CLUSTERING AND SELF ORGANIZED TOPOLOGY DEVELOPMENTS
TECHNIQUES FOR VEHICULAR AD-HOC NETWORK: A REVIEWS

Jiwa Abdullah¹, Hannes Masandig ²

¹jiwa@uthm.edu.my, ²hannes@uthm.edu.my

UNIVERSITY TUN HUSEIN ONN MALAYSIA

Abstract - Vehicular ad-hoc network does not rely on network infrastructure for its reliable communication; it requires well designs communication system, optimized and reliable protocols with efficient topology formations. This paper reviews clustering and self-organization techniques for vehicular ad-hoc network. Various clustering techniques have been investigated as well as studying the internal behaviour of the system and model them for the potential of introducing and developing of an intelligent rapid topology formation (IRTF) algorithm through advanced biologically intelligent system.

Keywords: Clustering, Self Organized, vehicular ad-hoc network, topology.

1 INTRODUCTION
Mobile Ad Hoc Networks (MANETs) are a class of infrastructure less network architecture which are formed by a collection of mobile nodes that communicate with each other using multi-hop wireless links. They eliminate the need for central management; hence each node must operate cooperatively to successfully maintain the network. Each node performs as a source, a sink and a router[1]. Typical application of MANET and VANET are includes rescues and military operations[2, 3].

Routing in wireless mobile ad-hoc networks should be time efficient and resource saving[4]. One approach to reduce traffic during the routing process is, to divide the network into clusters. Until now, there have been several approaches on cluster-based routing. Clustering methods allow fast connection and better routing and topology management of MANET (mobile ad hoc networks)[5]. The classification of the clustering techniques will be discussed in the later sections.

Routing protocol can be operated in unicast, multicast or geocasts. Unicast protocol is when one source transmits messages or data to one destination. Multicast is where one source is sending messages or data to several destinations through tree or mesh connections or topologies; and geocast is where one source is sending message or data to specific geographical location. Ad hoc network is a multi-hop wireless network, which consists of number of mobile nodes. MANET are networks which routing is based on multihop routing from a source to a destination node or nodes. These nodes generate traffic to be forwarded to some other nodes or a group of nodes. Due to a dynamic nature of ad hoc networks, traditional fixed network routing protocols are not practicable[6].

Scalable routing is one of the key challenges in designing and operating large scale Mobile Ad Hoc Networks (MANET) as well as vehicular ad hoc network (VANET). In order to ensure effective operation (since total number of nodes becomes larger and larger); the overhead of the employed routing algorithms should be low and independent of the total number of nodes[7]. Clustering provides a method to build and maintain hierarchical addresses in ad hoc networks [8].
The protocol in general can be according to figure 1[6]. The uniform type of the single channel work can be of either topology based or destination based. Whereas the non-uniform type are of neighbour selection or partitioning. The topology base protocol are like Global State Routing (GSR)[9-11] and Destination State Routing (DSR)[12, 13]. The destination based routing protocol on the other hand are like DSDV[14, 15], AODV[16], TORA[17], ABR[18], and WRP[19].

![Classification of Ad Hoc Network Protocol](image)

And the single and non-uniform channels were the ZRP[20], FSR[21], OLSR[22], CEDAR[23] and CBRP[24]. Other protocol are CGSR[25] and Epidemic[26] falls into unicast multi channel protocol categories.

The goal of this work is, to give a description of the cluster-based routing protocol; point out its advantages; investigates and comparing clustering routing protocols and to discuss challenges when deploying cluster-based routing. This paper organized in a way that it introduces the background of Mobile Ad Hoc Network (MANET) and Vehicular Ad hoc Network (VANET) at the introduction section. It follows by the related work on topology development and clustering studies for ad hoc routing protocol. Section 3 discusses about topology abstraction work designs approach proposal and followed by the Clustering algorithm designs of the system; and close by the conclusion.

2 CLUSTERING SYSTEM MODELLING

Clustering is a process that divides network into interconnected substructures, called clusters. This clustering approach is to build hierarchies of nodes; such that the network topology can be abstracted. The process is commonly referred to as clustering and the substructures that are collapsed in higher levels are called clusters. Each cluster has a cluster head (CH) as coordinator within the substructure. Each CH acts as a temporary base station within its zone or cluster and communicates with other CHs. Among the many challenges for ad hoc network is its scalability. When a flat topology network contains a large number of nodes, and control overhead, such as routing packets; its requires a large percentage of the limited wireless bandwidth.[8]

2.1 Cluster Modelling Scheme

System Model Scenario

Figure 2 shows the physical location of a mobile nodes before clustering while in figure 3 shows mobile (or vehicular) node structures after being clustered. One assumption we take is that each node capable of forwarding data intended for other nodes, passing data to next nodes until its reach the destination. This will simplifies the system for comparison assessments of clustering approach studies.
System Model and Assumptions
Clustering problem can be defined as undirected graph \( G = (V, E) \) [8, 27]; \( V \) representing a communication network where \( G \) is the vertices which is nodes in the network and \( E \) is the edges are the communication links. The clustering process divides \( V \) into a collection of subsets \( \{V_1, V_2, \ldots, V_k\} \) which not necessarily disjoint; \( V = \bigcup_{i=1}^{k} V_i \) such that each subset \( V_i \) induces a connected subgraph of \( G \).

Modelling Issues
To obtain better system hierarchy structure; four main issues shall be investigated: (1) the optimal ratio of the number of clusterheads (CH) and cluster members (CM) (CH versus CM); (2) the optimal hops between CMs and associated CHs (3) the dependency probability of CM where the probability of CM belongs to particular clusters and (4) the link stability between two nodes[7].

Figure 2: Physical location of nodes

The following are assumption also considered in modelling a system: (a) the wireless system is distributed on a flat two-dimension ground without any obstruction; (b) each node has a geographical position enable devices that their position can be measure. (c) the node uniquely identified by MAC address and (d) have the same transmission range.

Figure 3: Node clustering; before 3(a) and after 3(b)
**System Parameter**
The parameters also taken into considerations for optimization are; the number of CHs, \( n \); the optimal hops \( h \); dependency probability \( P_d \); and the links stability \( P_s \).

**Cluster Head selections**
The clusterhead selection can be in accordance to the following weighted \( W \) scheme[3, 28]:

\[
W_v = w_1 \Delta_v + w_2 D_v + w_3 M_v + w_4 P_v
\]  

The combination of weighing factor designs such that; \( w_1 + w_2 + w_3 + w_4 = 1 \). \( \Delta_v \) is the degree of difference obtained by calculating the number of neighbours of each node. The result of calculation is defined the degree of a node \( v \) and \( d_v \). The \( D_v \) is the sum of distances from a given node to all its neighbours. This is the energy consumptions; since farer neighbour requires more energy for communications. The \( M_v \) is the measure of mobility and \( P_v \) is cumulative time of the node being the clusterhead. Razaee [28] on the other hand uses the following form of equation as opposed to Roberto [3];

\[
W = aN + bR + cT + dP
\]  

Where \( N \) is the power level of the last received nodes; \( R \) is the percentage remaining battery life ratio of the node; \( T \) is the cumulative time during in which the node had been in the cluster and \( P \) is the transmission power which is used to be the major factors of the clusterhead selection criteria.

**Route Discovery**
Route discovery is done by using source routing. In the Cluster based routing protocol (CBRP) only clusterheads are flooded with route request package (RREQ). Gateway nodes receive the RREQs as well, but without broadcasting them. They forward them to the next clusterhead. This strategy reduces the network traffic. Initially, node S broadcasts a RREQ with unique ID containing the destination’s address, the neighbouring clusterhead(s)—including the gateway nodes to reach them and the cluster address list which consists of the addresses of the clusterheads forming the route [29]. Table 1 shows the algorithm of node \( n \) receiving a RREQ [4].

**2.2 Classification of Clustering Scheme**
Clustering are grouped into six different categories [30]. They are DS based[31], Low Maintenance Scheme[32], Mobility Aware[33], Energy Efficient[34], Load Balancing[35] and Combine Metrics based[36].

**3 TOPOLOGICAL ABSTRACTIONS AND HIERARCHICAL STRUCTURE**

**Topology structure formation:**
Some works related to topology hierarchical structure formation establishment abstraction have been studied by several researchers like Zhou et. al and Khun et. al [37, 38].
4 CLUSTERING ALGORITHM SCHEME

Clustering can be model as a graph partitioning problem. The key problem is on the identification of the virtual backbone network node assignments and deployments. However this could be solved with simple solution where VBN node are to be pre-assigned [39].

Table 1: Clustering Algorithm (simplified)

<table>
<thead>
<tr>
<th>Clustering Algorithm stages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1: Clusterhead selections</td>
</tr>
<tr>
<td>Step 2: Clustermember selections</td>
</tr>
<tr>
<td>Step 3: Advanced Clustering optimization – (using advanced computational algorithms; like Genetic algorithm [40], Simulated annealing[41], Particle swarm intelligent [42-44] and several other methods)</td>
</tr>
</tbody>
</table>

Comparison between clustering scheme:
Table 3 and 4 shows the comparison of the clustering scheme objectives and costs while clustering implementation.

5 QUALITY OF SERVICE AND NODE CONNECTION INDEX

MANET or VANET quality of service (QoS) can be modeled as $G = (E, Q(n_{ci}, B_{as}, D_{ase}, D_{mac}))$ where $E$ is a set of all mobile nodes in the network; and $Q$ is a set of QoS routing constraints which set the limits on the performance of the network. Each mobile node $i \in E$ has a unique identity and moves arbitrarily[1]. A circular plane, radius $R$ defines a coverage area within which each node could communicate with each other directly. Neighbours of node $i$ are defined as a set of nodes which are within radius $R$ and reachable directly from the node $i$. Every pair of neighbours can communicate with each other in both directions. The two host are considered neighbours if and only if their geographic distance is less than $R$ [3]. Hence, there exists a connectivity between neighbours $i$ and $j$ with the node connectivity index ($n_{ci}$).

If the pairs are moving towards each other or away from each other, the node pair connectivity index, $n_{ci}$ should be a positive value which describes the quality of connectivity between any two adjacent nodes. The least $n_{ci}$ value indicates a good quality connection, in which the node pair connectivity time is larger compared to high $n_{ci}$ value The node connectivity index, $n_{ci}$ is defined as:

$$n_{ci} = \begin{cases} 
  d - \left[ \frac{10^4 b}{10^3 c - npem} \right] & \text{for } P_2 < P_1 \\
  \frac{10^4 b}{10^3 c + npem} & \text{for } P_2 > P_1 \\
  0 & \text{for } P_2 = P_1 
\end{cases}$$

(3)

Where $npem$ is when node is moving towards each other whereas $npem$ is when nodes moving away from each other. This can be defined as:

$$npem = \frac{1}{(1/t_2 - t_1)} \sqrt{((1/P_2) - (1/P_1))}$$

(4)
and

\[ npcm = \left( \frac{1}{(t_2 - t_1)} \right) \left( \frac{1}{\sqrt{P_2}} - \frac{1}{\sqrt{P_1}} \right) \]  \hspace{1cm} (5)

The routing process takes place while transmission and connectivity is occurs. The routing can be models as R without intermediary node self-looping. The routing shall be in a single path interconnection between source to destinations. The Routing can be models such that:

\[ R(s, t) = \{s, ..., i, j, k, l, ..., t\} \]  \hspace{1cm} (6)

The node connectivity index (nci) metric can be defined as C :

\[
C = \begin{bmatrix}
nci_{0,0} & nci_{0,1} & \cdots & nci_{0,k-1} \\
ncki_{1,0} & ncki_{1,1} & \cdots & ncki_{1,k-1} \\
\vdots & \vdots & \ddots & \vdots \\
ncki_{k-1,0} & ncki_{k-1,1} & \cdots & ncki_{k-1,k-1}
\end{bmatrix}
\]  \hspace{1cm} (7)

Where k is number of node in the sub-cluster. Note that k is always greater than the interconnection edge of the nodes of one degree.

6 CONCLUSIONS

In this paper we addressed the clustering and self organized management in a mobile and vehicular ad hoc network. We outlined clustering technique scheme for topology establishment which improved communication, connection and data transmission between nodes. The benefit of introducing hierarchical structure of mobile nodes has been realized through this study especially for data communication of two or more nodes of both unicast and multicast environments and result of the proposed scheme showed obvious improvement. Thus; it is prooven that the system with self organized clustering for topological establishment approach is better than the mobile system without clustering approach communication protocol. Therefore further study of the system and its application through advanced optimization and intelligent study is feasible.

REFERENCES

29. Bansal, R.K. PERFORMANCE ANALYSIS OF CLUSTER BASED ROUTING PROTOCOL. in Computer. 2006: Thapar Institute of Engineering and Technology


**APPENDIX**

**Table 2: Route discovery algorithm**

```plaintext
IF N is member
    IF D is in the neighbour table
        send RREQ to D
    ELSE IF N is gateway to clusterhead C
        forward RREQ to C
    ELSE
        discard RREQ
    ENDIF
ELSE IF N is clusterhead
    IF RREQ already seen
        discard RREQ
    ELSE
        record ID in cluster address list of RREQ
        IF D is neighbour OR D is two hops away
            send RREQ to D
        ELSE
            FOR EACH neighbouring clusterhead C DO
                IF NOT C in address list of RREQ
                    record C in cluster address list of RREQ
                ENDIF
            ENDFOR
        ENDIF
    ENDIF
broadcast RREQ
ENDIF
```

ENDIF
Table 3: Clusters scheme

<table>
<thead>
<tr>
<th>Cluster scheme</th>
<th>Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS based</td>
<td>Reduce number of participating nodes in route search (or routing table maintenance)</td>
</tr>
<tr>
<td>Low Maintenance Scheme</td>
<td>Providing a cluster infrastructure for upper layer of applications</td>
</tr>
<tr>
<td>Mobility Aware</td>
<td>Use mobility behavior for cluster construction and maintenance</td>
</tr>
<tr>
<td>Energy Efficient</td>
<td>Avoid unnecessary energy consumption (or load balancing)</td>
</tr>
<tr>
<td>Load Balancing</td>
<td>Distribute the workload of a network into clusters</td>
</tr>
<tr>
<td>Combine Metrics based</td>
<td>Using cluster metrics configuration to determining the weighing factor in different applications</td>
</tr>
</tbody>
</table>

Table 4: Cost for Clustering

<table>
<thead>
<tr>
<th>Cluster scheme</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit control</td>
<td>Explicit clustering information exchanges between nodes</td>
</tr>
<tr>
<td>Re-clustering ripple effects</td>
<td>Relocation of clustering node will effects the whole clusters</td>
</tr>
<tr>
<td>Cluster formation (stationary assumptions)</td>
<td>Stractly assumption of nodes when clustering is to be forms- otherwise clusters will not be initiated/forms</td>
</tr>
<tr>
<td>Constant computation cycle</td>
<td>Number of cycles that the cluster formation are completed</td>
</tr>
<tr>
<td>Communication message complexity</td>
<td>Cluster formation and reformation requires exchanges of communication messages; the initial clusters formation and the reformation time may be differs</td>
</tr>
</tbody>
</table>