

Clustering-based Data Dissemination Protocol Using the Path Similarity for Autonomous Vehicles

MinSeok Seo ¹, SeungGwan Lee ^{2,*} and Sungwon Lee ³

¹ Department of Computer Science and Engineering, Kyung Hee University, Gyeonggi-do 17104, Korea; msseo@khu.ac.kr

² Humanitas College, Kyung Hee University, Gyeonggi-do 17104, Korea

³ Department of Software Convergence, Kyung Hee University, Gyeonggi-do 17104, Korea; drsungwon@khu.ac.kr

* Correspondence: leesg@khu.ac.kr; Tel.: +82-31-201-3711

Received: 24 December 2018; Accepted: 15 February 2019; Published: 19 February 2019

Abstract: The VANET (Vehicle Ad Hoc Network) is gathering attention for autonomous vehicles and the MANET (Mobile Ad Hoc Network) is attracting interest as well. Therefore, efforts have been made to overcome the challenges of the VANET in which the topology changes in real time and instability exists due to the difference in speed and physical phase. Particularly in the IoT era, the total amount of network nodes in addition to vehicle nodes is expected to increase dramatically. Therefore, a clustering algorithm for a mesh network capable of autonomous configuration is suitable for reducing the load of the central control device and data redundancy on the network, which is expected to increase as the IoT era progresses. However, since clustering algorithms based on the existing research have been developed for the current traffic situation, inefficiency is inevitable in the future autonomous navigation period in which traveling path prediction can be accurately performed. Therefore, this paper discusses a clustering algorithm and a data propagation algorithm between clusters using path information. The main content of this paper is as follows. First, we propose a clustering algorithm using path information and considering the existing research results. In the autonomous navigation period, if the path is predictable, the probability that the nodes in the same cluster are in the same block for a longer time than the conventional one can converge to 100%. Therefore, the survival time of the cluster can be dramatically improved. Second, we developed a data propagation algorithm that can increase the information propagation rate of the entire network using path information. The cluster temporarily stores the data to be disseminated and then disseminates it when it encounters another cluster of neighbors. Therefore, data can be disseminated even for noncontiguous clusters. To summarize, this paper proposes clustering-based data dissemination algorithms and protocols using vehicle pathways for autonomous navigation and compares them with clustering-based data dissemination algorithms using existing directions.

Keywords: vehicular ad-hoc network (VANET); clustering-based data dissemination; path-based clustering; mesh network; autonomous network; wireless access in vehicular environment (WAVE); vehicle-to-everything (V2X)

1. Introduction

In 2020, the number of vehicles in the United States is expected to increase by 247 million [1]. The number of Vehicle-to-everything (V2X) entities is also growing rapidly, and by 2020 more than 70 million vehicles worldwide are expected to utilize V2X communications for weather, traffic congestion, emergency vehicle access, and other purposes, compared to 10 million vehicles using V2X communications in 2015 [2]. Also, according to reference [3], the maximum payload that is

required in V2X communications is 1200 bytes. Assuming that 70 million nodes transmit data with these payloads once every 10 seconds, the maximum payload reaches 1,933 PB per month (peta bytes per month). In the IoT era, the total amount of network nodes as well as vehicle nodes is expected to increase significantly. According to reference [4], the IoT installed base is estimated to reach 30.73 billion by 2020. Rather than centrally control this vast amount of traffic through a cellular network, efforts have been made to reduce the load on the control by constructing a mesh network and a VANET between the road side unit (RSU) and the on-board unit (OBU). For example, there is geocast routing using zone of relevance (ZOR), directional data dissemination protocol using angles between nodes, and the distributed and mobility-adaptive clustering algorithm (DMAC) considering the speed difference of nodes [5–7].

Therefore, this paper proposes a clustering algorithm using path information and a data dissemination algorithm between clusters. The main content of this paper is as follows.

First, we propose a clustering algorithm using path information by improving the related research. In a related study, it was assumed that each vehicle node is G/G/1 Queue and the vehicle follows Brownian motion. In this case, the authors modeled the node's successful transmission rate as the probability that nakagami-m fading would not occur at all or would occur within a certain distance [8,9]. However, this model ignores the fact that vehicle traffic can be modeled as a Markov chain Monte Carlo system (MCMC) because the actual road environment does not know the traveling paths of other vehicles [10]. Therefore, the vehicle position cannot be considered to follow the complete Wiener motion, and it is correct to model each intersection with MCMC using numerous iterations.

However, if the path prediction is more accurate than the MCMC in the autonomous mobile era, the probability that nodes in the same cluster are in the same segment for a longer period of time converges close to 100%. As a result, you can dramatically improve the lifetime of your cluster. Additionally, as the lifetime of the cluster increases, the overhead of the cluster reconstructing process decreases and the ratio of lost data dissemination messages during the process can be reduced.

Second, we developed a data propagation algorithm that can increase the information propagation rate of the entire network by using path information. The cluster head is propagated using the member closest to the neighboring cluster as a gateway role, and the neighboring cluster is a cluster directed to another path. The cluster temporarily stores the data to be deployed, and then distributes it when it encounters another neighbor cluster. Since the cluster that receives the data is a cluster with a different path, the cluster that is newly encountered is likely to be a cluster that is encountered in another path. Thus, the data can spread even in non-contiguous clusters.

In summary, this paper proposes clustering-based data dissemination algorithms and protocols that use vehicle paths for autonomous navigation and compares them with existing clustering-based data dissemination algorithms using possibilities.

1.1. Clustering Algorithms

Clustering algorithms have been applied in many fields. For example, some basic algorithms such as from reference [11] and reference [12] are provided as meta-algorithms, and others have been used in software engineering as well [13]. However, the most recently used fields are related to self-organizing networks such as wireless sensor networks (WSN) or ad-hoc networks such as Refs. [14–16]. These studies noted the utility of managing clustering when combined with a mesh network. In other words, it has a smaller controller load than central management, but flooding occurs more often than for a fully-connected mesh network. Taking advantage of clustering on the ad-hoc network claimed in these papers, we note that clustering can have similar advantages over V2X mesh networks that share the basic characteristics of mesh networks.

In particular, VANET based on a clustering algorithm is suitable for reducing the load of the central control unit, which will increase during the IoT era, and will reduce the data redundancy on the network [17]. Distributed clustering is also applicable to location-unaware WSN, and VANET belongs in this category [18].

However, the existing clustering algorithms based on the current traffic environment are not optimized for autonomous driving when the accurate path of the vehicle can be known beforehand [19].

Therefore, this paper proposes a clustering algorithm that is specialized for autonomous driving environment.

1.2. Path Similarity

In this paper, path similarity means the degree of similarity between two paths. The path similarity is mainly used for graph matching such as skeleton graph matching as in Refs. [20–22].

However, path similarity is often not used in routing algorithms. Similar to the path similarity, geographical data such as directionality are used for routing. For example, Leontiadis [23] and Chen [24] correspond to geographical routing. These geographical routing schemes are suitable for overcoming frequent topology changes in WSNs with high mobility.

The reason why the path similarity is not used in the routing algorithm is that it cannot predict the entire movement path of each object in the system, such as WSN or VANET mentioned above.

Since existing clustering algorithms often target fixed or infrequently moving nodes, it is difficult to apply them directly to V2X with active mobility. If path similarity is available, it can increase the connectivity and maintainability of the cluster, but existing vehicles cannot know the full path information. So instead of path similarity, use directionality or the connection rate.

However, this paper focuses on the autonomous vehicle environment. In this case, each node, that is, each vehicle can know all the routes to the destination. Therefore, it is possible to utilize the route information which has been used only in the graph algorithm, which can clearly know the entire route, in the routing. This means that we can utilize path information that is remarkably accurate compared to the directionality or connection rate that can only estimate the path. In other words, it can identify nodes that are moving in a similar path, which contributes to the improvement of the survival of clustering. It is therefore possible to identify the nodes that are moving in a similar path. Thus, by constructing clusters among the nodes going to the same route in the future, the algorithm contributes to improvement of the survival of clustering.

Accordingly, in this paper, we propose a clustering algorithm based on path similarity based on the assumption that each object knows all its paths.

2. Related Works and Problems

2.1. DSRC/WAVE

Dedicated short-range communications (DSRC) is a long-range wireless communication method developed for the intelligent transport system (ITS) and is a protocol published by the US Federal Communications Commission (FCC) [25]. This method was originally developed only for bill collection and used in the 75 MHz frequency band but was changed to the 5.9 GHz frequency band to overcome low performance for V2X communication. Accordingly, it was listed in the IEEE standard as 802.11p with the name WAVE [26].

Figure 1 shows the protocol stack of WAVE [27]. The WAVE PHY layer corresponds to the OSI 1 layer, and WAVE MAC, multichannel operation, and the LLC layer correspond to the OSI 2 layer. The upper layers are similar to normal ethernet layers.

According to a previous research study, WAVE has a decline in transmission success rate depending on the density of nodes [28]. Therefore, if WAVE is used to construct a mesh network for V2X in the future, it is essential to reduce interference by reducing unnecessary retransmissions in the transport layer protocol.

However, WAVE does not respond to such a reduction in redundancy and focuses on smooth communication between vehicles with high speed. Therefore, the utility of a vehicle in a city where the vehicle is crowded is drastically reduced and a solution is needed [29]. Depending on the node density, severe delays may occur [30].

Since the WAVE protocol is now the de facto standard for V2X communication, it is likely to be based on WAVE even in the era of autonomous navigation in the future. Therefore, to use WAVE for V2X communication, it is also essential to compensate for the weakness of WAVE, but related research is not active.

As described above, the clustering-based data dissemination protocol is a solution to this issue. It is possible to configure the cluster autonomously without central management and reduce the redundancy by controlling the cluster head [31].

However, a node in VANET has very strong mobility. Therefore, the link is very unstable and vulnerable, and its topology changes from moment to moment [32]. Because of this, clustering-based algorithms are not easily maintained in a composed cluster. This is particularly the case in urban areas [33].

Thus, in this paper, we discuss how clustering-based data dissemination can increase the survivability of clusters and reduce the redundancy of a future internet environment.

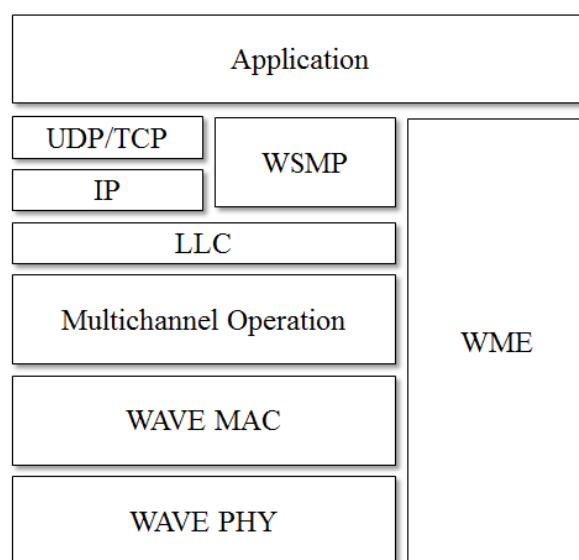


Figure 1. 802.11p WAVE Protocol Stack.

2.2. Data Dissemination Scheme based on Clustering and Probabilistic Broadcasting

In May 2018, Lei Liu announced the data dissemination scheme using broadcasting and clustering based on connection probability [34]. In this study, the authors proposed clustering and probabilistic broadcasting (CPB), which is the main concept when constructing clusters based on cumulative connection probabilities through continuous connectivity verification.

This paper is significant in that it is a clustering algorithm that maximizes connectivity in unstable V2X environments. However, the data rates they define are not consistent with the actual road environment.

$$P_{ij}^f = 1 - F_d(r_t, m, \psi) = e^{-\frac{m\gamma_T}{\psi} \sum_{i=1}^m \frac{(\frac{m\gamma_T}{\psi})^{i-1}}{(i-1)!}} \quad (1)$$

$$\begin{aligned} p(x|d_0, t) &= \Pr(x \leq D(t) \leq x + dx | d_0) \\ &= \frac{1}{\sigma\sqrt{2\pi t}} \sum_{n=-\infty}^{\infty} \left[\exp\left\{ \frac{\mu x'_n}{\sigma} - \frac{[x - d_0 - x'_n - \mu t]^2}{2\sigma^2 t} \right\} - \exp\left\{ \frac{\mu x'_n}{\sigma} \right. \right. \\ &\quad \left. \left. - \frac{[x - d_0 - x'_n - \mu t]^2}{2\sigma^2 t} \right\} \right] \end{aligned} \quad (2)$$

The authors define the probability P_{ij}^f that packet transmission between nodes i and j succeeds as Equation (1). In Equation (1), r_t is the threshold of the signal strength that can be received, ψ is the average signal strength, and m is the fading parameter. m is 1 if the distance between vehicles

is more than 150 m, 1.5 when it is more than 50 m and 3 if it is less than 50 m. F_d presents the probability density function of possibility that the nakagami-m fading will occur. That is, the probability that fading due to the nakagami-m distribution will not occur [35].

Additionally, the authors assumed that each vehicle follows the Wiener process. Equation (2) is used to derive the probability $p(x|d_0, t)$ that the vehicle will be at a certain radius.

$$LC_{ij} = F_{i,j}(t) = 1 - \int_0^R p(x|d_0, t)dx \quad (3)$$

Equation (3) defines the link connectivity LC_{ij} using P_{ij}^f derived from Equation (1) and $p(x|d_0, t)$ derived from Equation (2). The link quality lq_{ij} defined in the paper is calculated as Equation (4).

$$lq_{ij} = P_{ij}^f LC_{ij} \quad (4)$$

After the clusters are configured, they determine the forwarding probability as follows. First, a timer is started for a t_{int} interval which is a preset data transmission interval. Then, when the cluster member receives a disseminated message while the timer is running, it increments the counter t_c . The transmission probability P is determined according to the numerical value of the counter t_c as shown in Equation (5).

$$P = \begin{cases} 1, & C = 1 \\ \frac{2}{C^2}, & 2 \leq C \leq Threshold \\ 0, & C > Threshold \end{cases} \quad (5)$$

Note that P is a probabilistic variable that determines whether a cluster member will deliver it probabilistically when it receives a message, and does not mean a transmission success rate. That is, if P_{ij}^f in Equation (1) represents the success rate of the physical transmission of the communication, then P in Equation (5) is a variable of the probabilistic transfer algorithm which is one of the core solutions of CPB.

Here, $Threshold = [6V_r]$, and V_r is the speed ratio, which is the current vehicle speed V_c divided by the vehicle maximum speed V_f . The authors estimate this value using the number of vehicles on the road, but in this paper, we used the actual vehicle speed and the maximum speed when reproducing.

However, in reality, V2X communication not only includes various types of fading such as coherent fading, k-type fading as well as nakagami-m fading, but also an environment in which multiple constraints are applied in addition to fading. Therefore, it is difficult to verify the efficiency with simple arithmetic calculations. In this paper, we will compare this with a simulation environment which depicts a real environment rather than efficiency based on simple arithmetic operation.

Also, the authors did not take into account the probability that the route will change to link quality, an option which is favorable in existing environments such as highways, but can be extremely disadvantageous in urban areas with frequent path deviations. Clustering algorithms using a simple path loss model cannot guarantee future connectivity. When the cluster is disconnected frequently, the overhead is increased, and the message to disseminate is more likely to be lost.

Moreover, since all nodes must continuously check the connectivity, flooding may occur due to discovery messages in the city where the vehicle is driven. This is considered to be unsuitable for future Internet environments where there are numerous nodes besides vehicles.

In addition, the conventional algorithm does not consider roads with several lanes when determining the probability that a message is forwarded using density. In an actual driving environment, except for a highway, the lane serves as one of the critical variables that changes the route. Because of these problems, it seems that the existing algorithms are less effective in urban scenarios.

In particular, if we look at the left side of Figure 2, it can be confirmed that only one-dimensional positional variables can be considered since all lanes can be regarded as straight lines when the highway is generalized. However, when we look at the right side of Figure 2 considering roads with many lanes in actual cities, it can be seen that the lane information of the road not only refers to the position information between the left and right, but also affects the density of the node. That is, the number of lanes also affects estimations that utilize density [36].

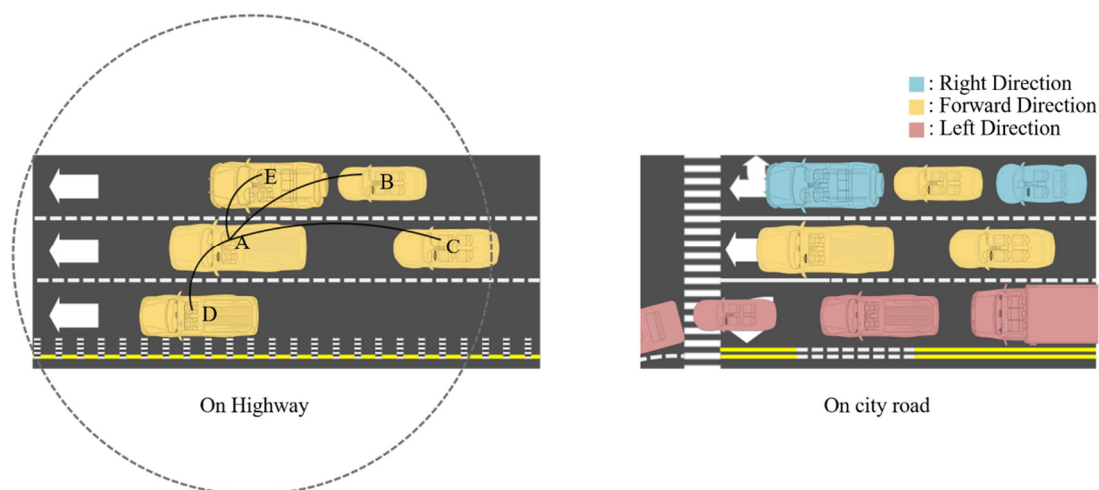


Figure 2. Difference between Highway and City Situations.

For example, a vehicle on a left-hand lane and a straight-ahead lane may have high connectivity, but in reality, it only takes a few seconds to move in a different direction. Clustering algorithms based solely on link probabilities are difficult to distinguish. However, if the exact path of each vehicle can be known in advance, it can be seen that the paths are compared with each other and the path changes soon.

Thus, in this paper, we will discuss how to obtain the advantages and overcome the disadvantages of existing research if we assume the above situation and have path information for each node in the future.

3. Proposed Method

In this section, we discuss the path-based clustering protocol proposed in this paper.

3.1. Term definitions

The definitions of the main terms used in this paper are shown in Table 1.

Table 1. Term Definition.

Term	Definition
Unclustered node	Node not joined to any cluster
Cluster head	A node that manages the cluster in a cluster
Cluster member	A node belonging to a cluster but not a cluster head
Path distance	Similarity distance between path information of each node
Data dissemination rate	The percentage of messages disseminated across the network
Survivability	The degree to which a cluster member can be maintained as a member of a cluster for a unit of time
Data Dissemination	Delivering messages to the entire network so that as many nodes as possible receive

3.2. Path-based Clustering Model

3.2.1. Concept Definitions

The proposed path-based clustering data dissemination protocol (PCDP) consists of two parts.

The first part is a clustering method that forms clusters based on the expected path of the vehicle. When constructing a cluster, the survivability of the cluster is improved by constructing a cluster between nodes with the smallest difference by comparing the planned paths of the vehicles. In particular, in the city area where the route is frequently used, there is a certain probability that the node will depart due to the change in route, even if the connection rate is high, which accumulates every time the route distribution section includes the intersection passes.

As can be seen in Figure 3, nearby vehicles can point in different directions at the next intersection. If the red vehicle in Figure 3 goes to the left, the yellow car goes straight, and the blue car goes to the right, there is a possibility that all of them belong to the same cluster according to the current communication intensity. However, they are clearly dispersed in other directions, and the cluster will collapse in an instant.

If the survivability of the cluster is low, the clustering process occurs frequently and the overhead increases rapidly. Data dissemination is also delayed due to the clustering process.

According to Yu Guoqiang, road conditions can be defined as a Markov chain in Bayesian terms [37]. This supports the fact that the better the future path, the better survivability of clusters.

In this paper, we propose a method for constructing a cluster that can last for a long time using path information.

The second part is a data dissemination method that can reduce the traffic redundancy and how much data can be delivered to the nodes on the network, that is, maximize the data dissemination rate. In this paper, we try to maximize the data dissemination rate by using the path-based cluster as a medium for message delivery. Each cluster has a high probability of heading to a different path, so there is a high probability that it will head to a path with a cluster that has not received the message. Each cluster moves to store the forwarded message and rebroadcasts the message each time it encounters a new neighbor cluster. This process maximizes the data dissemination rate of the entire network.

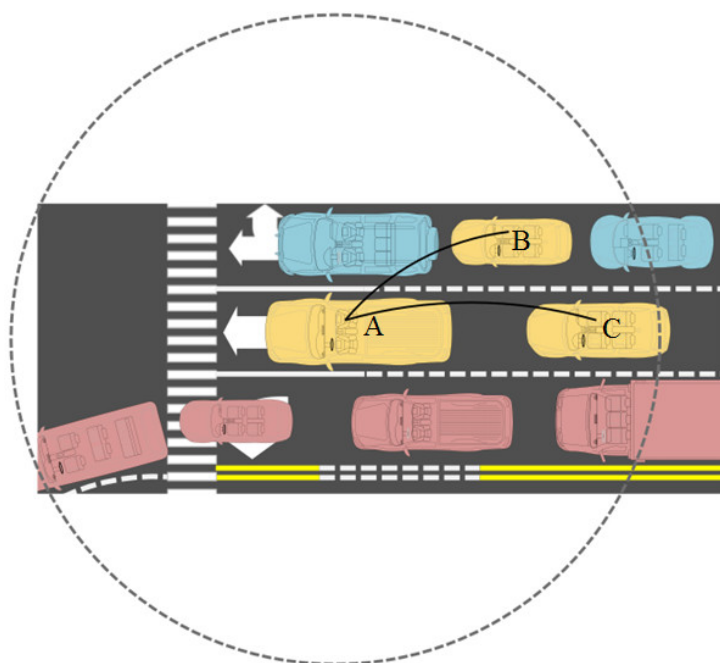


Figure 3. Cluster Topology in PCDP.

3.2.2. Use-Cases of PCDP

The clustering algorithms proposed in this paper are most effective when operating in cities with many intersections. In urban areas, path deviation of the node occasionally occurs, which means that the survivability of the cluster is degraded.

The robustness of the cluster is especially important in urban areas. Figure 4 shows a real part of the downtown area of Seoul, Korea and its implementation in the simulation of urban mobility (SUMO) traffic simulation. It can be seen that there are numerous paths in an area of only 1 km². In other words, there is a high possibility of deviation from the route in the city area. In such a scenario, for example, the following may occur:

There may be situations in which an ambulance crosses several intersections and the survivability of the cluster is expected to drop significantly. In addition, congestion may lead to multiple paths. Most often, there are intermediate destinations and they have different paths.

Moreover, in a situation where the route is controlled by autonomous driving, each vehicle may be assigned different routes in advance to avoid congestion. In such a situation, the path is inevitably staggered.

It is important to give priority to instantaneous connectivity in a situation where a route such as a highway is close to a straight line, but it is more appropriate to use the route information if it can be used in a situation where the route is likely to be divided.



Figure 4. Urban Scenarios with High Potential for Path Deviations.

3.3. Clustering Algorithm for Applying the Path-based Clustering Model

In this section, we will discuss the PCDP's clustering process and the message format required.

3.3.1. Path Distance Algorithm

This protocol uses the Levenshtein distance to find the difference in routes between vehicles. Assuming that the total path of the vehicle node n is p_n , the i -th road ID on p_n is r_{ni} and the length of p_n is l_n , the algorithm for calculating the distance of the path is shown in Algorithm 1. Also, Algorithm 1 has a time complexity of $O(n)$ because it computes the path distance using the finite n road information that each neighbor has.

Algorithm 1 Path distance algorithm

1. Set $i = 1, j = 1, d = 0$
 2. For each r_{mi} and r_{nj} of p_m, p_n
 3. If r_{mi} is the same as r_{nj} , then
 4. Add 1 to i and j
 5. Else then
 6. Add 1 to i and d
-

7. Add $(l_m - j)$ to d
8. Set d as d over $\max(l_m, l_n)$

The calculated path distance is used as an important factor for determining priority in the clustering process. Clustering can be performed as in Algorithm 2 below using the threshold th , where $0 < th < 1$. and having a time complexity of $O(n)$. In addition, Algorithm 2 has a time complexity of $O(n)$ since it calculates Path distance for finite n neighbor nodes.

Algorithm 2 Clustering algorithm with path distance

1. Set $b = 0$, $flg = \text{false}$, parameter th , assign p
2. If there is a cluster head among neighbors, then
 3. While not found proper cluster head and flg not set
 4. For each $m \in \{\text{neighbors}\}$
 5. Calculate path distance d_m
 6. If $d_m < b$, then
 7. Set $b = d_m$
 8. Set $p = m$
 9. If $b < th$, then
 10. Send cluster assignment request to p
 11. Else then
 12. Set timer t
 13. Else then
 14. Broadcast cluster head announcement
 1. Sub procedure t :
 2. If there is no d_m lower than th , then
 3. Send cluster head announcement

In this case, th is derived as Equation (6).

$$th = \frac{M}{N} \quad (6)$$

In Equation (6), M is the number of neighboring neighbor clusters and N is the number of neighboring nodes. That is, as the number of neighboring clusters is smaller than the number of neighboring nodes in the periphery, the node selects the cluster to join with care. In other words, it means to create a cluster if possible. In contrast, if the number of neighboring clusters is larger than the number of neighboring nodes, there are many clusters in the vicinity.

3.3.2. Clustering Process

The cluster status of this protocol can be divided into three categories: (1) unclustered node, (2) cluster member node, and (3) cluster head node.

Figure 5 shows this in a state machine diagram. Each state and condition that changes with each other is as follows.

First, all nodes are in the unclustered node state after initialization. The cluster head periodically broadcasts a head announcement message instead of a discovery message. When the unclustered node receives this information, it first confirms that it has collected all the neighbor's path information. Path similarity is measured using the path information of the announced heads in the state where path information exists. After measuring the similarity of the neighboring clusters

through the path data, the unclustered node requests to join the specific cluster with the best path distance. If the confirmation is answered, the unclustered node enters the cluster member state.

When becoming a cluster member, a node proves its connectivity by responding to the cluster head announcement periodically requested by the discovery answer. However, if an announcement is not received more than three times from the participating cluster head, we return to the unclustered node and perform discovery. In addition, when the path distance value to the cluster head of the neighbor cluster to which the cluster does not belong is better, the cluster head sends a join request message to the cluster head through the cluster assignment message. When a confirmation is received, it becomes a member of the cluster, just like an unclustered node.

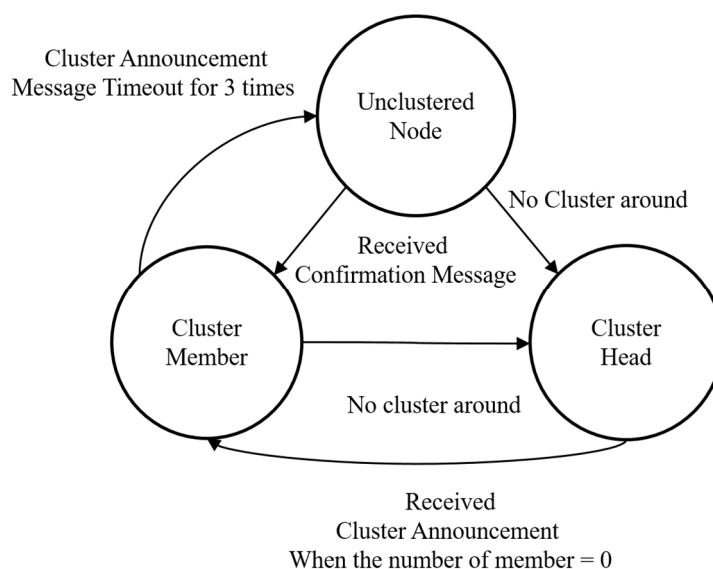


Figure 5. State Diagram of the PCDP Node.

However, if there are no nodes in the vicinity, the node will appoint itself as cluster head and implement cluster head announcement. In this state, if three neighboring nodes have not responded and after three timeouts, they become unclustered nodes again and perform discovery. When the node becomes a cluster head, its ID is used as a cluster ID. The neighboring nodes can confirm that the cluster head is the cluster head by confirming that the cluster ID is the same as the node ID.

If the node responding to the cluster head announcement message is an unclustered node, it requests the path data and performs the cluster head announcement again.

In contrast, if the node responding to the cluster head announcement message is another cluster head node and the cluster size of the other party is large, the path distance is checked. If the path distance is within the threshold, it returns to the unclustered node and performs discovery again.

Each node has a neighbor table as shown in Figure 6, as well as exchanges path information at the first contact. This causes overhead, but since the path information does not change, the exchange can be performed only once. Therefore, in the long term, the ratio of total traffic overhead is small compared to the overhead due to frequent connectivity verification.

Node ID	Cluster ID	Paths
14	156	312; 313; -313; -210; ...
156	156	312; 313; -313; -211; ...
...	...	...

Figure 6. Example of a Neighbor Table.

3.3.3. Message Format for a Clustering Process

The proposed PCDP has the following general message protocol. Figure 7 illustrates the general message protocol of the PCDP. All protocols of the PCDP are derived from the PCDP general message format. PCDP is a protocol based on WAVE Short Message and is basically an implemented protocol for the transmission method, such as Transmission Control Protocol (TCP) and User Datagram Protocol (UDP).

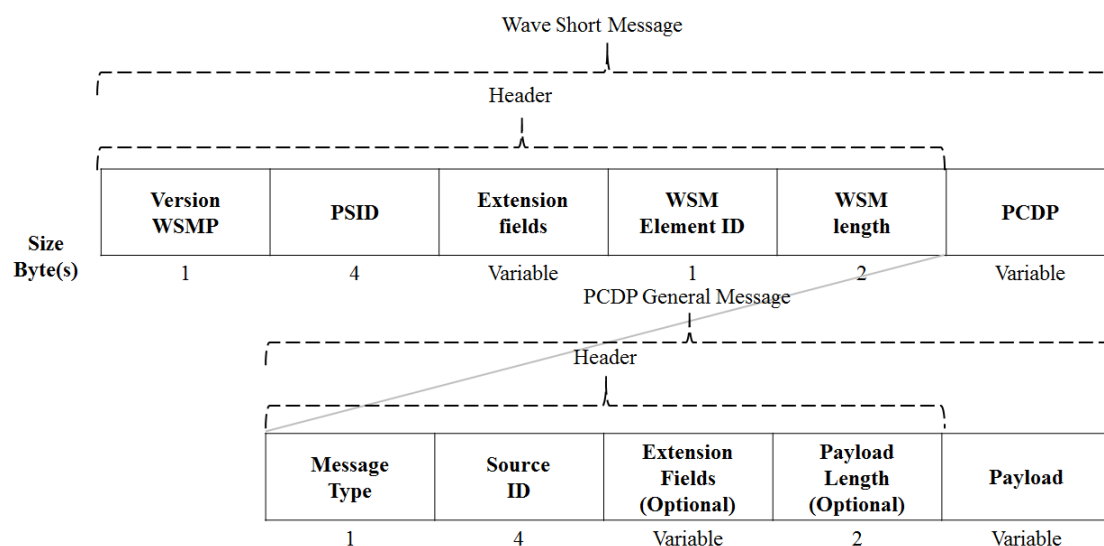


Figure 7. PCDP General Message Architecture.

The discovery message is shown in Figure 8. This is a message used in the discovery process described above, in which a neighbor is identified, or a new neighbor cluster is identified in a situation where a cluster is not assigned. In particular, since members become cluster members, ordinary members do not use discovery messages. This is because the cluster head is connected to 1-hop and the dissemination message is unconditionally forwarded to the head.

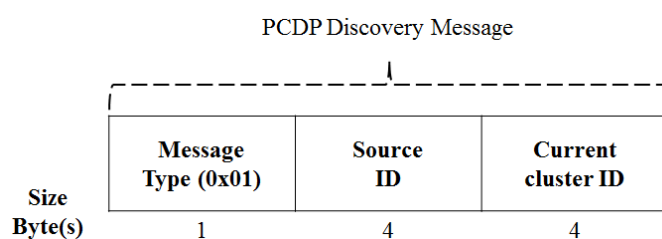


Figure 8. PCDP Discovery Message.

The discovery answer message in Figure 9 is the reply to the node when it receives the discovery message. This is used to accumulate the connection probability in the case of the existing CPB, but it is used only in the PCDP to identify the neighboring node.

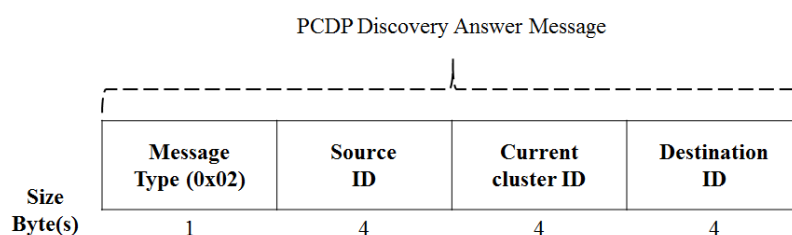


Figure 9. PCDP Discovery Answer Message.

The head announcement message in Figure 10 is used to notify the neighboring node that it is the head. Members use the discovery answer message to report to the head that they are connected.

If the unclustered node receives it, it asks the subscription according to the priority following the path information, as described above.

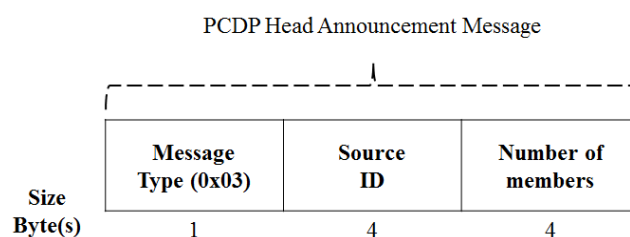


Figure 10. PCDP Cluster Head Announcement Message.

Figure 11 illustrates the cluster assignment request message format. For a node receiving an announcement from a neighboring cluster head, generally, an unclustered node determines whether to join by considering the path distance. If the node wants to join, the ID of the cluster head in the cluster ID field of the cluster assignment request message is unicast.

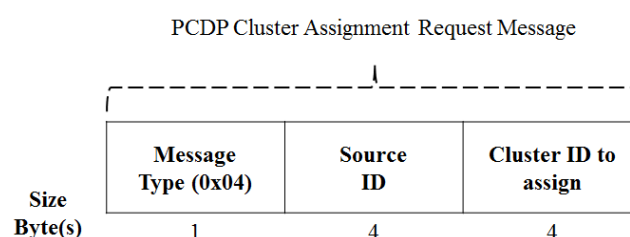


Figure 11. DCDP Cluster Assignment Request Message.

The format of the cluster confirmation message is shown in Figure 12. The cluster head that received the cluster assignment request from the neighbor node normally receives it and informs it through the cluster confirmation message that it will participate as a cluster member.

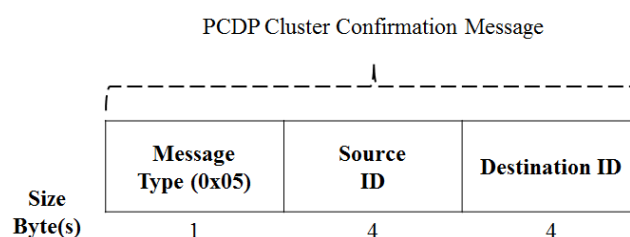


Figure 12. PCDP Cluster Confirmation Message.

Figure 13 shows the path announcement message format. When a node encounters a source ID for the first time, it automatically broadcasts a path announcement message, as shown in Figure 13. Each path is an integer of four bytes, which causes a large overhead when the first neighbor is encountered. However, this is a one-time message because the path has not changed. Therefore, it has a small effect in the long term.

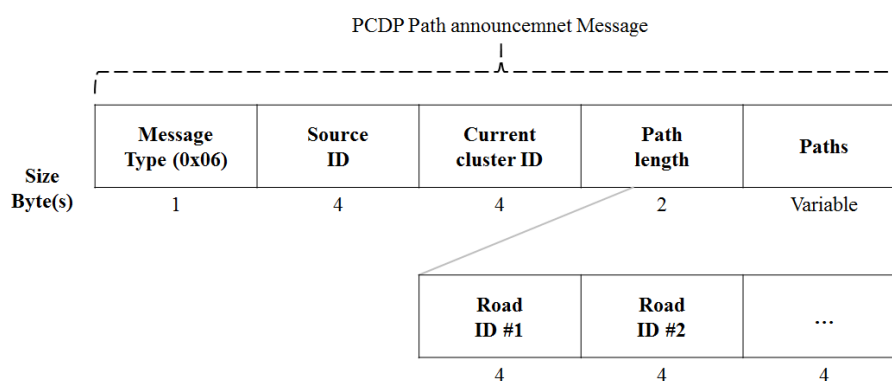


Figure 13. PCDP Path Announcement Message.

3.4. Data Dissemination Protocol

3.4.1. A Data Dissemination Algorithm

As described above, this protocol includes a data dissemination process. The main algorithm of the data dissemination process is given in Algorithm 3. Since Algorithm 3 can be described as a callback procedure that processes a message as it is received by each node, the time complexity is $O(1)$.

Algorithm 3 Data dissemination algorithm

Parameter m : received message, n : current node

1. If n is in a cluster, then
 2. If the source's cluster ID $\neq n$'s cluster ID, then
 3. Set m 's source and cluster ID field to n 's and broadcast m
 4. Else if n is the cluster head, then
 5. Set m 's source to n 's and broadcast m
 6. Else then
 7. Drop m
 8. Else then
 9. Set m 's source to n 's and broadcast m
-

When node n receives a dissemination message, it first checks to see if it belongs to a particular cluster. That is, it is verified that n is an unclustered node.

If the node is an unclustered node, the message is retransmitted because it is more important to reduce the redundancy than to increase the data propagation rate. When retransmitting, the source ID of the message is changed to itself and the cluster ID is -1, that is, an unclustered node.

If node n is a cluster head or a cluster member, we verify that the source cluster ID of message m is equal to itself. If it is not the same and if n is a head, we broadcast the field by changing the field of m . Additionally, m is stored in list lst , including the time it was passed. In the case of a message received from the same cluster member, retransmission is performed when there is a neighbor node belonging to another cluster. If not, message m is dropped.

The cluster head receives a neighbor attention message when it encounters a new neighbor cluster from its neighboring member nodes. At this time, if a message not previously transmitted is stored in lst , it is broadcasted. It also updates the timestamp of the message.

If timeout t_m has elapsed since the time of the first delivery of the message in lst , the message is deleted from lst . Timeout t_m is calculated as Equation (7).

$$t_m = \frac{2d_l}{v} \quad (7)$$

In Equation (7), v is the average speed (m/s) measured over the last minute of the head and d_l is the length of the normal block. In other words, this means that the message is stored and transmitted while passing through two blocks. Therefore, more messages are propagated to other clusters, and the information propagation rate of the entire network can be increased.

As shown in Figure 14, messages are stored in the list with the message specific ID. This field is an ID for internal identification, and it does not matter whether it is a non-overlapping line (such as a hash) using a source cluster ID that is first received and a timestamp of the first received time.

Message ID	Source Cluster ID	Timestamp	Payload
F230AC2D...	120	14:20:59	“Weather Forecast”
ACF3E32B...	152	14:23:59	“Accident Announcement”
...	...	...	...

Figure 14. Message Storage List.

Next, whenever a member encounters a new cluster, it forwards the cluster attention message to the head. At this time, the head broadcasts the stored message and records it in the forwarded message table.

Figure 15 shows the forwarded message table, which is described above. A cluster head does not broadcast if there is already a transmission record in the forwarded message table. If the message is timed out, the related record in the forwarded message table is dropped. Through this process, more redundant messages can be reduced.

Message ID	Cluster ID
F230AC2D...	135
F230AC2D...	148
...	...

Figure 15. Forwarded Message Table.

3.4.2. Message Format for a Data Dissemination Process

Figure 16 shows the packet structure used in the data dissemination process proposed in this protocol. As described above, in the data dissemination process, the receiving node changes the source ID and source cluster ID fields. The receiving node inputs its own ID into the source ID field when retransmitting and inputs the ID of the cluster to which the node belongs in the current cluster ID field.

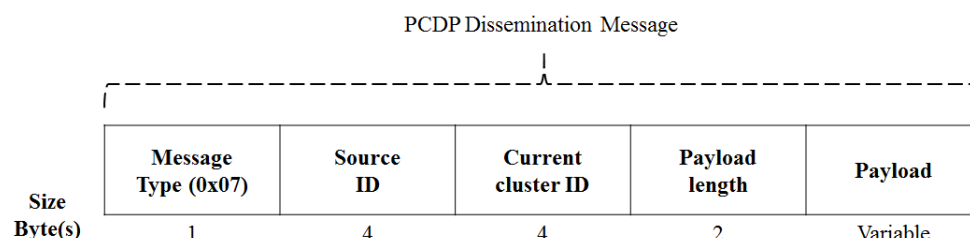


Figure 16. PCDP Dissemination Message.

If the member finds a cluster that is different from the neighbor of the neighbor, it sends a cluster detect message as shown in Figure 17. If it is not the head of the sender, it drops it and the head rebroadcasts the saved message as described above.

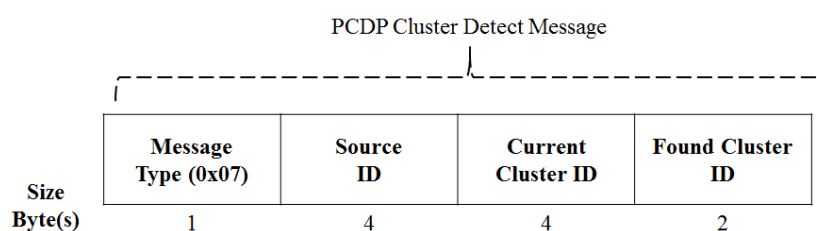


Figure 17. PCDP Cluster Detect Message.

4. Results

In this section, we compare the cluster holding time, cluster retention rate, message propagation rate, and overhead of the proposed clustering protocol compared to the existing CPB.

CPB is one of the most effective solutions to date on V2X's biggest challenge, which is mobility, and regards mobility as the biggest issue in this document. Because this is a problem that we pay attention to, this paper focuses on the comparison with CPB.

If the algorithms and protocols presented in this paper are superior to CPBs, it can be demonstrated that the solution to the autonomous V2X clustering solution using vehicle paths is more effective than the algorithms outlined so far.

4.1. Simulation Environment

Veins (Vehicles in Network Simulation) were used for the experiments. The software is an open source framework for OMNeT ++ and SUMO-based vehicular network simulation [38]. For this experiment, the intersection required for the experiment was constructed using SUMO [39].

Experiments were carried out by setting up a complete grid-shaped virtual zone rather than using actual roads to obtain formalized data. The shape of this area is shown in Figure 18.

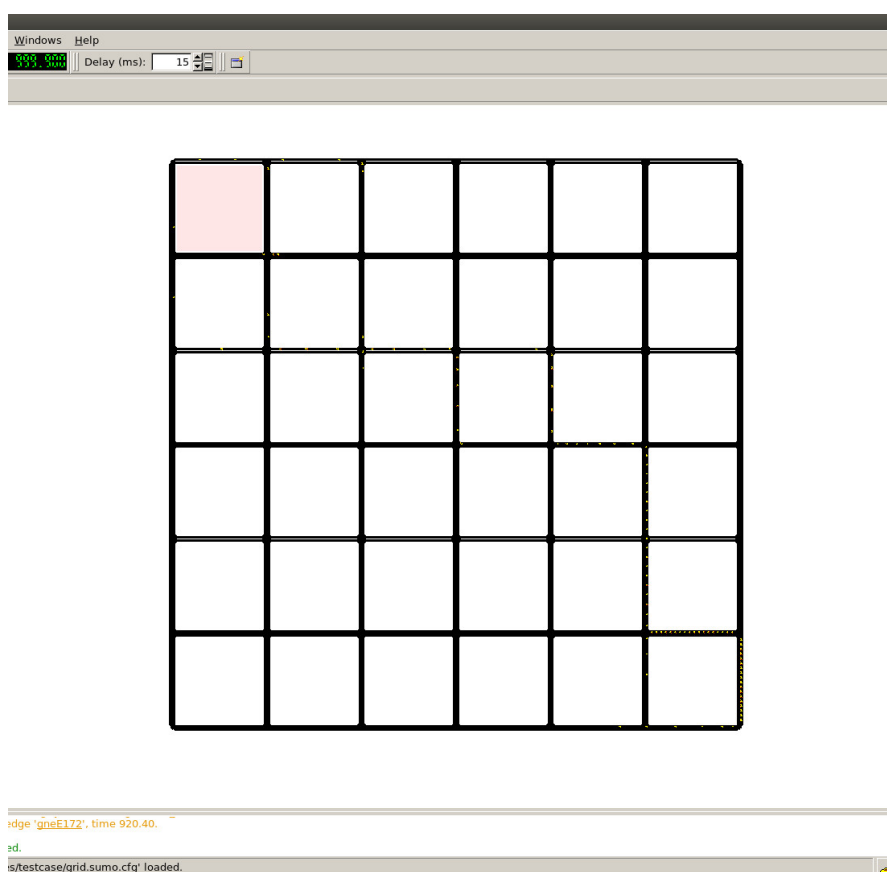


Figure 18. Simulation Area Built by SUMO.

Table 2. Simulation Variables.

Variable	Value	Unit
Area Size	36,000	m ²
Grid Size	1000	m ²
Size of Grid	6 × 6	
Maximum Speed	80	km/h
Number of Nodes Per Route	5	
Number of Routes	924	
Simulation Time	18,000	S
Bitrate	6	Mbps
Signal Strength	−20	dBm

Table 2 shows the variables used in the experiment. Roads 1000 m in length are arranged as 6 × 6 grids, and the simulation time is approximately 36,000 seconds using the whole route. Vehicles are designed to cross the number of paths that can come from the grid. In each case, five vehicles pass through, and there are 924 total routes.

4.2. Cluster Robustness Evaluation

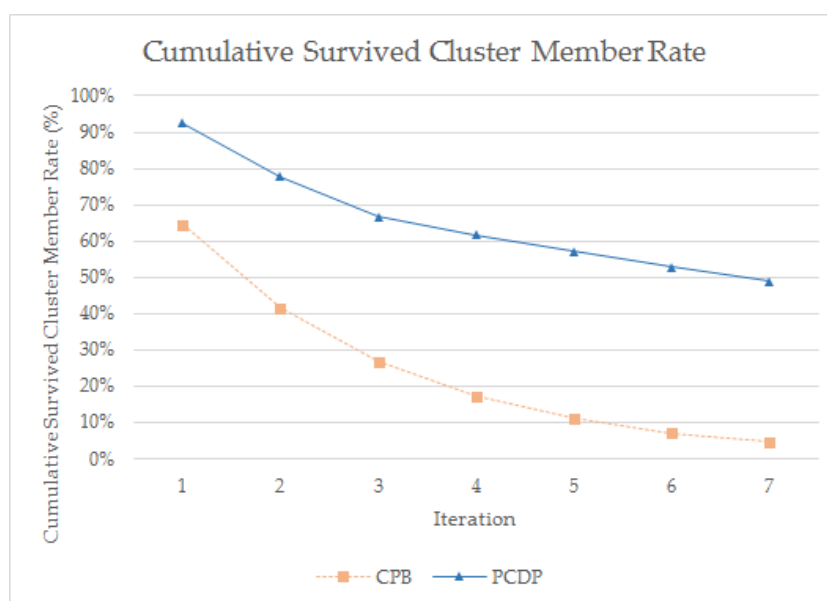
The survivability of the cluster is confirmed as follows. At the first road in the lattice-shaped experimental environment, 5620 cars start simultaneously. It then stores the first connected cluster members and the number of cluster members that are disconnected each time they cross an intersection.

The number of cluster members that have not been missed is then divided by the number of cluster members that were initially connected to each time they crossed the intersection. Because there are seven intersections in this experimental environment, we measure every seven iterations.

The results are shown in Figure 19. The number of cluster members in the CPB decreased sharply in the experimental environment with many intersections. In contrast, the PCDP showed that the number of members consisted of the first members, even after the intersections were steadily maintained.

Only 64.5% of the CPB maintained the first cluster on the entire network, even after the first iteration, and only 4.632% after the last iteration was the same as the first cluster member.

However, the percentage of PCDP in the same cluster after the first iteration reached 92.5%, and 57.92% remained after the last iteration.

**Figure 19.** Cumulative Survived Cluster Member Rate.

4.3. Overhead Evaluation

The measurement of protocol overhead proceeds as follows. Simulation with two protocols is performed to analyze the accumulated message amount per unit time.

The CPB protocol requires that all nodes send and receive discovery messages periodically, whereas PCDP does not have to do all nodes; only the cluster head periodically issues announcements. However, the PCDP must request and transmit path information whenever it encounters a new node. This is longer than discovery, but it is expected that overhead will be reduced in the long term because it is only required to be performed once per node.

The results are shown in Figure 20. If the node density is very low or the execution time is short, the overhead of the existing CPB may be smaller but insignificant, and the amount of overhead accumulated in the CPB is gradually increased.

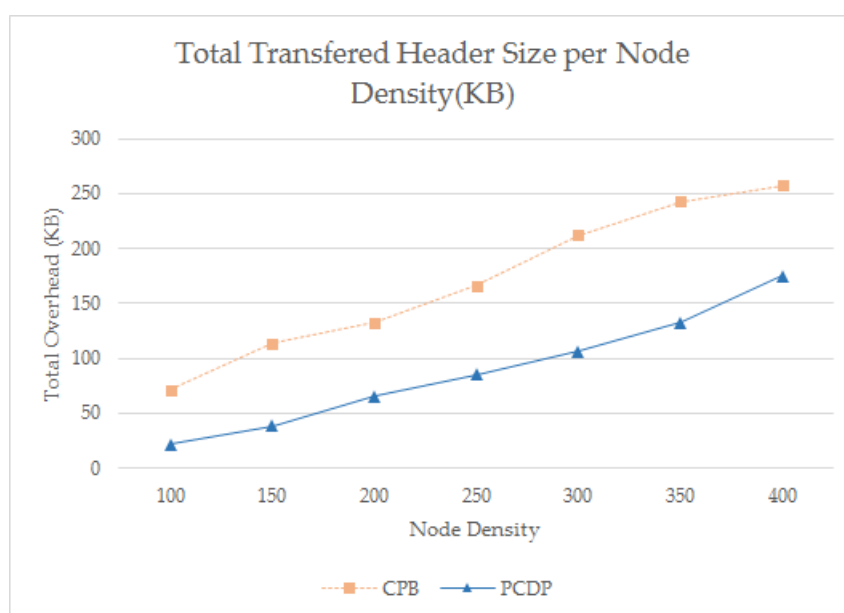


Figure 20. Total Transferred Header Size (Kb) per Node Density.

4.4. Data Dissemination Rate Evaluation

This section is about an evaluation for the data dissemination rate of the protocols. As previously stated, clustering data dissemination protocols should properly exclude messages from nodes other than their clusters to prevent flooding. If a protocol has been properly prioritized, it will have a higher data dissemination rate across the network compared to other protocols.

In this experiment, an arbitrary node is periodically selected every five seconds, and a dissemination message is sent from the corresponding node. When each message is disseminated and finally dropped and extinguished on the network, the number of nodes that existed in the area at the time of the first message transmission (i.e., the network size and the number of nodes that actually received the message) are measured.

Figure 21 is the message dissemination rate calculated by accumulating the number of nodes that successfully received the message and those that needed to receive the message. We see that the dissolution rate converges gradually as the two algorithms are repeated. Depending on the context within the scenario, the nodes may be close to each other or scattered, but the proposed protocol has a higher propagation rate than CPB. The converged dissemination rate is approximately 22.63% for CPB and 53.03% for the proposed protocol.

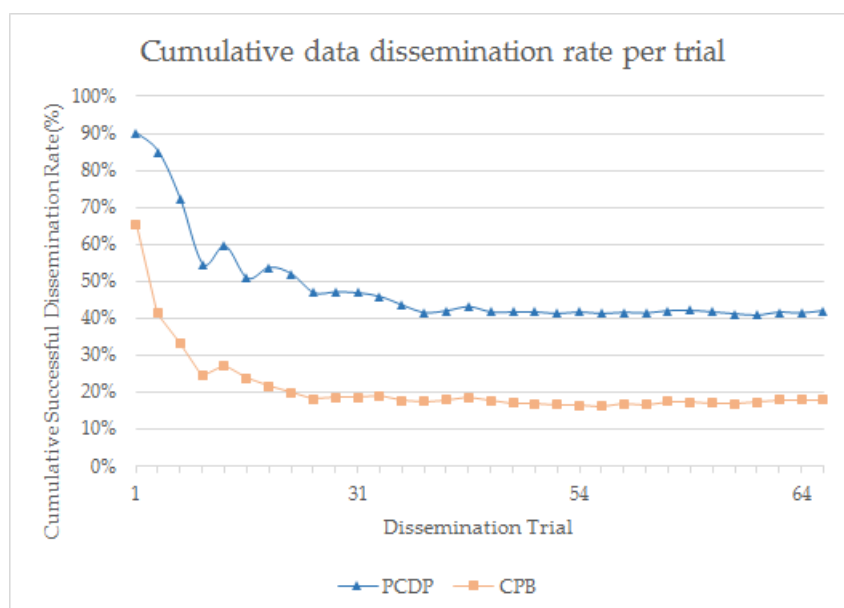


Figure 21. Cumulative Data Dissemination Rate per Dissemination Trial.

5. Conclusion

As the demand for V2X communication increases and the number of IoT entities increases, the need for VANET capable of autonomous configuration between nodes is emerging. Among them, the clustering-based algorithm is widely adopted as a compromise between a fully connected mesh network and a central control network via a controller.

The clustering algorithms can reduce the redundancy and control over the communication nodes without the central control by controlling the cluster head. However, it is fundamentally a method that is largely influenced by the survivability of the cluster. Therefore, there has been significant effort to overcome this problem, and the CPB protocol introduced in this paper is one such strategy.

Despite much effort, however, all the research is based only on the information currently available. Therefore, its effectiveness is greatly reduced in future Internet environments and cannot exceed the limit of the clustering algorithm. Therefore, this paper assumes an autonomous mobile automobile age network and develops algorithms with more efficiency, high robustness, and less overhead.

This paper presents the idea of clustering by using precise path information which has not been utilized or cannot be utilized. It is meaningful that the paper provided a clustering algorithm suitable for the autonomous vehicle era at present. In addition, this paper can be a stimulus to promote research in this field where such research is not yet active.

However, since it is difficult to conduct a large-scale autonomous driving experiment, there are limitations in not verifying efficiency when various kinds of fading are applied on actual roads.

Therefore, in the future, it is necessary to study the algorithm presented in this paper in actual road environment and overcome various physical limitations of vehicle nodes in V2X environment.

Author Contributions: All the authors contributed equally to this work. Conceptualization, Investigation, Writing – original draft, M.S.S.; Project administration, Supervision, Writing – review & editing, S.G.L. and S.L.

Funding: This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (2014-3-00547, Development of Core Technology for Autonomous Network Control and Management).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Arbib, J.; Seba, T. *Rethinking transportation 2020-2030: The disruption of transportation and the collapse of the internal-combustion vehicle and oil industries*; RethinkX: San Francisco, CA, USA, 2017.
2. The Next Generation Mobile Networks - NGMN, NGMN V2X White paper. Last modified 13 August 2018. Available online: https://www.ngmn.org/fileadmin/ngmn/content/downloads/Technical/2018/180813_Liaison_NGMN_C-V2X_v1.0.pdf (accessed on 10 November 2018).
3. 3rd Generation Partnership Project. *Study on LTE support for Vehicle to Everything (V2X) services (Release 14)*; Technical Report, 3GPP TR 22.885 V14.0.0 (2015-12); 3GPP: Valbonne, France, April 2015.
4. Columbus, L. Roundup of Internet of Things Forecasts and Market Estimates. *Forbes* **2016**. Available online: <https://www.forbes.com/sites/louiscolombus/2016/11/27/roundup-of-internet-of-things-forecasts-and-market-estimates-2016/#5e917ed4292d> (accessed on 5 December 2018).
5. Allal, S.; Boudjit, S. Geocast routing protocols for VANETs: Survey and guidelines. In Proceedings of 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), Palermo, Italy, 4–6 July 2012; pp. 323–328.
6. Schwartz, R.S.; Barbosa, R.R.; Meratnia, N.; Heijenk, G.; Scholten, H. A directional data dissemination protocol for vehicular environments. *Comput. Commun.* **2011**, *34*, 2057–2071.
7. Wolny, G. Modified dmac clustering algorithm for vanets. In Proceedings of 2008 Third International Conference on Systems and Networks Communications, Sliema, Malta, 26–31 October 2008; pp. 268–273.
8. Nakagami, M. The m-distribution—A general formula of intensity distribution of rapid fading. In *Statistical methods in radio wave propagation*; Elsevier: Los Angeles, CA, USA, 1960; pp. 3–36.
9. Whitmore, G. Estimating degradation by a Wiener diffusion process subject to measurement error. *Lifetime Data Anal.* **1995**, *1*, 307–319.
10. Ci, W.X.; Ahmed, S.K.; Zulkifli, F.; Ramasamy, A.K. Traffic flow simulation at an unsignalized T-junction using Monte Carlo Markov Chains. In Proceedings of 2009 IEEE International Conference on Signal and Image Processing Applications, Kuala Lumpur, Malaysia, 18–19 November 2009; pp. 346–351.
11. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosc.* **1984**, *10*, 191–203.
12. Hartigan, J.; Wong, M. Algorithm AS 136: A k-means clustering algorithm. *Appl. Stat.* **1979**, *28*, 100–108.
13. Lung, C.-H. Software architecture recovery and restructuring through clustering techniques. In Proceedings of the third international workshop on Software architecture, Orlando, FL, USA, 1–5 November 1998.
14. Abo-Zahhad, M.; Ahmed, S.M.; Sabor, N.; Sasaki, S. Mobile sink-based adaptive immune energy-efficient clustering protocol for improving the lifetime and stability period of wireless sensor networks. *IEEE Sens. J.* **2015**, *15*, 4576–4586.
15. Arain, Q.A.; Uqaili, M.A.; Deng, Z.; Memon, I.; Jiao, J.; Shaikh, M.A.; Zubedi, A.; Ashraf, A.; Arain, U.A. Clustering based energy efficient and communication protocol for multiple mix-zones over road networks. *Wirel. Pers. Commun.* **2017**, *95*, 411–428.
16. Lin, H.; Wang, L.; Kong, R. Energy efficient clustering protocol for large-scale sensor networks. *IEEE Sens. J.* **2015**, *15*, 7150–7160.
17. Vodopivec, S.; Bešter, J.; Kos, A. A survey on clustering algorithms for vehicular ad-hoc networks. In Proceedings of 2012 35th International Conference on Telecommunications and Signal Processing (TSP), Prague, Czech Republic, 3–4 July 2012; pp. 52–56.
18. Zanjireh, M.M.; Larijani, H. A survey on centralised and distributed clustering routing algorithms for WSNs. In Proceedings of 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), Glasgow, UK, 11–14 May 2015; pp. 1–6.
19. Gerla, M.; Lee, E.-K.; Pau, G.; Lee, U. Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds. In Proceedings of 2014 IEEE World Forum on Internet of Things (WF-IoT), Seoul, South Korea, 6–8 March 2014; pp. 241–246.
20. Bai, X.; Latecki, L.J. Path similarity skeleton graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1282–1292.
21. Xu, Y.; Wang, B.; Liu, W.; Bai, X. Skeleton graph matching based on critical points using path similarity. In *Asian Conference on Computer Vision*. Springer: Berlin, Germany, 2009; pp. 456–465.
22. Jin, T.; Jiang, B.; Luo, B. Graph Matching Based on Graph Histogram and Path Similarity. *J. Comput.-Aided Des. Comput. Graphics* **2011**, *9*, 3.

23. Leontiadis, I.; Mascolo, C. GeOpps: Geographical Opportunistic Routing for Vehicular Networks. In Proceedings of 2007 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, Espoo, Finland, 18–21 June 2007; pp. 1–6.
24. Chen, M.; Leung, V.C.M.; Mao, S.; Yuan, Y. Directional geographical routing for real-time video communications in wireless sensor networks. *Comput. Commun.* **2007**, *30*, 3368–3383.
25. Intelligent Transportation System Applications Using Dedicated Short-Range Communications, Federal Communications Commission Document, last modified 16 May 2001. Available online: <https://www.fcc.gov/document/intelligent-transportation-system-applications-using-dedicated-short-range> (accessed on 10 November 2018).
26. Group, I.W. IEEE Standard for Information Technology–Telecommunications and information exchange between systems–Local and metropolitan area networks–Specific requirements–Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 6: Wireless Access in Vehicular Environments. *IEEE Std* **2010**, doi:10.1109/IEEESTD.2010.5514475.
27. Uzcátegui, R.A.; De Sucre, A.J.; Acosta-Marum, G. Wave: A tutorial. *IEEE Commun. Mag.* **2009**, *47*, 126–133.
28. Campolo, C.; Vinel, A.; Molinaro, A.; Koucheryavy, Y. Modeling broadcasting in IEEE 802.11 p/WAVE vehicular networks. *IEEE Commun. Lett.* **2011**, *15*, 199–201.
29. Biddlestone, S.; Redmill, K.; Miucic, R.; Ozguner, Ü. An integrated 802.11 p wave dsrc and vehicle traffic simulator with experimentally validated urban (los and nlos) propagation models. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 1792–1802.
30. Gräfling, S.; Mähönen, P.; Riihijärvi, J. Performance evaluation of IEEE 1609 WAVE and IEEE 802.11p for vehicular communications. In Proceedings of 2010 Second International Conference on Ubiquitous and Future Networks (ICUFN), Jeju, Korea, 16–18 June 2010; pp. 344–348.
31. Momeni, S.; Fathy, M. Clustering In VANETs. In *Intelligence for Nonlinear Dynamics and Synchronisation*. Springer: Berlin, Germany, 2010; pp. 271–301.
32. Cai, X.; He, Y.; Zhao, C.; Zhu, L.; Li, C. LSGO: link state aware geographic opportunistic routing protocol for VANETs. *EURASIP J. Wirel. Commun. Netw.* **2014**, *2014*, 96.
33. Nzouonta, J.; Rajgure, N.; Wang, G.; Borcea, C. VANET routing on city roads using real-time vehicular traffic information. *IEEE Trans. Veh. Technol.* **2009**, *58*, 3609–3626.
34. Liu, L.; Chen, C.; Qiu, T.; Zhang, M.; Li, S.; Zhou, B. A Data Dissemination Scheme based on Clustering and Probabilistic Broadcasting in VANETs. *Veh. Commun.* **2018**, *13*, 78–88.
35. Zhang, X.; Cao, X.; Yan, L.; Sung, D. A street-centric opportunistic routing protocol based on link correlation for urban vanets. *IEEE Trans. Mob. Comput.* **2016**, *15*, 1586–1599.
36. Artimy, M.M.; Robertson, W.; Phillips, W.J. Assignment of dynamic transmission range based on estimation of vehicle density. In Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks, New York, NY, USA, 2 September 2005; pp. 40–48.
37. Yu, G.; Hu, J.; Zhang, C.; Zhuang, L.; Song, J. Short-term traffic flow forecasting based on Markov chain model. In Proceedings of IEEE IV2003 Intelligent Vehicles Symposium. Proceedings (Cat. No.03TH8683), Columbus, OH, USA, 9–11 June 2003; pp. 208–212.
38. Sommer, C. Veins The Open Source Vehicular Network Simulation Framework. Available online: <https://veins.car2x.org/> (accessed on 10 November 2018).
39. SUMO—Simulation of Urban Mobility, SUMO Official Web site, last modified 25 November 2018. Available online: <http://sumo.dlr.de/index.html> (accessed on 27 November 2018).

