval. At the beginning of the first time slice of each reclustering interval, the reclustering process is initiated, i.e. Algorithm 1 is executed again.

4. Performance Simulation

In this section, in order to verify the effectiveness of the proposed method, we compare the proposed multi-hop clustering algorithm based on link duration with the N-hop [20] and K-means [22] algorithms, respectively. At the same time, we obtained real historical aviation flight data for simulation experiments. The data obtained are the flight trajectory of all aircraft over Europe on June 22, 2022. We divided the data into 24 groups in hours and extracted flight trajectories for stable flight in each time period. All the extracted nodes fly at extremely high speeds, usually in the range of 700–900 km/h. Therefore, throughout the simulation, the nodes change their speed and position in the moving model of Eq. (2), and the network topology is constantly changing. Since the number of aircraft nodes in each time period is different, we divide the all-day data into five groups according to the number of nodes in each time period: 150–250, 251–350, 351–450, 451–550, and 551–700. The simulation time for each set of data is 60 minutes. We set the size of time slice to 1 min. At the beginning of each time slice, the speed and direction of the node's movement are modified. In each time slice, the node flies at a uniform speed at the specified speed. We assume that each node has the same maximum communication radius. Since this paper only considers the stability of the network topology, in the experiment, we assume that when the distance between two nodes is less than the maximum communication radius, it means that they can communicate directly and effectively. As described above, the maximum radius of the node is set to 250 km. Due to the high-speed movement of nodes, links between nodes may be frequently disconnected or new links may be created. Therefore, we set T_{Stime} to create links only between stable nodes. Here we set T_{Stime} to 10 min, that is, when the link duration between nodes is greater than 10 min, the two nodes will be considered relatively stable. Since the link duration between nodes is calculated based on the motion state of the current node, and the speed and direction of the node are not constant, the network topology may change greatly after a period of time, that is, the network cluster needs to be re-clustered. We make this time interval the reclustering interval and set it to 5 min.

The average cluster head duration, average cluster mem-

ber duration, number of cluster head changes, and average number of intra-cluster link changes can effectively evaluate the stability of clusters and the reliability of links. By comparing these four indicators, we can obtain the following experimental results based on the simulated aviation historical flight data obtained.

4.1 Average Cluster Head Duration

The cluster head duration represents the length of time during which a node was selected as the cluster head, that is, the time interval between when a node becomes a CH state and when it changes to a non-CH state. Average cluster head duration is the ratio of total cluster head duration to the number of cluster heads. The calculation method is as shown in Eq. (9).

$$AvgStime_{CH} = \frac{\sum_{i=1}^n Stime_{CH}(i)}{n} \quad (9)$$

Among them, $AvgStime_{CH}$ represents the average cluster head duration; $Stime_{CH}(i)$ represents the cluster head i duration; n represents the number of nodes that became cluster heads throughout the simulation period. According to the simulation, the experimental results are shown in Fig. 4.

According to the experimental result graph, the average cluster head duration of the proposed method is similar to the average cluster head duration of K-means and decreases with the density of nodes. When the number of nodes is 150–350, the average cluster head node duration of N-hop is similar to that of the proposed method and K-means, while when the number of nodes is greater than 350, the average cluster head duration of N-hop shows an upward trend. As the density of nodes increases, the distribution of neighbors across all nodes is similar. In the proposed method, because the distribution of nodes is similar, the node stability weights of the nodes in the center are similar. When the position of a node changes due to movement, the stability weight of the node changes, so that the past cluster head

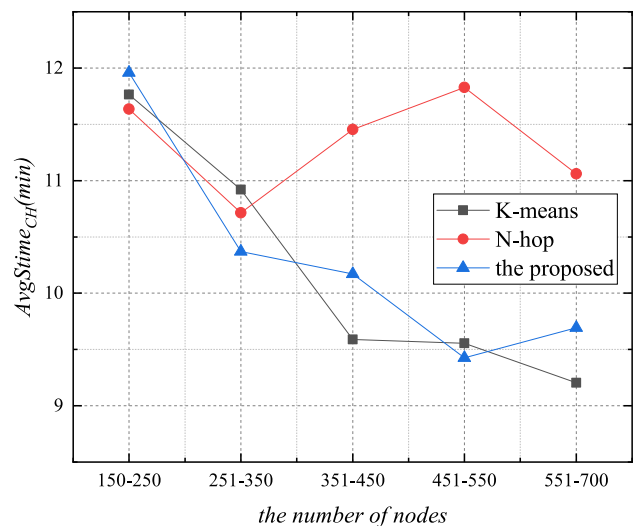


Fig. 4 Average cluster head duration simulation results.

node cannot become a cluster head node again. K-means mainly clusters the network by the position and speed of the nodes. Therefore, in the case of node position change due to node movement, K-means is difficult to select the past cluster head node as the new cluster head again. In N-hop, the cluster head is mainly determined by the relative speed of aggregation, so the determination of the cluster head is not affected by the node position. When the density of nodes increases, the relative speed of aggregation of different nodes will be more different. And because the speed of nodes in AANET is relatively stable, the cluster head is more stable. However, due to increased density, the neighbors of the cluster head may have similar weights to the cluster head, which can cause the cluster head to switch back and forth between some nodes. In terms of average cluster head duration, the proposed method is inferior to the other two methods. Overall, the average cluster head duration was reduced by 8.8% compared to N-hop.

4.2 Average Cluster Member Duration

Cluster member duration, defined as the time interval from when a node joins a cluster and when it leaves the cluster, is calculated similarly to the average cluster head duration. The average cluster member duration is the ratio of the combined duration of all nodes as a cluster member to the number of nodes, as shown in Eq. (10), where $AvgStime_{CM}$ represents the average cluster member duration, and $Stime_{CM}(i)$ represents the duration of node i as a cluster member, and m is the number of nodes.

$$AvgStime_{CM} = \frac{\sum_{i=1}^m Stime_{CM}(i)}{m} \quad (10)$$

From Fig. 5., we can see that the average cluster member duration of the proposed method and K-means is significantly higher than the average cluster member duration of N-hop, and the average cluster member duration increases with the increase of network nodes. When the number of nodes is greater than 350, the average cluster member duration is less affected by the density of nodes. This is mainly because when the node density is higher, the probability of neighbors appearing within the communication range of all nodes is also greater. Therefore, nodes are more likely to join the cluster by multi-hop. Since after K-means determines the cluster head, all nodes in the cluster head range will be directly connected to the cluster head and joined to the cluster, resulting in nodes at the cluster head communication edge directly connected to the cluster head. The node is disconnected from the cluster head due to movement. When the proposed method establishes a connection, the node selects the most stable link to join the cluster, so the average cluster member duration of the proposed method is slightly higher than that of K-means. The average cluster member duration of N-hop is much smaller than the other two methods. This is mainly because N-hop requires each node to obtain movement information for all nodes within its n-hop range. Then, the smallest aggregate mobile node is taken as the cluster

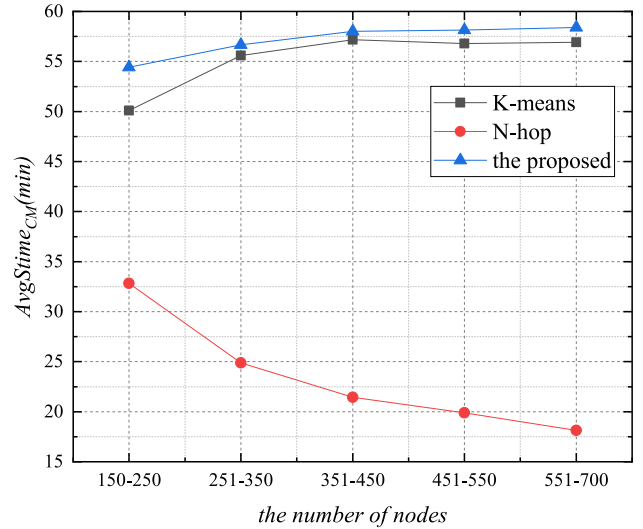


Fig. 5 Average cluster member duration simulation results.

head. This leads a situation that a node is not the minimum value within its n-hop, and all nodes within its n-hop range are not satisfied to be cluster heads. However, nodes that satisfy the cluster header will only add nodes within their n-hop range to the cluster, thus causing a large number of nodes to fail to become cluster member nodes. Therefore, the resulting cluster member duration is much smaller than the other two methods. Compared with K-means, the average cluster member duration of the proposed method is slightly improved, which is increased by 3.4%.

A comparison based on Fig. 4 and Fig. 5 raises the question that the average cluster head duration decreases as the number of network nodes increases, whereas the average cluster member duration is the opposite. The cluster head node is a small number of network nodes selected from a large number of network nodes according to each of these three algorithms. During each reclustering process, the cluster head node is reselected. This means that the cluster header may be different for each recluster interval. Other nodes are cluster members, as long as they can connect to the cluster head node through one or more hops. The higher the number of nodes in a scenario, the greater the probability that nodes will connect to the cluster head through neighbor nodes. The greater the node density, the more similar the weight of the cluster head and its neighbors, and the smaller the probability that the node that was originally the cluster head will become the cluster head again when the cluster head is reselected. Based on the definitions of average cluster head duration and average cluster member time, we will be able to understand why the average cluster head duration becomes lower and the average cluster member time becomes high.

4.3 Number of Cluster Head Changes

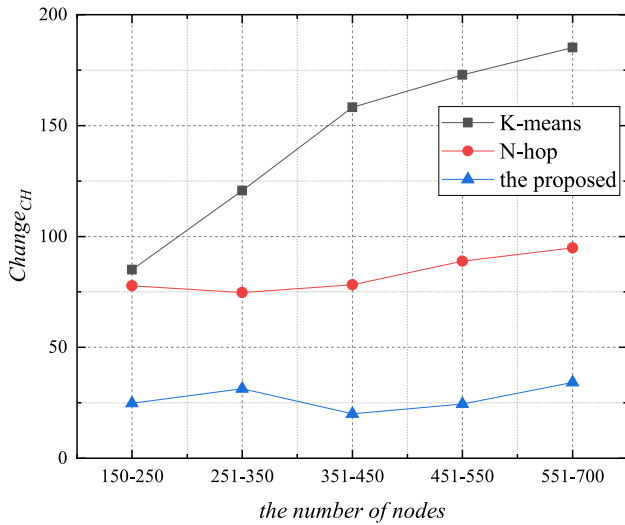
The number of cluster head changes refers to the number of times the cluster head node changes to other states during the simulation. It can be represented by Eq. (11),

Table 3 The average number of cluster heads per time slice.

the number of nodes	150-250	251-350	351-450	451-550	551-700
K-means	13.6	17.0	20.3	22.0	23.8
N-hop	12.7	8.4	13.2	15.3	16.0
the proposed	9.1	8.4	6.6	8.1	9.0

Table 4 Cluster head rate of change.

the number of nodes	150-250	251-350	351-450	451-550	551-700
K-means	0.521	0.591	0.648	0.655	0.647
N-hop	0.507	0.744	0.492	0.481	0.495
the proposed	0.227	0.312	0.252	0.251	0.317

**Fig. 6** Number of cluster head changes simulation results.

where $Change_{CH}$ represents the number of cluster head changes; $Change_{CH}(i)$ represents the number of cluster heads changed in time slice i ; and T represents the total number of time slices simulated.

$$Change_{CH} = \sum_{i=0}^T Change_{CH}(i) \quad (11)$$

Figure 6 shows the number of cluster head changes for three different methods. We also print the average number of cluster heads per time slice for the three methods, as shown in Table 3. In K-means, because there are more nodes in the scene, a larger K value needs to be set for the algorithm. As can be seen from the previous analysis, the position of the node changes because of the movement of the node. Therefore, during the next cluster formation, it is difficult for the cluster head to become a cluster head again, so the number of cluster head changes is large. In N-hop, nodes are selected by using the aggregate relative speed of all neighbor nodes within an n-hop range as the cluster head selection index. Because the movement characteristics of AANET nodes are stable and relatively similar, it will cause the cluster head to switch back and forth in certain nodes, resulting in a large number of cluster head changes. In the proposed method, we do not limit the maximum number of hops from a cluster member to a cluster head node, so the

proposed method does not generate a large number of cluster heads, which can effectively scale the scale and capacity of the network to increase the stability of the network. In addition, in the process of selecting the cluster head, we select the maximum node stability weight, so the cluster head is relatively more stable. According to Table 3, we can see that although our method does not set the maximum number of hops, the range of our cluster is not very large. The average number of members in each cluster of the proposed method is about twice that of N-hop, so the configuration cluster time of the proposed method will only be slightly higher than that of N-hop. In order to clearly illustrate that the cluster heads selected by the proposed method are more stable, we calculate the ratio of the number of cluster head changes to the number of cluster heads, as shown in Table 4. From the table, we can see that the cluster head change ratio of the proposed method is significantly lower than that of the other two methods. The experimental results show that compared with N-hop, the cluster head change rate of the proposed method is reduced by 49.1%.

4.4 Average Number of Intra-Cluster Link Changes

The high dynamics of the network topology are mainly caused by the frequent establishment and disconnection of links between nodes. The purpose of using clustering architecture in AANET is to improve the topological stability of the network, that is, to reduce the number of link changes in the cluster. Therefore, the average number of intra-cluster link changes is an important indicator to measure the performance of clustering. The average number of intra-cluster link breaks is the ratio of the number of changes of links in all clusters to the number of nodes during the entire simulation process, as defined in Eq. (12), where $AvgLink_{change}$ represents the average number of intra-cluster link breaks, $Link_{change}(i)$ is the number of times the node i link breaks during the simulation and m is the number of nodes.

$$AvgLink_{change} = \frac{\sum_{i=0}^m Link_{change}(i)}{m} \quad (12)$$

As shown in Fig. 7, when the density of nodes is low, the average number of intra-cluster link changes of the proposed method is significantly better than that of the other two methods. This is mainly because all nodes adjust the conditions for actively creating links based on the sum of

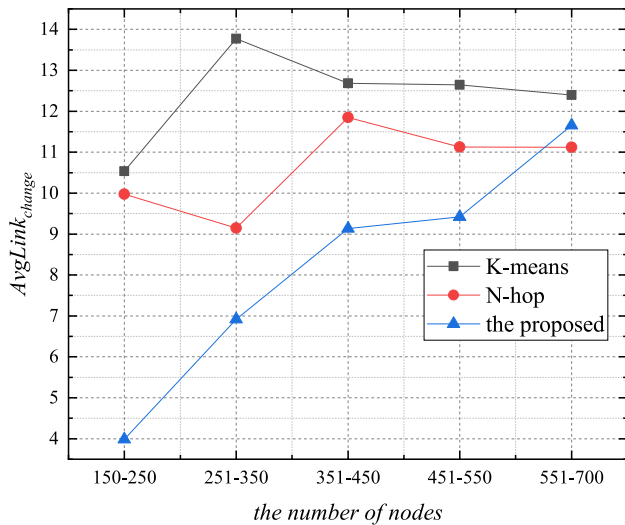


Fig. 7 Average number of intra-cluster link changes simulation results.

the number of neighbors during the clustering phase. In addition, T_{Stime} is used to limit the establishment of unstable links. Therefore, the average number of intra-cluster link breaks in the proposed method is lower. When the node density is too high, more links are created within the cluster. When a node location and speed changes, it causes a large number of link breakdowns. When the node density is large, the average intra-cluster link change of N-hop is the least. The main reason for this is that N-hop has a large number of orphaned nodes that do not establish links. As a result, the average number of intra-cluster link changes is lower. For K-means, when the node density is greater, the proportion of nodes at the cluster head communication edge decreases. Therefore, the number of link changes in the cluster after averaging will be reduced. According to the experimental results, the average number of intra-cluster link changes of the proposed method is reduced by 19.0% compared with N-hop.

5. Conclusion

In this paper, aiming at the problem of high dynamic change of AANET topology, we propose a multi-hop distributed clustering algorithm based on link duration. In this algorithm, we use link duration as a measure of inter-node stability. In addition, we use a gravity factor k to adjust the conditions for actively creating links based on the distribution of node neighbors. Simulation experiments are carried out based on real historical aviation flight data, and we compare the proposed method with other methods on each performance index. In addition to the average cluster head duration, the proposed method improved the performance indicators by 3.4%, 49.1%, and 19%, respectively. In future work, we will study node routing algorithms on the basis of clustering to improve the performance of the network.

Acknowledgments

This work is supported by Tianjin Education Commission Research Project (2020KJ025) and Key Program of National Natural Science Foundation of China (U2033210).

References

- [1] T. Bilen, H. Ahmadi, B. Canberk, and T.Q. Duong, "Aeronautical networks for in-flight connectivity: A tutorial of the state-of-the-art and survey of research challenges," *IEEE Access*, vol.10, pp.20053–20079, 2022, doi: 10.1109/ACCESS.2022.3151658.
- [2] E. Sakhaee, A. Jamalipour, and N. Kato, "Aeronautical ad hoc networks," *Proc. IEEE Wireless Commun. Netw. Las Vegas, NV, USA*, pp.246–25, 2006.
- [3] J. Zhang, T. Chen, S. Zhong, J. Wang, W. Zhang, X. Zuo, R.G. Maunder, and L. Hanzo, "Aeronautical ad-hoc networking for the internet-above-the-clouds," *Proc. IEEE*, vol.107, no.5, pp.868–911, May 2019, doi: 10.1109/JPROC.2019.2909694.
- [4] D. Wang, Y. Wang, S. Dong, G. Huang, J. Liu, and W. Gao, "Exploiting dual connectivity for handover management in heterogeneous aeronautical network," *IEEE Access*, vol.7, pp.62938–62949, 2019, doi: 10.1109/ACCESS.2019.2916920.
- [5] M. Mukhtaruzzaman and M. Atiquzzaman, "Clustering in vehicular ad hoc network: Algorithms and challenges," *Computers & Electrical Engineering*, vol.88, p.106851, Dec. 2020, doi: 10.1016/j.compeleceng.2020.106851.
- [6] S. Ghosh and A. Nayak, "Multi-dimensional clustering and network monitoring system for aeronautical ad hoc networks," *Proc. ICUFN, Sapporo, Japan*, pp.772–777, 2015.
- [7] T. Bilen and B. Canberk, "Handover-aware content replication for mobile-CDN," *IEEE Netw. Lett.*, vol.1, no.1, pp.10–13, March 2019, doi: 10.1109/LNET.2018.2873982.
- [8] D. Zhang, K. Zheng, D. Zhao, X. Song, and X. Wang, "Novel quick start (QS) method for optimization of TCP," *Wireless Netw.*, vol.22, no.1, pp.211–222, Jan. 2016, doi: 10.1007/s11276-015-0968-2.
- [9] D. Zhang, G. Li, K. Zheng, X. Ming, and Z.-H. Pan, "An energy-balanced routing method based on forward-aware factor for wireless sensor networks," *IEEE Trans. Ind. Informat.*, vol.10, no.1, pp.766–773, Feb. 2014, doi: 10.1109/TII.2013.2250910.
- [10] M. Chatterjee, S.K. Das, and D. Turgut, "WCA: A weighted clustering algorithm for mobile ad hoc networks," *Cluster Comput.*, vol.5, no.2, pp.193–204, 2002, doi: 10.1023/A:1013941929408.
- [11] A. Salim, A.M. Khedr, and W. Osamy, "IoVSSA: Efficient mobility-aware clustering algorithm in internet of vehicles using sparrow search algorithm," *IEEE Sensors J.*, vol.23, no.4, pp.4239–4255, Feb. 2023, doi: 10.1109/JSEN.2022.3233903.
- [12] A. Ali, F. Aadil, M.F. Khan, M. Maqsood, and S. Lim, "Harris hawks optimization-based clustering algorithm for vehicular ad-hoc networks," *IEEE Trans. Intell. Transp. Syst.*, vol.24, no.6, pp.5822–5841, June 2023, doi: 10.1109/TITS.2023.3257484.
- [13] D. Zhang, H. Ge, T. Zhang, Y.-Y. Cui, X. Liu, and G. Mao, "New multi-hop clustering algorithm for vehicular ad hoc networks," *IEEE Trans. Intell. Transp. Syst.*, vol.20, no.4, pp.1517–1530, April 2019, doi: 10.1109/TITS.2018.2853165.
- [14] A. Akbar, M. Ibrar, M.A. Jan, L. Wang, N. Shah, and H.H. Song, "SeAC: SDN-enabled adaptive clustering technique for social-aware internet of vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol.24, no.5, pp.4827–4835, May 2023, doi: 10.1109/TITS.2023.3237321.
- [15] J. Guo, H. Gao, Z. Liu, F. Huang, J. Zhang, X. Li, and J. Ma, "ICRA: An intelligent clustering routing approach for UAV ad hoc networks," *IEEE Trans. Intell. Transp. Syst.*, vol.24, no.2, pp.2447–2460, Feb. 2023, doi: 10.1109/TITS.2022.3145857.
- [16] G. Sun, D. Qin, T. Lan, and L. Ma, "Research on clustering

routing protocol based on improved PSO in FANET,” *IEEE Sensors J.*, vol.21, no.23, pp.27168–27185, Dec. 2021, doi: 10.1109/JSEN.2021.3117496.

- [17] F. Shi, Y. Shi, and L. Lai, “A clustering algorithm of Ad-Hoc network based on honeycomb division,” 2011 IEEE International Conference on Granular Computing, Kaohsiung, Taiwan, 2011, pp.863–866, doi: 10.1109/GRC.2011.6122549.
- [18] J. Yang, K. Sun, H. He, X. Jiang, and S. Chen, “Dynamic virtual topology aided networking and routing for aeronautical ad-hoc networks,” *IEEE Trans. Commun.*, vol.70, no.7, pp.4702–4716, July 2022, doi: 10.1109/TCOMM.2022.3177599.
- [19] E. Sakhiae and A. Jamalipour, “Stable clustering and communications in pseudolinear highly mobile ad hoc networks,” *IEEE Trans. Veh. Technol.*, vol.57, no.6, pp.3769–3777, Nov. 2008, doi: 10.1109/TVT.2008.919606.
- [20] Z. Zhang, A. Boukerche, and R. Pazzi, “A novel multi-hop clustering scheme for vehicular ad-hoc networks,” *Proc. MobiWac*, Florida, USA, pp.19–26, 2011.
- [21] M. Royer, “An enhanced 1-hop clustering algorithm for publish/subscribe systems in AANET,” 2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC), Prague, Czech Republic, pp.1–25, 2015, doi: 10.1109/DASC.2015.7311542.
- [22] T. Bilen, P.J. Aydemir, A.E. Konu, and B. Canberk, “Customized K-means based topology clustering for aeronautical ad-hoc networks,” 2021 IEEE 26th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), pp.1–5, 2021, doi: 10.1109/CAMAD52502.2021.9617810.
- [23] T. Bilen and B. Canberk, “Learning-vector-quantization-based topology sustainability for clustered-AANETs,” *IEEE Netw.*, vol.35, no.4, pp.120–128, July/Aug. 2021, doi: 10.1109/MNET.011.2000688.
- [24] M. Shabbazi, M. Simsek, and B. Kantarci, “Density-based clustering and performance enhancement of aeronautical ad hoc networks,” 2022 International Balkan Conference on Communications and Networking (BalkanCom), Sarajevo, Bosnia and Herzegovina, pp.51–56, 2022, doi: 10.1109/BalkanCom55633.2022.9900681.
- [25] D. Liu, J. Zhang, J. Cui, S.-X. Ng, R.G. Maunder, and L. Hanzo, “Deep-learning-aided packet routing in aeronautical ad hoc networks relying on real flight data: From single-objective to near-pareto multiobjective optimization,” *IEEE Internet Things J.*, vol.9, no.6, pp.4598–4614, March 2022, doi: 10.1109/JIOT.2021.3105357.
- [26] B. Heng, H. Zhang, and G. Huang, “Analysis of link dynamics in aeronautical ad hoc networks,” *J. Systems Engineering and Electronics*, vol.35, pp.420–424, 2013, doi: 10.3969/j.issn.1001-506X.2013.02.32.



Zheng Chen received the B.S. degree in Huaiyin Institute of Technology in 2021. He is currently pursuing a M.S. degree at Civil Aviation University of China.



Hongyu Yang received the B.S., M.S. and Ph.D degrees in 1992, 1997 and 2003, respectively. He got his Ph.D degree in Tianjin University, China. He is a professor now working in Civil Aviation University of China.



Laiwei Jiang received the B.S., M.S. and Ph.D degrees in Harbin Institute of Technology in 2009, 2011 and 2017, respectively. During 2010–2011, she stayed in Tokyo Institute of Technology to study as an exchange student. During 2013 to 2014, she stayed in Columbia University as a Visiting Ph.D student. She is a lecture now working in Civil Aviation University of China.