Novel Clustering Method Based on K-Medoids and Mobility Metric

Y. Hamzaoui[1], M. Amnai[2], A. Choukri[3], Y. Fakhri[1]

[1] LARIT, Networks and Telecommunications Team, Faculty of Sciences Kenitra (Morocco)
[2] Hassan 1 University, Laboratory IPOSI National School of Applied Sciences (ENSA), 26000 Settat (Morocco)
[3] Cadi Ayyad University, S.A.R.S Group, ENSA Safi (Morocco)

Received 2 June 2017 | Accepted 3 October 2017 | Published 3 November 2017

Abstract

The structure and constraint of MANETS influence negatively the performance of QoS, moreover the main routing protocols proposed generally operate in flat routing. Hence, this structure gives the bad results of QoS when the network becomes larger and denser. To solve this problem we use one of the most popular methods named clustering. The present paper comes within the frameworks of research to improve the QoS in MANETs. In this paper we propose a new algorithm of clustering based on the new mobility metric and K-Medoid to distribute the nodes into several clusters. Intuitively our algorithm can give good results in terms of stability of the cluster, and can also extend life time of cluster head.

Keywords

MANETs, Clustering, Mobility, K-medoids.

DOI: 10.9781/ijimai.2017.11.001

I. Introduction

Ad hoc wireless networks (MANETs) are a new architecture of wireless networks, where a set of mobiles can be deployed easily in a minimum configuration, with free movement and without the existence of the centralized device. It allows a user with a terminal to access to information at any time and from any place. These mobiles cooperate with each other to overcome the Ad hoc network constraints such as dynamic topology, lack of centralized monitoring points, limited bandwidth, etc. The non-centralized management and dynamic topology in Manets require nodes (mobiles) to behave as routers in order to maintain routing information. Many routing protocols have been proposed [1] on different types of application. The research has not ceased to have efficient protocols that adapt to all mobility models [2]. The routing of the information in the Manets can be classified into two types: Flat and Hierarchical. In the first type all network nodes play the same role, this can overload the network, as well as cause other problems such as scalability and complexity, when the network becomes wider and denser. The second type of routing is used to support networks that are wide and dense. The clustering has a hierarchical structure that makes it possible to group geographically the nodes that are neighbors. This allows to each node storing all information about its group and only some information of other groups (clusters). This approach can reduce the cost of routing of information in large and dense networks. Research has not ceased to have efficient protocols that can support this type of network and structure. In order to achieve this goal, researchers propose various metrics for building and organizing clusters in Manets. These metrics will be the basis for the clustering construct algorithms integrated in the routing protocols in order to optimize and improve them. The mobility metric allows to study the motion of the nodes, it allows to quantify the dynamic topology of the nodes in the Manets. A good metric of mobility makes it possible to better differentiate between the models of mobility [2], and to reflect the real behavior of the nodes. As well as facilitating the performance improvement of protocols.

* Corresponding author.

E-mail addresses: Younes_ieee@hotmail.com (Y. Hamzaoui), mohamed.amnai@uhp.ac.ma (M. Amnai), choukriali@gmail.com (A. Choukri), akhri-youssef@univ-ibntofail.ac.ma (Y. Fakhri).

In this article we propose a new stability metric that refines the degree of mobility, taking into account any type of motion in a coverage area of a mobile node, this metric will be the basis of a cluster building algorithm by using k-medoid to create the groups of mobile, this new approach can generate the more stable cluster and cluster head. In the rest of this paper we start with a presentation of some related work. In the second section we present the problem formulation, and in the third section we define the clustering. Then in the fourth section we present the definition of the distance, in the fifth section we explain our mobility metric, after in the sixth section we present the K-medoid algorithm before we propose the clustering algorithm in seventh section. In eighth section we present the conclusions.

II. Related Work

Routing is among the most important processes in Manets, their performance is related to the density and mobility of the nodes. Several clustering algorithms have been designed and proposed, that are based on different parameters to organize the network into several groups with their cluster head. In this section we will present some interesting works based on the creation of clusters by the use of density and mobility metrics.

In [3] a new approach is described, it is implemented into OLSR (Optimized Link State Routing) protocol to build a cluster and elect a clusters head, it is based on the calculation of the less mobile node, hence the nodes are affected by the less mobile node. According to the results, this algorithm made a great improvement of cluster stability compared with other density approaches.

A novel clustering algorithm for OLSR is proposed in [4], it gives birth to a new OLSR protocol, based on a metric that combines the two major characteristics of ad hoc networks: mobility and density. The simulations results show, that the proposed approach enhances the routing process and produces a small number of stable (less mobile) cluster heads.

The improvement of an existent study is the goal of authors in [5], it defines a new metric measuring of degree of mobility (for each node) based on incoming and outgoing nodes from a coverage area. The others use this metric to select the MPRs (Multipoint Relays) of OLSR protocol, the new OLSR protocol named Mob2-OLSR is compared with standard OLSR and Mob-OLSR, the results of simulation shows the amelioration of the performance of the Mob2-OLSR compared with others protocols.

In [6], the authors propose a novel energy aware clustering algorithm for the optimized link state routing (OLSR) protocol. This algorithm takes into account the node density and mobility and gives major improvements regarding the number of elected cluster heads and increase the network lifetime.

III. Problem Formulation

Several routing protocols ensure a high level of quality of service in Manets networks characterized by low density and low mobility, however when the network becomes denser [7] or more agitated [8], the network generates more control message, routing table, etc. Consequently the network becomes more sensitive, and it doesn’t ensure a minimum quality of service. Several studies on routing optimization adopt the clustering method to reduce the costs produced by the MANETs structures. It is based on the distribution of mobiles nodes into groups. In the literature we can have several clustering techniques; the K means clustering technique is among the best methods known [9] used in MANETs. This method is sensitive to outliers and the center of a cluster by calculating k-Means clustering can be empty. Thus, the extreme objects that have the upper values can significantly distort the distribution of clusters and their centers [10]. However these weaknesses can be avoided, if we apply a more robust method than K-Means [11] [12]. This one will allow clusters to be more stable and the center will be better chosen. In this paper we propose the use of K-medoid to produce optimal clusters. This method is based on a metric of mobility proposed by [4] and improved by our approach. In [4] the author has noticed the impact of nodes that join and / or leave the coverage area of a node on the evaluation of their degree of mobility (see Fig. 1(b)). Consequently he proposed the formula (2) which calculates the degree of mobility of a node. This brings relevant results in terms of QoS. Intuitively we added to this formula other measurement parameters that can better reflect the degree of stability of a node. These parameters are: The compute of the nodes that converge and that diverge during a determined duration of time (see Fig.1 (b), (c)). The stability of a node is not enough to elect it as a cluster head. Thus, we will add another parameter that will guarantee a longer lifetime of the cluster head. This parameter is the residual energy of a node, so that a stable node is the one that ensures the performance of these two variables (energy, stability).

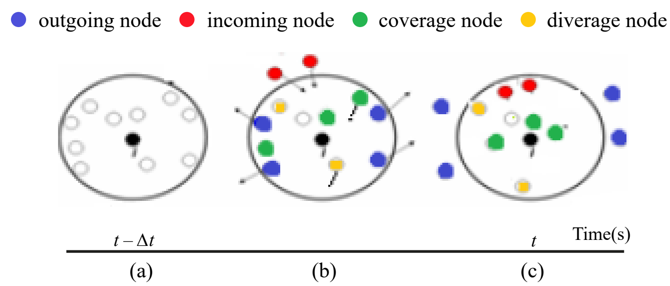


Fig 1. Illustration of movement inside the coverage area of a node.

IV. Clustering

Spatial clustering algorithms can be classified into four categories. There are the partition based, the hierarchical based, the density based and the grid based [13], [14]. According to [15] clustering in ad hoc networks can be defined as a theoretical arrangement of dynamic nodes corresponding to one or more specific properties in different subsets called “Cluster”. An element of a cluster is characterized by a strong similarity to components of its group, and a strong dissimilarity with respect to members of other groups [16]. Each cluster is identified by a particular node called “Cluster head”. Clustering allows a node to store only part rather than all the information of the network topology. This simplifies the processing of the global topology [17]. This reduces the size of routing tables and thereafter the reduction of the control messages generated by the routing system.

The use of clustering in Manets has several advantages [3], usually a cluster structure allows the node to play one of three roles (Fig. 2):

- Cluster head: A cluster head is elected in the cluster formation process for each cluster. Each cluster should have one and only one cluster head.
- Gateway : A node is called a gateway node of a cluster if it knows that it has a bidirectional or uni-directional link to a node from another cluster.
- Members: All nodes within a cluster except the cluster head are called members of this cluster.

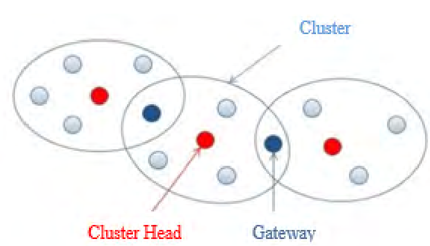


Fig 2. Node types in a cluster.

V. Distance

The choice of the distance is a key issue for classification methods. To offer a relevant measure of similarity between elements, it is necessary to well use the available information at the nodes. The Minkowski distance is the most used to determine the similarity between elements [18]:

$$d(x_i, x_j) = \left(\sum_{k=1}^{P} \left|v_k(x_i) - v_k(x_j)\right|^l\right)^{1/l} \tag{1}$$

Where $v_k(x_i)$ is the value of the object x_i on the variable v_k. Depending on the values taken by the parameter *l*, we talk about:

- Euclidean distance ($l = 2$);
- Manhattan distance ($l = 1$);
- Chebychev distance ($l = \infty$).

We note that the metrics used to analyze the ad hoc performances such as density, mobility and energy can be used to express distance.

VI. Novel Metric of Stability

The metric that we propose helps to calculate the degree of mobility (stability) of a node; this metric can be used in the calculation of the MPR (case of OLSR protocol) in selecting stable routes (case of AODV protocol) or in clustering process. In our proposal we introduce other parameters as the number of nodes incoming and outgoing in the area coverage of node studied (see Fig. 1). These parameters are the calculation of the numbers of nodes converging and diverging in the coverage area of the node studied.

In [4], only incoming and outgoing nodes in the radio range of the studied node are used to calculate the degree of mobility in the following formula (2):

$$M_i^{\lambda}(t) = \lambda \frac{\text{NodesOut}(t)}{\text{Nodes}(t-\Delta t)} + (1-\lambda)\frac{\text{NodesIn}(t)}{\text{Nodes}(t)} \quad (2)$$

In our proposed approach, we noticed that in a scenario of movement (see Fig.1 (a)), it may be that more nodes move towards the studied node. The latter will be as a gravity point (convergence point). Thus these movements towards it will give them stability over time and vice versa (in case of divergence). While we have not only limited on incoming and outgoing nodes, on the other movement being within the radio range of the studied node, we try to refine the calculation of the stability metric of a node.

We selected four types of movements that have influence on the stability or mobility of a node; we find the node that leaves the coverage area, the node that joins the zone, the node that approaches to the studied node and the node that finally moves away from the examined node and that stays in their coverage area. The first two parameters will be retained by the collected control messages. The last two will be calculated by the calculating power for two successive received messages (eg Hello message). We define the following parameters that characterize our metric:

- N_{con} defines the number of nodes that converge on the studied node.
- N_{div} defines the number of nodes that diverge towards the outside of the studied node.
- N_{in} defines the number of nodes within the area of the studied node.
- N_{out} defines the number of nodes out of the coverage area of the studied node.

Following these four types of movement we have created a metric that will rank the nodes between any of these four metrics of stability. We define each as shown in Table I:

TABLE I. Classifying of Metrics

Metric 1: Better Stability	$\begin{cases} Ncon > Ndiv \\ Nin > Nout \end{cases}$	$\gamma \in [0\ 0.25[$
Metric 2: Stable	$\begin{cases} Ncon < Ndiv \\ Nin > Nout \end{cases}$	$\gamma \in [0.25\ 0.5]$
Metric 3: Less stable	$\begin{cases} Ncon > Ndiv \\ Nin < Nout \end{cases}$	$\gamma \in]0.5\ 0.75[$
Metric 4: Poor Stability	$\begin{cases} Ncon < Ndiv \\ Nin < Nout \end{cases}$	$\gamma \in [0.75\ \ 1]$

In the four classifications, we have the first case that reflects a better stability for the node in question and the latest which represents a poor stability of the studied node, intuitively the second is better than the third, because N_{in} is greater than N_{out} and secondly, even with $N_{div} > N_{con}$ the diverging node stays in the coverage area of the studied node.

We determine subsequently the metric degree of stability that will calculate for each category the best stability node, we use in the formula (3) the coefficient of flow defined in [14], we divide the coefficient in 4 intervals (Table I) and metric degree of stability of node i will be as in (3):

$$M_t^{\lambda}(t) = \lambda\left(\frac{N_{div}}{N_{con}+N_{div}+N_{out}} + \frac{N_{out}}{N_{con}+N_{div}+N_{out}}\right) + (1-\lambda)\left(\frac{N_{con}}{N_{con}+N_{div}+N_{in}} + \frac{N_{in}}{N_{con}+N_{div}+N_{in}}\right) \quad (3)$$

VII. The Description of K-medoid Algorithm

K-Medoid is a partitioning technique of clustering of object into k clusters, each cluster is presented by a Medoid, it is the most optimal located object in a cluster. The PAM (Partioning Around Medoids) is the first K-Medoid algorithm introduced [19]. Initially, the k number of desired clusters is an input and a set of k nodes is taken randomly to be the initial representative medoid of k clusters. The final medoid (object) calculated by PAM is the most centralized position of all objects in a cluster. Thus the PAM algorithm examines in each step, all nodes (one by one) from the input dataset (nodes) that are not currently a medoid and see if they should be one. That is, the algorithm determines whether there is a node that should replace one of the existing medoids to minimize the total error (4). A node is assigned to the cluster represented by the medoid to which it is closest (minimum distance). The PAM algorithm is shown in Fig. 3. We assume that K_i is the cluster represented by medoid t_i. Suppose t_i is a current medoid and we wish to determine whether it should be exchanged with a non-medoid t_h. We wish to do this swap only if the overall impact to the cost (sum of the distances to the cluster medoid) represents an improvement. The total error by a medoid change S_{ih} is given by (4):

$$S = \sum_{h=1}^{k} \sum_{n_i \in C_h} dis(n_h, n_i) \quad (4)$$

S is the sum of absolute error for all objects in the data set (nodes).

```
Algorithm of PAM
Input
D={t1, t2,.........,tm) // set of nodes
K // Number of desired clusters
Arbitrarily select k medoids from D;
   Repeat
      For each th not a medoid do
         For each medoid ti do
            Compute square error function Sih;
               Find i,h where Sih is smallest;
               If  Sih< current Sih then;
                  Replace medoid ti with th;
   Until Sih>=current Sih;
   For each ti ∈D do
Assign ti to Kj where dist(ti, tj) is the smallest over all medoids.
```

Fig 3. Algorithm of PAM

VIII. Novel Approach of Clustering

The introduction of changes to the standard algorithm is a need, to use the k-medoid method for grouping nodes of MANETs into clusters;

this change aims to determine the K parameters of the algorithm. Firstly it is assumed that each node in Manets is an own cluster; then followed by a sequence of ascending partitions, in the end we reunite the nodes from the same neighborhood in the same cluster. Until reaching a final number of clusters. Thereafter, the K-Medoid algorithm will be used to generate more stable clusters with their cluster head. The parameters used in the medoid calculation algorithm (PAM) are the stability and the residual energy of a mobile, so each node is identified by a vector (stability, energy). This vector will be the basis for expressing the distance between nodes.

Fig. 4 presents our algorithm based on K-medoid to create the partition of the cluster nodes in Manets.

Proposed algorithm
Input: Mobile ad hoc network of n nodes.
Output: Network virtually partitioned into P clusters
Step 0

1. Initialization with n medoids $(M_1^0 \ldots M_n^0)$ each node is a cluster
2. Creation of an initial partition $P_0 = \{(C_1^0 \ldots . C_k^0)\}$

a. Initialize *l* to 1 (l *is iteration index*)

b. Assign to M_l^0 its two-hop neighbors ;

$$C_l^0 = \{x_i \in \text{network} \mid d(x_i, M_l^0)\};$$

c. Remove from list of medoids the $n_{C_l^0}$ nodes assigned to medoid M_l^0;

d. Move to the medoid $l + n_{C_l^0}$

e. Repeating steps b to d until all node are affected ;

f. Calculation of new medoids of k cluster obtained $(M_1^1 \ldots M_k^1)$: using the PAM algorithm;

Step t

3. Creation of a new partition $Pt = \{(C_1^t \ldots . C_k^t)\}$ by assigning to each medoid its two-hop neighbors;
4. the medoids affected by other medoids are removed from the list of medoids;
5. the isolated medoid are assigned to the list of medoids ;
6. Calculate the medoid of k clusters obtained $(M_1^0 \ldots M_k^0)$: using PAM algorithm ;
7. Repeat steps 3 to 6 until that a stable partition is achieved (structure of partition P_{t+i} equals that the P_{t+i+1}) or reach n iterations.

Fig 4. Proposed Algorithm of Clustering Method.

IX. Conclusion

In order to reduce the routing information costs and increase the QoS in MANETs, we have used the strong method to group the nodes into several clusters. In this paper we have proposed a novel approach of clustering based in K-medoid, and using the new metric of the calculation of degree of mobility. In future work we validate our approach by their implementation in protocol routing. Consequently, this can improve the QoS in MANETs.

References

[1] A. Boukerche, B. Turgut, N. Aydin, M.Z. Ahmad, L. Bölöni, and D. Turgut. Routing protocols in ad hoc networks: a survey. Computer Networks, Vol. 55, No. 13, pp. 3032–3080, September 2011.

[2] These de Yawut en (2009). Titre : Adaptation à la mobilité dans les réseaux ad hoc, institut de Recherche en Informatique de Toulouse.

[3] M. Dyabi, M. Saadoune, A. Hajami, H. Allali. OLSR Clustering Algorithm Based on Nodes Mobility, International Review on Computer and Software, (IRECOS), Vol. 10, No. 1, pp. 36-43, January 2015.

[4] A. Loutfi, M. Elkoutbi. Optimizing the process of OLSR clustering based on mobility and density in ad hoc networks. International Conference on Multimedia Computing and Systems (ICMCS), pp. 522 – 526, May 2012.

[5] N. Lakki, A. Ouacha, A. Habbani, M. Ajana el khadar, M. El Koutbi, J. El Abbadi. A new approach for mobility enhancement of olsr protocol. International Journal of Wireless & Mobile Networks (IJWMN) Vol. 4, No. 1, February 2012.

[6] A. Loutfi, M. Elkoutbi, J. BenOthman, A. Kobbane. An energy aware algorithm for olsr clustering. Annals of telecommunications, Vol. 69, No. 3, pp. 201–207, April 2014.

[7] Kwar, B.J., N.O. Song and L. Miller. On the scalability of ad hoc networks. IEEE Communication Letters, Vol. 8, No. 8, Aug. 2004.

[8] C. A. Santiváñez, B. McDonald, I. Stavrakakis, and R. Ramanathan. On the scalability of ad hoc routing protocols, in Proceedings of the IEEE Infocom, pp. 1688–1697, June 2002.

[9] A. M. Bagirov, J. Ugon, and D. Webb. Fast modified global k-means algorithm for incremental cluster construction, Pattern Recognition, Vol. 44, pp. 866-876.

[10] Jiawei Han, Micheline Kamber. Data Mining Concepts and Techniques, second edition, Elsevier Science, 2007.

[11] Norazam Arbin, Nur Suhailayani Suhaimi. Comparative Analysis between K-Means and K-Medoids for Statistical Clustering, Third International Conference on Artificial Intelligence, Modelling and Simulation, 2015.

[12] Noor Kamal Kaur, Usvir Kaur, Dheerendra Singh. K-Medoid Clustering Algorithm- A Review, International Journal of Computer Application and Technology (IJCAT) Vol. 1, No. 1, pp. 42-45, 2014.

[13] Y. Sun, Radio Network Planning for 2G and 3G, Master of Science in Communications Engineering, Munich University of Technology, 2004.

[14] R. Whitaker, S. Hurley. Evolution of planning for wireless communication systems, Proceedings of the 36th Annual Hawaii International Conference (HICSS'03), pp 296, January 2003.

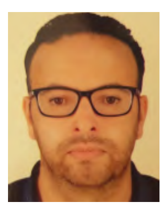
[15] Angione, G., Bellavista, P., Corradi, A. and Magistretti, E. A k-hop Clustering Protocol for Dense Mobile Ad-Hoc Networks. 26th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'06). Portugal, pp. 10, 2006.

[16] Daniel T. Larose. Discovering Knowledge in Data: An Introduction to Data Mining, John Wiley & Sons, Inc., Hoboken, New Jersey. 2005.

[17] Inn, E. and Winston, S. Mobility-Based D-Hop Clustering Algorithm for Mobile Ad Hoc Networks. IEEE Wireless Communications and Networking Conference, Vol. 4. Atlanta, GA, USA, Georgia, pp. 2359-2364, 2004.

[18] A. Choukri, A. Habbani, M. Elkoutbi. Classification par similarité: OLSRKmeans, WCCS14 The 2nd World Conference on Complex Systems, Agadir Morocco, 2014.

[19] L. Kaufman and P. J. Rousseeuw. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, 1990.



Younes Hamzaoui

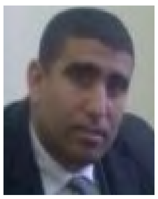
Younes Hamzaoui obtained his DEUG degree on chemical phsysics in 2000 from Abdmalik Saadi University, Tanger, Morocco. Then, he received his degree on IEEA (Computers, Electronics, Electrical and Automation) in 2003. After in 2007, he obtained his DESA degree on information system and telecommunication. The author is a professor of computer science in high school. He is a PhD student of management of QoS in routing protocol, also, he is doing his research in the Laboratory in Computer Science and Telecomunications (LaRIT) in the Faculty of Sciences of Kénitra.

Mohamed Amnai

Mohamed Amnai received his bachelor's degree in 2000, in IEEA (Computers, Electronics, Electrical and Automation) from Molay Ismail's University, Errachidia city. Then, he obtained his master's degree in 2007, from Ibn Tofail University, Kenitra city. In 2011, he received his PhD on Telecommunication and computer science, from Ibn Tofail University in Kenitra city, Morocco. Since March 2014 he is a Professor, he has been an Assistant at National School of Applied Sciences Khouribga, Settat University, Morocco. The author is also an associate member of Research Laboratory in Computer Science and Telecommunications (LARIT), Team Networks and Telecommunications Faculty of Science, Kenitra, Morocco. Also, the author is an associate member of laboratory IPOSI National School of Applied Sciences, Hassan 1 University, Khouribga, Morocco.

Ali Choukri

Ali Choukri is an Assistant professor at National School of Applied Sciences. He was awarded a Master of Computer Science and Telecom from the University of Ibn Tofail, Kénitra-Morocco in 2008. He got ENSET (higher Normal School of technical teaching) degree in 1992. He has a PhD from the School of Computer Science and Systems Analysis (ENSIAS). He is working within the MIS team of the Laboratory SIME, for studying ad hoc mobile intelligent communication systems, and wireless sensor networks. His research interests are in the areas of ubiquitous computing, Internet of Things, Delay/Disruption Tolerant Networks, Wireless Networks, QoS routing, Mathematical modeling and performance analysis of networks, Control and decision theory, Game theory, trust and reputation management, Distributed algorithms, Metaheurtistics and optimization, Genetic algorithms.



Youssef Fakhri

Youssef Fakhri received his Bachelor’s Degree (B.S) in Electronic Physics in 2001 and his Master’s Degree (DESA) in Computer and Telecommunication from the Faculty of Sciences, University Mohammed V, Rabat, Morocco, in 2003 where he developed his Master’s Project at the ICI Company, Morocco. He received a PhD in 2007 from the University Mohammed V - Agdal, Rabat, Morocco in collaboration with the Polytechnic University of Catalonia (UPC), Spain. He joined the Faculty of Sciences of Kénitra, Department of Computer Science and Mathematics, IbnTofail University, Morocco, as an Associate Professor in March 2009, he is the Laboratory head at LaRIT (Laboratory for Research in Computing and Telecommunications) in the Faculty of Kénitra, and Member of Pole of Competences STIC Morocco.