UAVs-assisted Data Collection in Vehicular Network

Mohamed Ben Brahim*†, Hakim Ghazzai*, Hamid Menouar*, Fethi Filali*
*Qatar Mobility Innovations Center (QMIC), Qatar University, Doha, Qatar
†HANA Research Lab, University of Manouba, Manouba, Tunisia
Email: {mohamedb, hakimg, hamidm, filali}@qmic.com

***Abstract*—In vehicular network, nodes generate and transmit timely measured data by embedded sensors. Some data needs to be gathered by remote devices. The limited communication range of the vehicular wireless system requires to proceed through multi-hop data routing to collect periodic fresh data. The hop-by-hop based data journey and the dynamic topology increase the overall packet delivery delay, as well as the packet loss ratio. Unmanned Aerial Vehicles (UAVs) could be used in relaying vehicular data, particularly for sparse networks. Once they join the network, they are considered as normal network nodes however with some specific link characteristics. In this paper, a link-aware data collection approach is investigated. More precisely, an optimization problem maximizing a weighted multi-objective utility including the wireless link data rate, the wireless link stability, and the data progress towards the destination is formulated and solved using a Distributed Minimum Spanning Forest (DMSF) approach. The outcomes of the proposed approach and the impact of the UAVs assistance are evaluated. The present approach outperforms the other algorithms and the usage UAVs enhances the DMSF-based solution especially for low-density network.**

Keywords–Unmanned aerial vehicles; Vehicular network; Data collection; Distributed minimum spanning forest.

I. INTRODUCTION

Vehicular networks are promoting a wide range of applications rendering the driving experience safer and more efficient. Indeed, using embedded communication capabilities, the vehicles are able to exchange timely and accurate information reflecting their status and mobility. Neighbor vehicles use the received information to mitigate collisions [1] and avoid road hazards [2]. Vehicles may disseminate traffic-related information in the relevant traffic stream to early warn the incoming travelers to seek alternative routes. An Intelligent Transport System (ITS) station is usually equipped with an On-Board Unit (OBU) which is an embedded computer interfacing with a multitude of measurement-units and sensors mounted in the vehicle, e.g., Global Positioning System (GPS), proximity ultra-sound and short-range radars, etc.

Collecting data in real-time in mobile network is challenging. Unlike data exchange or data dissemination, data collection's purpose is to route data towards a gathering node from all the network nodes. The achievement of this task should take into account the underlying network characteristics. Indeed, the vehicles are moving with relatively high-speed compared to legacy mobile ad hoc network (MANET). In addition, obstructing objects and signal attenuation due to mobility and radio interference impose a wise use of the network resources. Moreover, the height of vehicle-embedded antennas and the frequency bands used in vehicular network differentiate vehicular networks from cellular networks with infrastructure support [3]. These constraints require an optimized design of communication protocols to transfer data between mobile nodes and to ensure reliable data gathering by remote infrastructure nodes.

The UAVs have recently seen an increasing usage in civilian and military applications [4]. This trend is explained by the ease of deployment and the low cost of maintenance. Among the promising applications, the UAV may contribute in the monitoring and assistance of the traffic in ITS [5]. Indeed, the drone could be equipped with wireless communication system and participates in relaying traffic data or generating traffic alerts to warn ground road users, as illustrated in Figure 1. Worthy to mention here that the energy-related issues of the drones are intentionally skipped and the aerial nodes are considered as simply hovering over the road junction at a fixed altitude. the mobility of UAVs is planned to be studied in a future work.

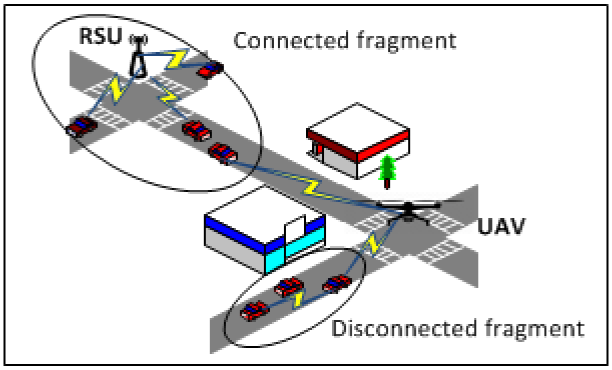

Figure 1. UAV relays data between two disconnected network fragments.

The present work proposes a new data collection approach using Distributed Minimum Spanning Forest (DMSF) algorithm: a modified version of the typical Distributed Minimum Spanning Tree (DMST) algorithm [6]. The proposed approach is based on two phases: the paths construction, and the effective data collection:

- The distributed routes construction phase proceeds with a parallel building of multiple trees having a minimum overall utility. These multiple trees enable the establishment of the routing paths and balance the load of the data traffic over the network. The utility considered in the construction of the DMSF is a weighted multi-objective function achieving a trade-off between three metrics: i) the achieved data rate over the wireless link, ii) the inter-vehicles link stability, and iii) the progress towards the destination. Hence, each node selects the relay node providing the best throughput for rapid data transfer, the highest link stability to avoid link disappearance mainly during big size packet transmission, and the closest route to the destination to reduce delays due to processing at the relay levels.

ISBN: 978-1-61208-643-9

- Once the data collection session is due, nodes transmit the received data from children nodes and their data content to their parent nodes in the tree-routing structures. The process is executed by each node once it has a data to send or to forward. The data is forwarded towards the roots of the trees. Since the roots are one-hop neighbors of the sink node, a simple data forwarding ensures the successful delivery of data. When all the data is collected by the sink or the collection session is expired, the non-delivered data is discarded and a new routing topology is calculated to start a new data collection round.

The proposed approach seeks to ensure periodic data collection generated by a big number of mobile nodes through a fast and distributed routes construction. This kind of periodic data gathering is required especially for network monitoring and analysis purposes. The performances of the proposed approach are evaluated against existing algorithms.

The remainder of this paper is organized as follows: Section II provides an overview of the existing work in data collection. Section III presents the system model. Section IV starts by presenting the methodology followed by the paper. Afterwards, it presents the problem formulation. Section V introduces the proposed DMSF algorithm to solve the data collection problem. The next section presents the simulation environment and some selected numerical results. Finally, Section VII concludes the paper.

II. RELATED WORK

For many use cases, we need to collect data from mobile nodes within the network by a collection host, aka sink. At the collection time, the traveling vehicles may or may not be close-enough to the collection node in order to deliver their data. Hence, several studies tackled this issue and investigated the possible ways to ensure data handing to the collection station. In [7], the authors presented a survey of the position-based data routing in vehicular networks. The position information availability and accuracy is a basic factor of the algorithm performances. The forwarding decision, which is usually greedy or an improved form of greedy, ensures a progress towards the destination node. Although judged to outperform topology-based algorithms, position-based routing algorithms present some limitations. Authors conclude that there is not a position-based protocol performing well in both urban and highway environments and compare the performance of the protocols mainly in terms of overhead, availability, resilience, and latency. In [8], authors leveraged a hybrid communication scheme using 3GPP/LTE and the IEEE 802.11p-based technologies to disseminate safety messages in Vehicular Ad hoc NETwork (VANET). The approach is based on the construction of multi-hop clusters and the offload of data transfer between farther clusters to the cellular network. The hybrid communication scheme outperforms the basic clustering and the flooding-based forwarding algorithms. Even though it resolves the hole problem especially in sparse networks, the proposed approach could not be the best option in the present scenario because no alternative communication technology other than the Dedicated Short Range Communication (DSRC) wireless technology is available.

Another approach for data collection based on a mobile agent is proposed in [9]. The proposed solution takes into account the lossy nature of the network links and uses mobile agents in the network to collect data from vehicles in one-hop fashion. The approach reaches a higher data collection ratio but is still limited in terms of data freshness, which is ensured by the present approach, and may be useful for special cases of deferred offline data processing scenarios. An earlier published algorithm, named ADOPEL [10], is proposed using the reinforcement learning technique for the data collection in VANET. The approach uses a distributed Q-learning technique to update Q-values required to select the relaying nodes to forward data towards the collection node. The algorithm is fully distributed and outperforms the non-learning approach. However, the fast topology updates negatively affect the convergence of the learning strategies. Indeed, the required time to learn the best relay is likely to exceed the link duration of all or some of the available links at a given time. In this paper, we overcome this issue in DMSF through periodic assessment and re-weighting of the wireless links. In [11], a data routing optimization approach based on multi-objective metrics and minimum spanning tree (MST) is proposed. The solution investigates an interference-aware routing scheme by fostering the better links and radio channels for routing the data. The used metrics to compute the link weight are end-to-end delay, link duration probability, and co-channel interference. The algorithm needs a global awareness of the network topology (e.g., convergence of routing tables) by every node to be able to compute paths to other nodes. This constraint limits the approach from scaling and from distributed construction of the spanning tree. However, this constraint is released in DMSF approach since only relay node selection is needed to forward data towards collection node.

III. SYSTEM MODEL

The nodes in vehicular networks are smart agents which interact with each other through direct ad hoc wireless links, i.e., V2V and occasionally with infrastructure or Road Side Unit (RSU), i.e., V2I. Mobile nodes mutually exchange status information which are useful for efficient mobility and traffic cooperative awareness. The focus of this paper is on the collection of data measured by the vehicles within an urban area. Therefore, we consider a geographical area with multiple roads and intersections. In the center of the area, an RSU is placed to regularly collect the data from N connected vehicles driving around. Every vehicle aims to periodically transfer a data of size M bytes to the RSU at a transmission power level P_{tr}. The obstacles and signal attenuation do not always allow direct communication between all the nodes and the RSU. Hence, a multi-hop data transfer is required.

To study the network behavior under different conditions and evaluate the ITS applications, several research studies investigated various channel propagation models of DSRC [12] [3]. In [3], authors carried out extensive measurement campaigns in different communication scenarios: the Line-of-Sight (LoS) and Non-LoS (NLoS) cases and for urban, suburban, highway, and rural areas characterized by different traffic densities. The resulting model is adopted for this paper, it is well-defined, and its parameters are explicitly given. By adopting the log-distance power law, the path loss expression is given by:

$$PL(d) = PL_0 + 10\nu \; log_{10} d + S, \qquad d_{\min} \leq d \leq d_{\max}, \quad (1)$$

where d is the distance between the transmitter and the receiver in meter and ν is the path loss exponent related to the propagation environment. The parameter S models a zero-mean random variable with normal distribution and standard deviation σ_S modeling the large-scale fading. The term PL_0 is given by:

$$PL_0 = PL_0(d_0) - 10\nu \; log_{10} d_0, \tag{2}$$

where d_0 is the reference distance and $PL_0(d_0)$ is the path loss value at the reference distance. Its value is given along with other parameters in Table I. This model is derived from measurement campaigns where $d_{\min}$ and $d_{\max}$ are bounding the model validity domain.

The communication link between the UAV and the ground nodes is modeled differently. The two nodes are in LoS with a certain probability P_r which depends on the UAV's altitude h_u [4]. P_r is given by:

$$P_r = \frac{1}{1 + Bexp(-C[\theta(h_u, d_{uv}) - B])}, \tag{3}$$

where B and C are environment-dependent constants, θ represents the elevation angle and is given by $\theta = \frac{180}{\pi} sin^{-1}(\frac{h_u}{d_{uv}})$, and d_{uv} denotes the Euclidean distance between the UAV and the ground node. The probability of having a NLoS link is equal to $1 - Pr$. Hence, the path loss expression in dB of the Air-to-Ground (A2G) link is given by [4]:

$$PL^{A2G}[dB] = P_r PL^{LoS} + (1 - P_r) PL^{NLoS}, \tag{4}$$

where PL^{LoS} is the path loss effect of LoS link and is expressed in dB as $PL^{LoS} = 10\gamma log_{10}(\frac{4\pi f d_{uv}}{C_l}) + L^{LoS}$, with γ is the path loss exponent, f is the frequency of the carrier, C_l is the speed of light, and L^{LoS} is an additional attenuation of the environment. The second term of the path loss PL^{NLoS} is the effect of NLoS link and is expressed in dB as $PL^{NLoS} = 10\gamma log_{10}(\frac{4\pi f d_{uv}}{C_l}) + L^{NLoS}$ with $L^{NLoS}$is an additional attenuation for the NLoS environment.

It's important to remind here that the coherence time of the channel is very small compared to the time of data collection and construction time of the tree, therefore, we investigate the channel based on the average statistics and we consider the path loss effect only.

IV. PROBLEM FORMULATION

The investigated problem can be modeled using graph theory where nodes are the graph nodes and wireless links are mapped to graph edges. At a given time, a generic undirected graph $G(V, E)$ with weighted edges results from a spatial distribution of mobile nodes. An edge matching two nodes is characterized with different attributes, namely *data rate*, *link stability*, and *closeness to sink*, which are quantified and detailed in the following subsection. The target of the next step is to leverage the link state of each node in order to compute a path from each node towards the sink in a distributed fashion. Indeed, getting information about the locations and states of all the nodes and the whole network topology in a centralized manner is an unpractical and unrealistic assumption mainly during short intervals of time. Therefore, we proceed with a distributed approach adapted to the network topology evolution. The constructed paths aim at enhancing packets delivery within a time-bounded delay while maintaining a fair exploitation of the network's resources.

A. Methodology and Metrics

The present paper focuses on the data collection task, and more specifically on a periodic data collection rounds triggered at a prefixed time period. Since every node is invited to append its own data to the received data, if available, before forwarding it towards the sink node, proceeding with broadcasting data overwhelms the network and may result in a *broadcast storm* especially in dense networks [13]. To avoid these harmful effects, every node operates through unicasting its data to a single next hop while taking into account the link states to maximize the chance of a time-bounded routing and successful delivery of packets.

To ensure a link-aware approach, three factors are considered while quantifying the weights of the wireless links. These properties are directly impacting the link quality and hence, its ability to deliver data without being distorted. Considering two nodes i and j of the network, the attributes of the link (i, j) are:

• Achieved throughput: Links with higher data rates are privileged in the construction of the data routing paths to accelerate the data collection procedure. Following is the achieved rate of link (i, j) based on Truncated Shannon Bounds (TSB) [14] and denoted by $R_{i,j}$:

$$R_{i,j} = \begin{cases} R_{max} & \text{if } SINR_{i,j} \geq SINR_{max}, \\ 0 & \text{if } SINR_{i,j} \leq SINR_{min}, \\ B \log_2 (1 + SINR_{i,j}) & \text{otherwise.} \end{cases} \tag{5}$$

where R_{max} is given by:

$$R_{max} = B \log_2 (1 + SINR_{max}), \tag{6}$$

where B is the system bandwidth, $SINR_{i,j}$ is the signal-to-noise-plus-interference ratio at the receiver of the link (i, j), given by:

$$SINR_{i,j} = \frac{P_{tr} H(\bar{d}(i,j))}{I + N}, \tag{7}$$

where P_{tr} denotes the transmit power level, H is the channel propagation model which is a function of average-distance between transmitter and receiver. It corresponds to the inverse of the path loss between the nodes i and j which is given in (1). I and N quantify the average interference and the noise power levels at each receiving node, respectively. Finally, $SINR_{min}$ and $SINR_{max}$ are the $SINR$ thresholds for the discretization of the data rate in DSRC technology. Their simulation values are given in Table I.

• Link stability: The relative high speed, dynamic topology, and the mobility patterns of the vehicular network impose constraints on the link stability. Hence, wireless link lifetime between two nodes is expected to persist for short time prior to be obstructed by static and/or mobile obstacles until it disappears from the available set of links. In the general case, the motion of the node in the network is hard to predict, even though some mobility patterns could be drawn in specific areas and during specific times. Hence, the position $(x_{i,t}, y_{i,t})$ of the node i at an instant t is given by:

$$(x_{i,t}, y_{i,t}) = f(x_{i,0}, y_{i,0}, t), \tag{8}$$

where f is a generic function taking as parameters the position of node i at instant $t = 0$ and the current time. Taking into account the locations, speed values, and driving directions of two given nodes at an instant t, and under the assumption that vehicles are generally driving into linear road segments with nearly constant speeds specifically for very short time-scale, the following mathematical linear system derives the expected link duration between two arbitrary vehicles i and j:

$$\begin{cases} x_{j,t} = v_{x,0}t + x_{j,0} & (9) \\ y_{j,t} = v_{y,0}t + y_{j,0} & (10) \end{cases}$$

where $x_{j,t}$ and $y_{j,t}$ are the coordinates of node j at an instant t, $x_{j,0}$ and $y_{j,0}$ are the reference coordinates of node j at $t = 0$, and $v_{x,0}$ and $v_{y,0}$ are the components of relative velocity of node j with respect to node i. With this setup, the coordinates of $node\ i$ are given by:

$$\begin{cases} x_{i,t} = x_{i,0} & (11) \\ y_{i,t} = y_{i,0} & (12) \end{cases}$$

where $x_{i,t}$ and $y_{i,t}$ are the coordinates of node i at an instant t, $x_{i,0}$ and $y_{i,0}$ are the reference coordinates of node i at $t = 0$. From (9)−(12), the euclidean distance between the nodes enables the determination of the expected link duration, denoted by LD and returns an important indicator about the radio link stability. Indeed, the resolution of the equation below, returns the LD:

$$LD_{i,j} = t, \text{when } \sqrt{(x_{i,t} - x_{j,t})^2 + (y_{i,t} - y_{j,t})^2} = D_{\max}, \tag{13}$$

where $D_{\max}$ is the maximum communication range between the nodes i and j, which depends on the channel path loss modeled in (1).

• Closeness to sink: The main goal is to gather the data from mobile nodes driving along a geographic area to a sink. In order to reach the destination node in a reasonable time, data should progress towards the sink at every hop. Otherwise, packets may spend a long time floating from a node to another without reaching the destination. To model the progress of data towards the destination, a metric characterizing the link aptitude to approach the sink is considered. Every node is labeled with a number of hops required to reach the collection point. If a node i selects the node j as its next hop, and the latter is $\bar{N}_h(j)$ away from the destination then, the created link has an advance attribute, denoted by $adv_{i,j}$, and is given by:

$$adv_{i,j} = \bar{N}_h(j) + 1, \text{if } i \text{ selects } j \text{ as next hop.} \tag{14}$$

Once the main factors impacting the link quality are defined, the links are labeled with weights resulting from a weighted combination of the aforementioned metrics. The best subset of the set of links maximizing the overall weights is used to connect the nodes with their respective successors to reach the collection node.

B. Problem formulation

Consider that each edge of the graph $G(V, E)$, that we denote by $e_k, k = 1, \dots, |E|$ where $|.|$ denotes the cardinality of a set, is characterized by a weight denoted by w_k. The objective is to divide the graph $G(V, E)$ into N_t disjoint trees to create a forest. This forest ensures a faster building of disjoint trees with shorter routes. The total number of trees N_t is depending essentially on the number of nodes that are within the coverage of the sink and characterized by a good link quality, e.g., $w_k \geq w_{th}$, where w_{th} is a link weight threshold defined by the operator. Define $T_t \equiv G(V_t, E_t) \subset G(V, E)$, where $t = 1, \dots, N_t$, a tree composed of a set of vertices denoted by V_t and a set of edges denoted by E_t. Accordingly, the forest will be formed by N_t trees. Denote by $\epsilon_{k,t}$ the binary variable indicating whether an edge k of E belongs to set of edges E_t. Hence, $\epsilon_{k,t}$ is expressed as:

$$\epsilon_{k,t} = \begin{cases} 1 & \text{if } \ e_k \in E_t, \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Consequently, the optimization problem aiming at minimizing a multi-objective utility composed of the three metrics discussed above can be formulated as:

$$\underset{\substack{\epsilon_{k,t}, k=1,\dots,|E| \\ t=1,\dots,N_t}}{\text{minimize}} \sum_{k=1}^{|E|} w_k \sum_{t=1}^{N_t} \epsilon_{k,t} \tag{16}$$

subject to:

$$\epsilon_{k,t} \leq \delta_k, \tag{17}$$

$$\sum_{t=1}^{N_t} \epsilon_{k,t} \leq 1. \tag{18}$$

The solution of the optimization problem given by (16)−(18) will associate the edges e_k to the N_t trees to connect the maximum of nodes that will forward their data packets to the roots of each tree such that the total weight is minimized. Using the metrics expressed in (5), (13), and (14), the weight of a given link joining two nodes i and j and corresponding to the edge $k \in E$ can be expressed as:

$$w_k \equiv w_{i,j} = -\alpha \frac{R_{i,j}}{R_{\max}} - \beta \frac{LD_{i,j}}{LD_{\max}} + \gamma \frac{adv_{i,j}}{adv_{\max}}, \tag{19}$$

where $R_{\max}$, $LD_{\max}$, and $adv_{\max}$ are determined by Monte Carlo simulation campaigns. They are used to normalize the magnitude of the different metrics. Note that the parameter adv_{max} represents the maximum possible number of hops to reach the sink node, which is equal to the number of nodes in the network. The parameters α, β, and γ are three freedom degrees identified by the operator. Their sum is equal to one and they have to be set properly according to the operator's need. For instance, increasing α will promote the rate metric at the expense of the others. Constraint (17) ensures that an edge e_k can be considered in the association if the corresponding $SINR$ provides at least the minimum requirement for a seamless transmission. This condition is identified by the binary parameter δ_k which is given as follows:

$$\delta_k = \begin{cases} 1 & \text{if } \ SINR_{i,j} \geq SINR_{\min}, \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

where $SINR_{i,j}$ is SINR at the receiver of the link k that can be deduced from (7) while $SINR_{\min}$ is a SINR minimum threshold. Finally, constraint (18) imposes that an edge can be associated to only one tree.

V. DATA COLLECTION WITH DMSF

The data collection is a specific case of more general routing problem. Indeed, for data gathering, the destination node is usually not an arbitrary node but is often the sink node having pre-defined public properties, e.g., its location is known by nodes belonging to the network. At every data collection round, the nodes have to check their link states to find the best relevant next hops in order to deliver their data content. Recall that there is no agent that has real-time global view of the network topology to execute a centralized computing of the best paths. Moreover, a given node does not have to be aware of all the network links to build its best path. Instead, it requires a local awareness of its links associating it with one-hop neighbor-nodes to make sure that its path to the sink node is loop-free. Finally, data sent by the distributed nodes should reach a subset of nodes, called the roots of the forest trees, which are responsible to relay data directly to the sink.

Algorithm V.1 Distributed Minimum Spanning Forest Algorithm

```
$V_R \leftarrow Best\_Sink\_Neighbors$. {The set of reliable neighbors of sink node (RSU)}
Update adjacency matrix.
$BSE \leftarrow All\_links$. {The set of basic edges}
$BRE \leftarrow \emptyset$. {The set of branch edges}
$RJE \leftarrow \emptyset$. {The set of rejected edges}
while $|| BSE || > 0$ do
  $WKN \leftarrow random\_subset(V)$.
  for $n \in WKN$ do
    $OE \leftarrow outgoing\_edges(fragment(n))$.
    $e \leftarrow \min(OE)$.
    $BRE \leftarrow BRE \cup \{e\}$.
    {fusion fragments and update ranks}
    $frag\_1 \leftarrow right\_fragment(e)$.
    $frag\_2 \leftarrow left\_fragment(e)$.
    if $frag\_1.root \notin V_R$ & $frag\_2.root \notin V_R$ then
      Use default DMST update.
    else
      if $frag\_1.root \in V_R$ & $frag\_2.root \notin V_R$ then
        $frag\_1 \leftarrow fuse(frag\_1, frag\_2)$. {frag_1 absorbed frag_2}
      else
        if $frag\_1.root \notin V_R$ & $frag\_2.root \in V_R$ then
          $frag\_2 \leftarrow fuse(frag\_2, frag\_1)$. {frag_2 absorbed frag_1}
        else
          continue. {frag_2 and frag_1 could not be fused}
        end if
      end if
    end if
    $BRE \leftarrow BRE \cup \{e\}$.
    Update $RJE$.
    $BSE \leftarrow BSE \setminus (\{e\} \cup REJ)$.
  end for
end while
```

Figure 2. DMSF algorithm.

The main idea of the proposed approach is to build multiple trees with minimum overall weights and rooted at a set of predefined nodes to create a Forest of N_t trees. We call this subset of nodes, i.e., the roots, as V_R where $|V_R| = N_t$. The tree-based topology of the collection algorithm focuses on ensuring a convergence of the collected data towards the destination and using loop-free paths. Moreover, each node can join a single tree and forward its own data and other received data exclusively to one *parent* node. Working with minimum spanning forest instead of a single minimum spanning tree is preferred because it leads to short paths to reach the destination. In addition, considering multiple roots allows the network to avoid bottlenecks in the last hops and ensures a load balance throughout the network.

Figure 2 illustrates a high-level description of the proposed DMSF algorithm. For a periodic data collection, where all the vehicles in a given geofence area are expected to transmit their generated and occasionally received data, nodes should first build routing paths especially when the exchanged data size is important. In order to build loop-free paths leading to the sink node, the algorithm starts with the selection of the members of V_R. Then, nodes initiate the construction of fragments based on the best available link weights and fuse fragments into bigger ones. The process continues in a fully distributed way until all the nodes, reachable from at least one of the V_R members, join a tree of the forest. It may happen that some nodes do not join any of the forest's trees if the network graph is physically disconnected. Hence, these nodes will not transmit data until joining routing trees to save communication bandwidth.

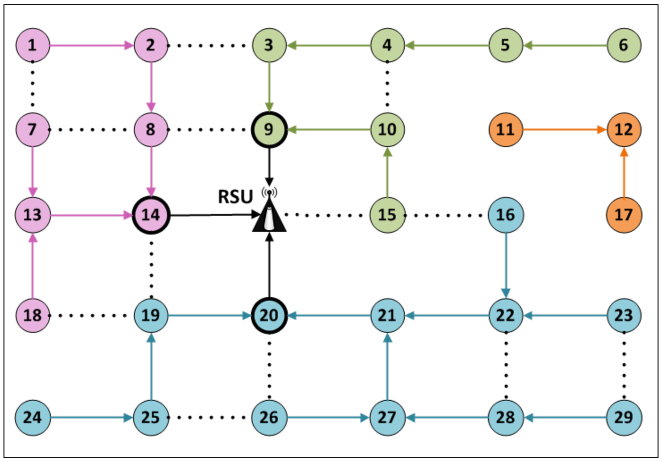

Figure 3. Illustration of routing paths resulting from DMSF algorithm for a regular vehicles distribution. The set of nodes $V_R = \{9, 14, 20\}$ represents the roots of the formed N_t trees.

Figure 3 shows an illustration of the paths built for a network graph with some disconnected parts, e.g., the tree $\{11, 12, 17\}$. The built trees reflect the fully distributed and parallel behavior of the algorithm. The root selection and rank updates of the fragments are similar to DMST algorithm. However, some updates are carried out to promote the data progress towards the RSU. These updates force the nodes member of V_R to be the roots of the formed trees. It is also worthy to mention that DMSF approach does not ensure the shortest path in terms of hop count however, it fosters the stable routes with higher throughput along with the convergence to the destination. For instance, the node 26 selects a more stable parent which is node 27 with hop-count equal to four instead of node 20 at two hops to RSU. The same scenario occurs with node 15, it chooses to go through longer route and did not join V_R because of the weak connectivity it has with the sink node.

Once the nodes finish building their routes, a data collection is initiated by the nodes joining the trees. Every node appends its locally generated data with the received data from children nodes in the routing trees, if any, and forwards data to its parent node. The roots of the trees deliver the received data to the collection point as soon as they receive data from their respective children nodes. When the next session is due, the DMSF algorithm is initiated again and the data gathering is performed by its end, and so forth. Or, in order to increase the data collection speed, new DMSFs can be created based on previously constructed DMSFs during previous session. This approach is left for a future extension of this work.

VI. PERFORMANCE ANALYSIS

This section describes the simulation environment and discusses the performance results of the proposed algorithms.

A. Simulation Setup

An urban traffic area is considered under the form of a grid network with controlled intersections. An RSU is deployed in the central junction of the studied area. The UAVs are hovering at a fixed altitude h_u above the junctions. The mobility traces of ground nodes were generated using the Simulator of Urban MObility (SUMO) [15] and the algorithms were implemented and tested in Matlab. Table I lists the simulation parameters that have been used.

TABLE I. SIMULATION PARAMETERS.

Parameter	*Value/Description*
area dimensions	1km x 1km
Number of junctions	5 x 5
Road segment length	250 m
System bandwidth	10 Mhz
Transmission power (P_{tr})	23.8 dBm
$PL_0(d_0)$	58.81 dB
Path Loss Exponent (ν)	1.83
σ_S	4.48 dB
$SINR_{min}$	-6.37 dB
$SINR_{max}$	7.35 dB
h_u	100 m

B. Performance results

In this section, the performance of the algorithms are quantified. Indeed, the data collection ratio, the required time to achieve the collection task, and the average number of hops to deliver data packets are studied for DMSF, DMST, and Greedy Perimeter Coordinator Routing (GPCR) [7] protocols. The latter is a position-based routing protocol. It leverages the planar property of the graph formed by road segments and junctions. GPCR uses the restricted greedy forwarding and a repair strategy to deal with local maximum problem [16]. Afterwards, the impact of using UAVs in assisting data collection using DMSF algorithm is evaluated for a low-density network.

1) Data Collection ratio: The considered data collection scenario in this paper is periodic and asynchronous. This means data could be concentrated in intermediate nodes which need to forward received data from children nodes to respective parent nodes. If the link quality is not considered in the selection of the next hop, many packets risk to be lost. Figure 4 shows the variation of the data collection ratio with the number of vehicles for the three algorithms. The proposed DMSF algorithm achieved clearly higher data delivery, slightly better than DMST particularly in low and medium traffic. For dense network, DMST performance drops significantly because of the increasing routing tree and the dynamic nature of the network. GPCR is a position-based routing algorithm which forwards data to the nodes ensuring always a better progress towards the destination node. It uses a restricted greedy forwarding. When the restricted greedy forwarding fails to find a next hop, a recovery strategy such as the $right\ hand\ rule$ is used. It is clear that the delivery ratio of this algorithm is limited for low-density network. However, it grows up with the network density because of the higher probability to get better relay node to reach the sink node. Even though, GPCR algorithm did not achieve DMSF packet delivery ratio.

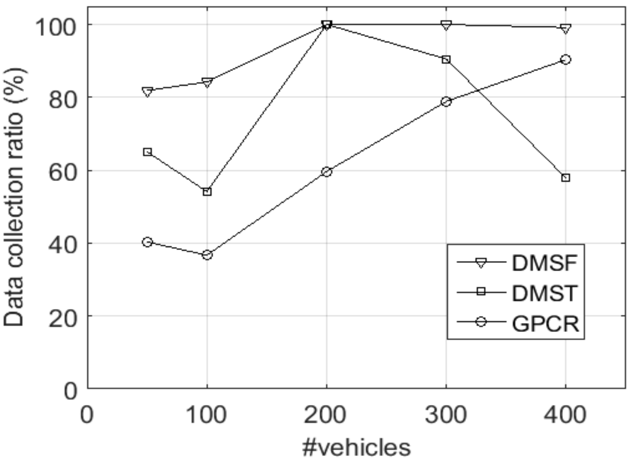

Figure 4. Data collection ratio for different traffic densities

2) Overall Delay: One major factor of data collection in mobile networks is the time bound within which the collection task has been achieved.

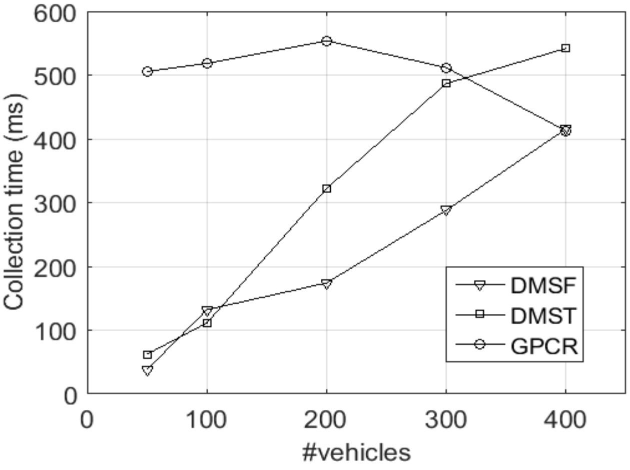

Figure 5. Data collection time for different traffic densities

The faster the collection is performed, the better the approach is. Indeed, in topology-based approaches, the computed paths are subject to change within a short time. Hence, a slow data forwarding will probably face outdated paths and end up with discarding packets or recomputing the path. Figure 5 shows the collection time in milliseconds to route the data from all the network nodes towards the sink. Of course, only nodes able to find a relay will send data. The other nodes will discard their data and wait to the next data collection session. For different densities of the network, the DMSF performs better than DMST, which is, in its turn, performing better than GPCR in terms of overall data collection delay for low and medium traffic densities. In dense network, GPCR overall delay outperforms DMST. In GPCR, the extra delay is caused by the long paths traveled by some packets when entering the recovery mode. The difference between DMSF and DMST in terms of collection time is due to the partitioned routing structures (i.e., trees) in DMSF leading to load balance and bottleneck avoidance in the neighborhood of the sink node. These factors lower the packet processing time in intermediate nodes, and hence lower the end-to-end journey delay of the packets.

3) Average number of hops: The average number of hops denotes the length of the paths of delivered packets traveled through. Figure 6 shows that GPCR presents shortest paths in average. This is due to low delivery ratio when the network is sparse, and the greedy nature of the protocol to approach the sink node in dense network. For DMSF, the results are slightly higher because of the consideration of the link quality and the distributed nature of the paths construction. DMSF performs better than DMST because the latter uses a single routing tree causing the leaf nodes to be more distant from the tree root.

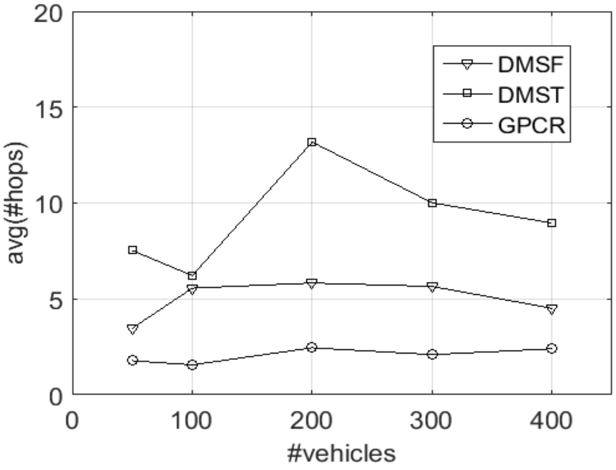

Figure 6. Average hops' number for different traffic densities

4) UAVs impact: For low density traffic, the network is fragmented. This causes a degraded performance of the data collection. Exploiting the existing UAVs as relay nodes between two fragments of the network has a positive impact on the system performance. Figures 7-9 depict the impact of using UAVs with DMSF algorithm in low-density network compared to DMSF with only ground nodes. The hovering of UAVs above the junctions allow them to link the network fragments and hence increase the packet delivery ratio which approaches 98% when using 24 drones. The overall delay and the average number of hops within the network did not change significantly which means the the UAVs enhance the network behavior with a very low cost.

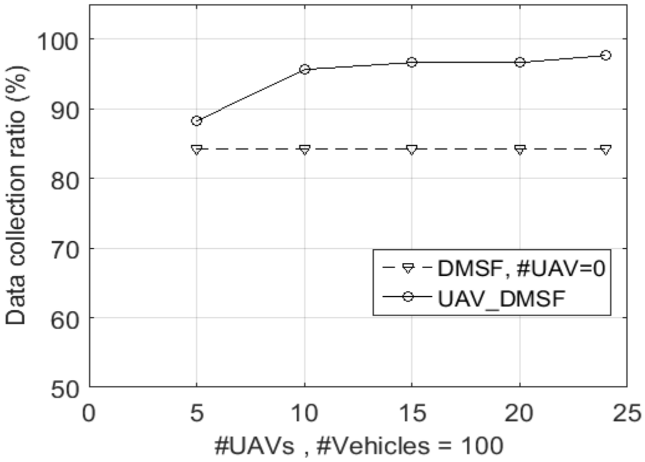

Figure 7. Impact of UAVs on data collection ratio

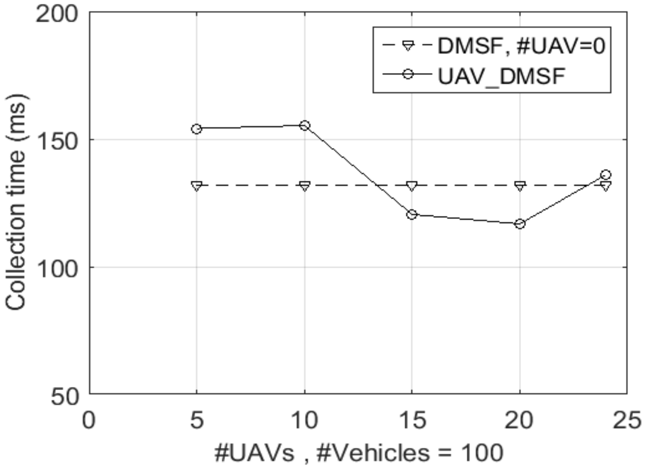

Figure 8. Impact of UAVs on data collection time

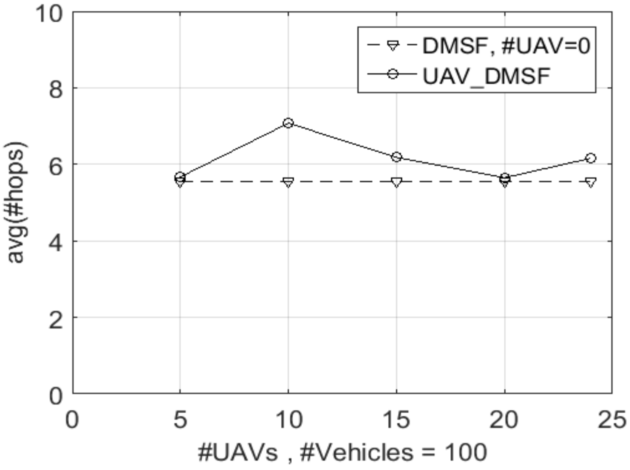

Figure 9. Impact of UAVs on average hops' number

ISBN: 978-1-61208-643-9

VII. CONCLUSION

Vehicular network enables a set of applications which make the driving experience safer and more efficient. Mobile nodes continuously transmit data which is gathered by remote devices. In this paper, the data collection challenge had been tackled in an urban area through the proposal of DMSF algorithm. The proposed solution achieves a very high data delivery ratio and scales better than GPCR and DMST algorithms in terms of collection time. The provided assistance of UAVs in relaying data between disconnected network fragments clearly enhances the data collection task with a very low cost in terms of overall delay. In a broader view, the present solution could help in enhancing the network nodes connectivity and end-to-end data delivery by assuring a local link-aware next-hop selection.

ACKNOWLEDGMENT

This publication was made possible by NPRP grant #NPRP8-2459-1-482 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] “ETSI TS 101 539-3 V1.1.1(2013-11) - Intelligent Transport Systems (ITS); V2X Applications; Part 3: Longitudinal Collision Risk Warning (LCRW)application requirements specification,” http://www.etsi.org/deliver/etsi_ts/101500_101599/10153903/01.01.01_60/ts_10153903v010101p.pdf, NOV 2013, accessed: 2018-04-21.

[2] “ETSI TS 101 539-3 V1.1.1(2013-11) - Intelligent Transport Systems (ITS); V2X Applications; Part 1: Road Hazard Signalling (RHS) application requirements specification,” http://www.etsi.org/deliver/etsi_ts/101500_101599/10153901/01.01.01_60/ts_10153901v010101p.pdf, AUG 2013, accessed: 2018-04-21.

[3] H. Fernández, L. Rubio, and V. M. Rodrigo-Peñarrocha, “Path loss characterization for vehicular communications at 700 MHz and 5.9 GHz under LOS and NLOS conditions,” IEEE Antennas and Wireless Propagation Letters, vol. 13, May 2014, pp. 931–934, DOI: 10.1109/LAWP.2014.2322261.

[4] H. Ghazzai, A. Feidi, H. Menouar, and M. L. Ammari, “An exploratory search strategy for data routing in flying ad hoc networks,” in Proceedings of the IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC) Oct 8–13, 2017, Montreal, QC, Canada, 2017, pp. 1–7, DOI: 10.1109/PIMRC.2017.8292474.

[5] H. Ghazzai, H. Menouar, and A. Kadri, “Data Routing Challenges in UAV-assisted Vehicular Ad hoc Networks,” in Proceedings of VEHICULAR 2017: the 6th International Conference on Advances in Vehicular Systems, Technologies and Applications July 23–27, 2017, Nice, France, 2017, pp. 85–90, ISBN: 978-1-61208-573-9.

[6] R. G. Gallager, P. A. Humblet, and P. M. Spira, “A Distributed Algorithm for Minimum-Weight Spanning Trees,” Journal ACM Transactions on Programming Languages and Systems (TOPLAS), vol. 5, no. 1, Jan. 1983, pp. 66–77, DOI: 10.1145/357195.357200.

[7] J. Liu et al., “A survey on position-based routing for vehicular ad hoc networks,” Springer Telecommunication Systems, vol. 62, no. 1, May 2016, pp. 15–30, Online ISSN: 1572-9451.

[8] S. Ucar, S. C. Ergen, and O. Ozkasap, “Multihop-Cluster-Based IEEE 802.11p and LTE Hybrid Architecture for VANET Safety Message Dissemination,” IEEE Transactions on Vehicular Technology, vol. 65, no. 4, Apr. 2016, pp. 2621–2636, DOI: 10.1109/TVT.2015.2421277.

[9] H. Huang, L. Libman, and G. Geersp, “An Agent Based Data Collection Scheme for Vehicular Sensor Networks,” in Proceedings of the IEEE 24th International Conference on Computer Communication and Networks (ICCCN) August 3–6, 2015, Las Vegas, NV, USA, 2015, DOI: 10.1109/ICCCN.2015.7288374.

[10] A. Soua and H. Afifi, “Adaptive data collection protocol using reinforcement learning for VANETs,” in Proceedings of the IEEE 9th International Wireless Communications and Mobile Computing Conference (IWCMC) July 1–5, 2013, Sardinia, Italy, 2013, pp. 1040–1045, DOI: 10.1109/IWCMC.2013.6583700.

[11] P. Fazio, F. D. Rango, C. Sottile, and A. F. Santamaria, “Routing Optimization in Vehicular Networks: A New Approach Based on Multiobjective Metrics and Minimum Spanning Tree,” International Journal of Distributed Sensor Networks, vol. 9, no. 11, 2013, pp. 1–13, DOI: 10.1155/2013/598675.

[12] A. Al-Hourani, S. Chandrasekharan, G. Baldini, and S. Kandeepan, “Propagation measurements in 5.8GHz and pathloss study for CEN-DSRC,” in Proceedings of the IEEE International Conference on Connected Vehicles and Expo (ICCVE) Nov 3–7, 2014, Vienna, Austria, 2014, pp. 1086–1091, DOI: 10.1109/ICCVE.2014.7297518.

[13] N. Wisitpongphan, O. K. Tonguz, and J. Parikh, “Broadcast storm mitigation techniques in vehicular ad hoc networks,” IEEE Wireless Communications, vol. 14, no. 6, Dec. 2007, pp. 84–94, DOI: 10.1109/MWC.2007.4407231.

[14] A. Burr, A. Papadogiannis, and T. Jiang, “MIMO Truncated Shannon Bound for System Level Capacity Evaluation of Wireless Networks,” in Proceedings of the IEEE Wireless Communications and Networking Conference Workshops (WCNCW) July 1–1, 2012, Paris, France, 2012, pp. 1040–1045, DOI: 10.1109/WCNCW.2012.6215504.

[15] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, “Recent development and applications of SUMO - Simulation of Urban MObility,” International Journal On Advances in Systems and Measurements, vol. 5, no. 3&4, Dec. 2012, pp. 128–138.

[16] C. Lochert et al., “Geographic Routing in City Scenarios,” ACM SIGMOBILE Mobile Computing and Communications Review, vol. 9, no. 1, Jan. 2005, pp. 69–72, DOI: 10.1145/1055959.1055970.