

University of Texas Rio Grande Valley

ScholarWorks @ UTRGV

Computer Science Faculty Publications and
Presentations

College of Engineering and Computer Science

2-2018

EDDA: An Efficient Distributed Data Replication Algorithm in VANETs

Junyu Zhu

Chuanhe Huang

Xiying Fan

Sipei Guo

Bin Fu

The University of Texas Rio Grande Valley

Follow this and additional works at: https://scholarworks.utrgv.edu/cs_fac



Part of the [Computer Sciences Commons](#)

Recommended Citation

Zhu, Junyu; Huang, Chuanhe; Fan, Xiying; Guo, Sipei; and Fu, Bin, "EDDA: An Efficient Distributed Data Replication Algorithm in VANETs" (2018). *Computer Science Faculty Publications and Presentations*. 26. https://scholarworks.utrgv.edu/cs_fac/26

This Article is brought to you for free and open access by the College of Engineering and Computer Science at ScholarWorks @ UTRGV. It has been accepted for inclusion in Computer Science Faculty Publications and Presentations by an authorized administrator of ScholarWorks @ UTRGV. For more information, please contact justin.white@utrgv.edu, william.flores01@utrgv.edu.

Article

EDDA: An Efficient Distributed Data Replication Algorithm in VANETs

Junyu Zhu ^{1,2}, Chuanhe Huang ^{1,2,*}, Xiyang Fan ^{1,2}, Sipei Guo ³ and Bin Fu ⁴

¹ School of computer, Wuhan University, Wuhan 430072, China; 2011@whu.edu.cn (J.Z.); wow@whu.edu.cn (X.F.)

² Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430072, China

³ School of Mathematics and Statistics, Central China Normal University, Wuhan 430079, China; gsp@mail.ccnucnu.edu.cn

⁴ Department of Computer Science, The University of Texas Rio Grande Valley, Edinburg, TX 78541, USA; bin.fu@utrgv.edu

* Correspondence: huangch@whu.edu.cn; Tel.: +86-27-6877-5602

Received: 11 January 2018; Accepted: 8 February 2018; Published: 10 February 2018

Abstract: Efficient data dissemination in vehicular ad hoc networks (VANETs) is a challenging issue due to the dynamic nature of the network. To improve the performance of data dissemination, we study distributed data replication algorithms in VANETs for exchanging information and computing in an arbitrarily-connected network of vehicle nodes. To achieve low dissemination delay and improve the network performance, we control the number of message copies that can be disseminated in the network and then propose an efficient distributed data replication algorithm (EDDA). The key idea is to let the data carrier distribute the data dissemination tasks to multiple nodes to speed up the dissemination process. We calculate the number of communication stages for the network to enter into a balanced status and show that the proposed distributed algorithm can converge to a consensus in a small number of communication stages. Most of the theoretical results described in this paper are to study the complexity of network convergence. The lower bound and upper bound are also provided in the analysis of the algorithm. Simulation results show that the proposed EDDA can efficiently disseminate messages to vehicles in a specific area with low dissemination delay and system overhead.

Keywords: VANETs; data dissemination; bounded number of messages; distributed consensus; sensor

1. Introduction

Vehicular ad hoc networks (VANETs), which aim to improve transportation safety and enable data services for in-vehicle consumption, have attracted much interest [1]. By enabling vehicles to communicate with one another and creating a large network with vehicles acting as the network nodes, various types of information (e.g., traffic conditions, advertising news and e-coupons) can be shared among vehicles. Interconnected by means of vehicle-to-vehicle (V2V) communications, VANETs can exchange information about the states of vehicles and roads, as well as provide infotainment services [2,3]. Especially, with the considerable demands of sensing and transmitting data for services, the vehicles can be regarded as mobile sensors, which can sense real-time vehicular surroundings [4].

Data dissemination is a promising application in VANETs, where messages are carried and forwarded by vehicular sensor nodes cooperatively toward their destinations. Through V2V communication in VANETs, the safety-related and commercial contents can be efficiently disseminated among a large number of vehicular sensor nodes. Thus, VANETs have significant potential to enable diverse applications associated with transportation safety, traffic efficiency and infotainment. There are many data dissemination applications, such as the important example of VANET safety

application, local danger warning (LDW). LDW creates a warning message for every detected hazard and informs the driver of its conditions [5]. In LDW applications, the information dissemination protocol has to implement strategies that provide reliable communication. There are also data dissemination applications and services that provide ad and infotainment information to users in VANETs, such as digital billboards [6], an ad service provider architecture for urban vehicular networks; an electric coupon system [7], which contains a bonus point-based coupon scheme and derives an optimal strategy that allows each user to determine how many bonus points she/he should ask for when passing the coupon; and FleaNet [8], a novel epidemic-based architecture that supports query dissemination, matching and notification routing and scales up to thousands of nodes without disrupting existing services. To fulfill the attractive applications by V2V data dissemination, messages can be disseminated via the vehicular environment. Extensive research works have been conducted to facilitate data dissemination, as it is of great importance in VANETs [9,10].

To achieve efficient data dissemination, data replication has been recognized as an effective approach for data delivery in vehicular networks [11]. Data replication enables multiple copies of the same data carried by different vehicles to be transmitted to a target destination simultaneously. Thus, the data will be distributed to a specific area in a quick manner. A variety of distributed algorithms has been proposed based on data replication [12]. Additionally, as VANETs are characterized by the lack of centralized control, distributed coordination among the nodes and averaging consensus problems are challenging issues in networks [13,14]. Due to the adaptability to dynamic network topologies, gossip algorithms, a distributed averaging approach based on rounds of communications between neighbor nodes, have received considerable attention because of their simplicity and robustness in noisy and uncertain environments [15]. The algorithms are usually compared by evaluating their speed of convergence, which can be measured by the number of transmissions. As gossip algorithms can waste significant energy by essentially passing around redundant information, significant efforts have been devoted to decrease the resource consumption, mainly through acceleration of the convergence rate. Similar to gossip algorithms, in replication-based data dissemination, each node possesses the message by utilizing a limited amount of local information to allow distributed knowledge of global network properties. In this research, we study distributed averaging to evaluate the efficiency of network convergence in replication-based data dissemination.

1.1. Our Goal

Dynamic data replication in distributed network systems can accelerate information spread in a specific area. However, possible issues with replication-based routing include: network congestion in clustered areas leading to a long delay, being wasteful of network resources (including bandwidth, storage and energy), and network scalability. Since network resources may quickly become constrained, deciding how to replicate the messages plays critical roles in many routing protocols. Meanwhile, some algorithms, such as epidemic and gossip algorithms, could cause significant network overhead by essentially passing around redundant information multiple times, which brings additional communication overhead.

On the other hand, by increasing the diversity of pairwise exchanges, data replication algorithms improve upon the convergence speed of data dissemination. As load balancing is an important goal in ad hoc networks, we let the data carrier distribute the data dissemination task to multiple vehicles in the area, and we hope every vehicle in the network can carry an approximately equal amount of data dissemination tasks. To achieve this, at each step, every vehicle computes a weighted average (mean of the two numbers) of its own value with values received from some of the other nodes until the network converges to a global average. The global average is calculated as the total number of message copies divided by the number of vehicles being averaged. The average calculation can be utilized to measure the convergence rate and complexity when network consensus is achieved in replication-based data dissemination. It is interesting to investigate how such nodes average their values, trying to achieve a general consensus in the shortest possible time.

To reduce unnecessary transmissions and improve communication overhead, the number of messages that can be replicated should be bounded when a message is disseminated to a specific area. Besides, through distributed averaging, the dissemination delay will decrease when the computing and communication burdens are distributed among the vehicles, which motivates us to combine distributed averaging with replication-based data dissemination. Therefore, we need to develop a data replication scheme that can control the number of message copies and combine distributed averaging with data dissemination to evaluate network convergence rate through data replication.

1.2. Main Contributions

Graph theory is used to help understand the topological properties of VANETs, where the vehicles and their communication links can be modeled as vertices and edges in the graph, respectively. According to the network traffic density, we divide the VANET topology into three types of graphs: linear graph, arbitrary graph and complete graph. In urban areas, due to fast speed of the vehicles, the network topology changes from time to time, so do the communication links. We can treat this type of network topology as arbitrary graph. In the scenario of a highway, assume vehicles move at a constant velocity along the road and every vehicle has full knowledge of its neighbors right next to it. In this situation, the network topology can be seen as a special case of arbitrary graph, that is a linear graph. We talk about the cases of arbitrary graph and linear graph in this study, while the case of complete graph will be discussed in the future work.

This paper investigates the problem of a vehicle node disseminating a message to a large number of recipients in a specific area in VANETs, and the objective is to disseminate the message to the destination area with low dissemination delay and reduced communication overhead. As shown in Figure 1, a message is generated at the source node, carried and forwarded by the passing vehicles to vehicles in the destination region. During the message transmission process, the message can be forwarded from one vehicle to another by V2V communications. When two vehicle nodes meet, they will update their values of the corresponding message after an average operation until the network consensus is reached. When the network reaches a consensus, we say that it is ϵ -balanced. In this context, we propose a distributed data replication algorithm, EDDA, for arbitrary graph and linear graph.

Before more details are given, we give a description of the ϵ -balanced status to help understand the proposed idea. Assume each vehicle i in the region carries ϵ message M and a parameter $n_i \geq 1$, which indicates how many message copies i can spread. After average operations, if $n_i (i = 1, \dots, m)$ carried by vehicle i satisfies the conditions $\sum_{i=1}^m n_i = n$ and $|n_i - n_j| \leq \epsilon$, we say that the network is ϵ -balanced. Here, n indicates the total number of copies, n_i indicates the number of copies held by node i and $\epsilon \in (0, 1]$.

To speed up the dissemination process, the proposed data replication algorithm lets the sender distribute the data dissemination task to vehicles in the same network by allowing each vehicle node to randomly choose one of their neighbors and average their current values, which indicate the amount of their assigned dissemination tasks. The pair of nodes computes the pairwise average, which then becomes the new value for both nodes. By iterating this pairwise averaging operations, it is easy to show that all the nodes converge to the global average in a completely distributed manner. Then, the paper analyzes the replication algorithm from the perspective of the approximation of how many nodes can receive the message when the system enters into a balanced status. It also presents the convergence speed of the network in the cases of arbitrary graph and linear graph. More specifically, we measure the complexity of network convergence by the number of communication stages in a distributed computing environment. A communication stage indicates an average operation among a set of independent edges in the connected graph. In each stage, each node is involved in at most one average operation. Theoretical analysis and simulation results show that the proposed algorithm can improve the efficiency of data dissemination and make the network converge to consensus quickly.

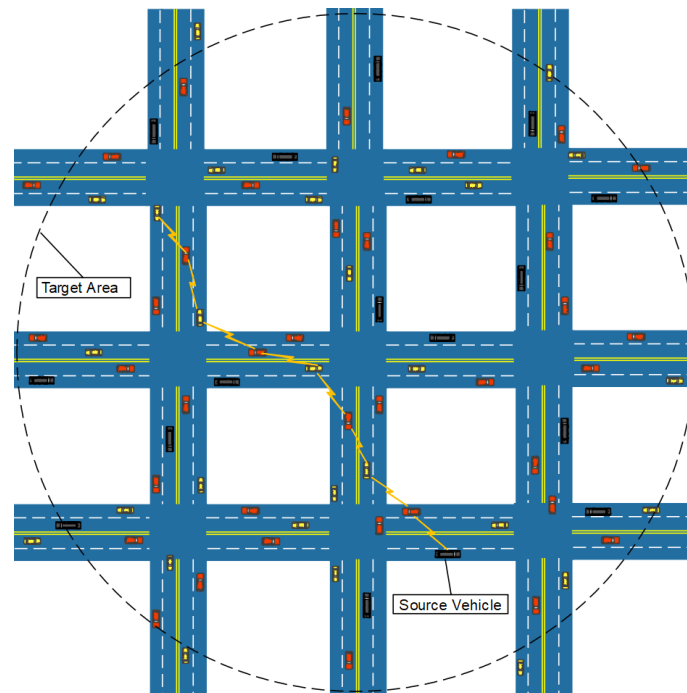


Figure 1. Data dissemination area.

Besides, the dissemination delay and overhead can be measured respectively by network convergence speed and the number of communication stages needed by the network to be ϵ -balanced. Through cooperative communication of vehicle nodes, the data dissemination process will speed up; thus, the delay for the network to converge to a balanced status can be reduced. The number of communication stages is also reduced as the computing and communication burdens can be distributed among the nodes, which can be seen as a reflection of reduced communication overhead.

To summarize, the main contributions are described as follows.

- (1) We apply graph theory to different scenarios in VANETs and divide the network into arbitrary graph, linear graph and complete graph. The cases of arbitrary graph and linear graph are discussed in this study. We propose a general system model for disseminating a bounded number of message copies in the network. Under this model, we develop EDDA, an efficient data replication algorithm that can be applied to both arbitrary graph and linear graph, in which the number of messages that can be replicated is limited and a network balanced status will be achieved.
- (2) We derive the theoretical analysis to obtain the approximate number of nodes that would receive the message when the system achieves an ϵ -balanced status. The convergence speed of the algorithm is also presented. As a special case of arbitrary graph, detailed analysis of the upper bound and lower bound for linear graph is provided, to show the efficiency of the proposed algorithm. The effectiveness of our algorithm has been validated by extensive simulations.

1.3. Paper Organization

The rest of this paper is organized as follows. Section 2 overviews the related work on data replication and average consensus problems in vehicular ad hoc networks. Section 3 first describes the analytical model of data replication, then presents a data replication algorithm, which can be applied to arbitrary graph and linear graph, and finally analyzes the approximation property of the proposed algorithm. Detailed theoretical analysis of the algorithm on convergence speed of arbitrary graph and linear graph is given in the following Section 4. Section 5 presents the simulation environment

and performance metrics to evaluate the performance of the compared algorithms. Finally, Section 6 concludes the paper.

2. Related Work

In this section, we give a brief overview of the related works. First, we discuss the research work on data dissemination in VANETs, then we introduce current research on the average consensus problem.

2.1. Data Dissemination Algorithms

Similar to other ad hoc networks, VANETs often suffer intermittent connectivity due to the high mobility of vehicles [16]. There are many existing works considering and studying how to develop efficient algorithms, to achieve low dissemination delay and system cost in VANETs [17–19]. As broadcast is the basic mechanism of VANET communication, flooding is the most common method in data dissemination. While flooding can achieve the maximum wireless coverage and rapid data dissemination, it may cause a serious broadcast storm. Torres et al. [20] compared ten different flooding-based algorithms and proposed an improved flooding scheme to cope with variant vehicle density situations. Yang et al. [21] first challenged the accuracy of the innovative assumption that was widely adopted in delay performance analysis of network-coding-based epidemic routing in delay-tolerant networks. Taherkhani et al. [22] proposed two strategies, DySch and TaSch. According to the strategies, the safety and service messages were assigned a priority value based on the characteristics of messages; such that the messages were disseminated dynamically and heuristically.

Although many technologies are proposed as potential solutions to support various communications in the network, V2V communication is still an important component of the VANET system. Related works using pure V2V communication are summarized as follows. Goonewardene et al. [23] designed a vehicle precedence algorithm to adaptively identify the nearby one-hop neighbors and select optimal cluster heads based on vehicle locations and velocities. Luo et al. [24] formed clusters based on geographically-divided grids, but they did not consider velocity and direction, which are important for accommodating the dynamic nature of VANETs. In [25], Ohta et al. used positions and moving direction of vehicles for clustering. Lin et al. [26] presented a novel moving-zone-based architecture and a corresponding routing protocol for message dissemination in VANETs by using pure V2V communications; the authors claimed that this was the first study that applied moving object techniques to vehicular networks. Kumer et al. [27] aimed to achieve lower message dissemination delay and reduce redundant rate. Bi et al. [28] used the time and angle method to determine the replicated node when RFID was not installed, to avoid network disconnection and improve message disseminating. Rivoirard et al. [29] proposed a clustering scheme that combines the information on the road configuration, vehicle mobility and link quality in order to build a structure relying only on the vehicles. Nilsson et al. [30] presented a measurement-based analysis of multilink shadowing effects in a V2V communication system with cars as blocking objects. Basheer et al. [31] reviewed several existing VANET safety applications and revised the disseminating methods for safety messages between vehicles without infrastructure to cover a wide area in a quick and reliable way.

In data dissemination, replication-based protocols have gained much attention in the scientific community, as they can allow for substantially better message delivery ratios than in forwarding-based protocols. These types of routing protocols allow for a message to be replicated; each of the replicas, as well as the original message itself are generally referred to as message copies or message replicas. Data replication-based protocols, such as epidemic routing [32], the PROPHET routing protocol [33] and MaxProp [34], have been instrumented to intentionally minimize one of four metrics: average delay, missed deadlines, maximum delay and communication overhead.

Some algorithms controlled the replication rules in data dissemination. Spyropoulos et al. [35] proposed spray-and-wait (SW), a routing protocol that attempts to gain the delivery ratio benefits of replication-based routing, as well as the low resource utilization benefits of forwarding-based routing.

Spray-and-wait can decrease the network resource consumption while it has similar performance compared to epidemic routing. However, it does not take into account bandwidth and network capacity. In contrast, Balasubramanian et al. [36] proposed a resource allocation protocol for intentional delay tolerant network (RAPID), which decided whether to replicate packets whenever two vehicles encountered each other with respect to a given performance metric. RAPID explicitly calculated the effect of replication on the routing metric while considering resource constraints. To tackle the problems that data dissemination could not effectively address broadcast storm and network partition problems simultaneously, Akabane et al. [37] proposed a suitable urban multi-hop broadcast protocol (TURBO) that relied exclusively on local one-hop neighbor information to deliver messages under dense and sparse networks. Yan et al. [38] designed a data dissemination scheme, which can disseminate data to a desired number of receivers in VANET (DOVE). The scheme was inspired by processor scheduling treating roads as processors to optimize the workload assignment and improve the efficiency of on-road dissemination. Takahashi et al. [39] proposed an advanced routing scheme that controls the maximum number of replicas according to the distance between the source node and the nearest base station. They also showed how to decide the maximum number of replicas by analyzing the message delivery reliability of an existing DTN routing scheme. Li et al. [40] evaluated how many idle resources of LTE could be provided for safety services and how safety applications impacted LTE traditional users and proposed to reserve the idle radio resources in LTE for vehicular safety services. Additionally, the weighted-fair-queueing (WFQ) algorithm was proposed to schedule beacons for safety services using the LTE reserved resources. Li et al. [41] investigated the budget-constrained and delay-bounded roadside units' (RSUs) placement problem in VANETs and formulated the budget constrained and delay-bounded placement (BCDP) problem to the budgeted maximum coverage problem, which was proven NP-hard. Heuristic algorithms were proposed to solve the problem, to reduce the RSU installation cost and provide a wide coverage for data dissemination. Chen et al. [42] investigated the data dissemination process in a cooperative communication mode. They fully utilized the capacity of VANETs' mobility and then developed an analytical framework to model the capacity achieved by the vehicles of interest (VoIs) in VANET. Li et al. [43] proposed an adaptive quality of service (QoS)-based routing for VANETs (AQRV), in which the intersections were adaptively chosen, through which data packets pass to reach the destination. To achieve the QoS constraints, the routing selection issue was mathematically formulated as a constrained optimization problem, and an ant colony optimization (ACO)-based algorithm was proposed to solve this problem. However, few of the schemes for data dissemination consider controlling the number of message copies in the network.

2.2. Average Consensus Problem

In cooperative control, a critical problem is to design appropriate protocols such that the group of agents can reach consensus on the shared information in the fixed and dynamically-changing interaction topologies. A consensus problem means convergence to a common value in the literature [44]. As a special case, an average consensus problem means convergence to the average of initial values, which is important in distributed decision-making for multiagent systems [45]. Average consensus is a fundamental problem in both distributed ad hoc networks and wireless sensor networks. The consensus status can be achieved through an iterative process of data transmission. During the past few decades, the problem attracted increasing interest in various research fields, and many gossip-based algorithms have been proposed to solve the consensus problem in wireless sensor networks [46,47].

Khosravi et al. [48] presented a gossip-based broadcast framework for strongly-connected topologies such as sensor networks, to solve the average consensus problem in an asynchronous environment where the network members might have different clock rates. Angelia et al. [49] first showed convergence to consensus under a bounded delay condition and some connectivity and intercommunication conditions imposed on the multi-agent system and then provided a bound on the time required to reach the consensus. Fabio and Sandro [50] investigated the consensus algorithms in

large-scale networks; different from average preserving algorithms, they allowed it to reach consensus at a point that may be different from the average of the initial states. The advantage of such algorithms was that they did not need bidirectional communication among agents, and thus, they applied to more general contexts. To overcome the drawbacks of the standard packet-based gossip algorithms, Aysal et al. [51] studied a broadcast-based gossiping algorithm for wireless sensor networks and proved that the random consensus value was the average of initial node measurements and that it could be made arbitrarily close to this value in the mean squared error sense under a balanced connectivity model and by trading off convergence speed with the accuracy of the computation. Boyd et al. [52] analyzed the averaging problem under the gossip constraint for an arbitrary network graph and proposed a distributed subgradient algorithm that solves the optimization problem over the network. More importantly, the authors established a tight relation between the averaging time of the algorithm and the mixing time of an associated random walk, which could be utilized to design fast averaging algorithms for two popular networks: wireless sensor networks, modeled as geometric random graphs, and the Internet graph under the preferential connectivity (PC) model. Wu et al. [53] proposed and analyzed a family of broadcast gossip algorithms for strongly-connected directed graphs. If the network was symmetric (undirected) or if nodes knew their out-degree, these algorithms were guaranteed to converge to the average consensus both in expectation and in the mean-squared sense. Nedić et al. [54] assumed the underlying graph was the complete graph and that all the weights were equal; the convergence of the weighted-averaging dynamics was established only for double-stochastic weights. To improve the result of previous work in [54], the authors [55] investigated the properties of the weighted-averaging dynamic for the consensus problem and established convergence of the weighted-averaging algorithm for general time-varying graphs. Shi et al. [56] studied finite-time convergence of deterministic gossiping and showed that there existed a symmetric gossip algorithm that converged in finite time if and only if the number of network nodes was a power of two. They also proved that there always existed an asymmetric gossip algorithm with finite-time convergence for any number of nodes with asymmetric updates.

The aforementioned schemes for data dissemination and network consensus in vehicular networks together improve reliability, efficiency and persistency of network performance. However, few of them consider controlling the number of message that can be replicated in the network. Thus, the existing results from the known mechanisms cannot be applied to achieve the goals of this paper. This paper investigates the network consensus problem in a bounded number of message passing, which can be achieved through a number of average operations. Meanwhile, it mainly targets reducing redundant data transmission and alleviating network burden by distributing the communication and computing burden loads among vehicles. Toward this goal, we develop a mathematical framework for data dissemination and propose an efficient distributed data replication algorithm.

3. Bounded Number of Data Replication in Message Passing

In this section, we first give an example to explain how the bounded number of data replication works in Section 3.1. Section 3.2 presents definitions and models used in the bounded number of messages passing. Then, Section 3.3 proposes an efficient data replication algorithm to reduce dissemination delay and redundant transmissions. An approximation of the number of nodes that receive the message is derived in Section 3.4.

3.1. An Example

In this part, we will show that data replication and the bounded number of message copies are quite useful in accelerating the information spread in a specific area. On the one hand, they can improve upon convergence speed of the network and realize network equilibrium. Besides, when compared with the basic broadcast mechanism, giving a restriction on the number of message copies helps to reduce redundant data dissemination. To explain how the bounded number of data replication works, an example is given as below.

We consider three different scenarios of VANETs, urban environment, highway and parking lot, which respectively correspond to arbitrary graph, linear graph and complete graph. The way that the mechanism works in different scenarios is illustrated separately.

Case 1: Assume in the common urban environment that the traffic can be described as arbitrary graph. If a vehicle carries message M , it needs to spread the message to its neighborhood district. Assume the total number for which the message can be replicated is $n = 100$ and the number of vehicles in this area is 20. The number of messages can be copied by each vehicle is denoted by n_i . We have $n_1 = 100$ and $n_j = 0$ for $j > 1$. Then, Vehicle 1 with $n_1 = 100$ communicates with a vehicle j within its communication area and $n_j = 0$, and then, the message parameters of the two vehicles take the average, with $n_1 = n_j = \frac{n_1 + n_j}{2} = 50$. The rest may be deduced by analogy. The ideal result is that each vehicle in this area carries message M with $n_i = 5$. Until then, the network enters into a balanced status.

Case 2: Assume the message passing happens on the highway, where the traffic can be described as linear graph. We keep the same parameters. It takes less average operations than arbitrary graph for the network to enter into a balanced status.

Case 3: Assume the message passing happens in a parking lot. As there are many vehicles in the parking lot, we consider the traffic as a complete graph. Then, there might be many vehicles in one vehicle's communication range. When the vehicle that carries message M with $n_i = 100$ receives more than one communication request, it chooses the vehicle with n_j if the gap between n_i and n_j is the largest. We hope that every vehicle in the neighborhood area receives message M and holds parameters with $|n_i - n_j| \leq \epsilon$, which means the network enters into a ϵ -balanced status and the network equilibrium is realized.

3.2. Definitions and Model

Assume vehicle i in the region carries message M and a parameter $n_i \geq 1$, where n_i indicates the number of copies assigned to be spread by vehicle i . The total dissemination tasks of all the participant vehicles should meet the condition, which is $\sum_{i=1}^m n_i = n$. Here, n indicates the maximum number of copies that M can be replicated and spread in the network. If $n_i, n_j (i, j \in 1, \dots, m)$ carried by any two vehicles i and j satisfy the condition $|n_i - n_j| \leq \epsilon$, we say the system is ϵ -balanced. An ϵ -balanced status will be obtained after a series of average operations. We define some concepts in this section.

When a node carries message M and it controls at most a copies of message M to be distributed in the network, it must have $a \geq 1$, and each node with a nonzero value is at least one.

We need to define the concept of potential (energy) in order to analyze the number of stages for the system entering into a balanced status and need the following lemma.

Lemma 1. Assume that a, b, c , and d are real numbers with $a + b = c + d$. Then, we have (1) $(a^2 + b^2) - (c^2 + d^2) = 2(b - d)(b - c)$ and (2) $(a^2 + b^2) - (c^2 + d^2) \geq 0$ if $a \leq c \leq d \leq b$.

Proof. By the condition of the lemma, we have $b - d = c - a$ and $b - c = d - a$. Thus, $(a^2 + b^2) - (c^2 + d^2) = (b - d)(b + d) + (a - c)(a + c) = (b - d)(b + d - c - a) = 2(b - d)(b - c)$. This proves (1). The second part of this lemma can be easily proven with (1). \square

We will utilize graph theory to help understand the topological properties of VANETs. The vehicles and their communication links can be respectively treated as vertices and edges in the graph. The graph definition is described as below.

Definition 1. For a set of vehicles, their connected graph is an undirected graph $G(V, E)$ such that each node represents a vehicle and an edge between two nodes indicates that the corresponding vehicles are within the distance of communication.

- In urban areas, due to fast speed of the vehicles, the network topology changes from time to time, so do the communication links. We can formalize this type of network topology as arbitrary graph.
- In the scenario of a highway, assume vehicles move at a constant velocity along the road and every vehicle has full knowledge of its neighbors right next to it. In this situation, the network topology can be seen as a special case of arbitrary graph, that is linear graph.

Definition 2. Let M be a message. Let $G(V, E)$ be the connected graph for a set of vehicles. If each node i has a parameter n_i to control the number of copies of message M that i can replicate, then $G(V, E)$ associated with n_i becomes a graph with a bounded number of message copies.

Definition 3. Let $G(V, E)$ be a connected graph. Each node of G is assigned a nonnegative number n_i . The nodes of G are ϵ -balanced in the corresponding bounded message graph if the following conditions are satisfied:

- Each node of G with $n_i > 0$ satisfies $n_i \geq 1$.
- For every two nodes with $n_i, n_j > 0$, $|n_i - n_j| \leq \epsilon$, and
- There is no edge between nodes of values n_i and n_j in G , respectively, such that $n_i \geq 2$ and $n_j = 0$.

Definition 4. Let \mathbb{R} be the set of real numbers and \mathbb{N} be the set of nonnegative integers. Define the following concepts:

- A real average function $A(.,.)$ is a mapping $\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \times \mathbb{R}$, such that for two numbers $a \leq b$, $A(a, b) = (\frac{a+b}{2}, \frac{a+b}{2})$ if $a + b \geq 2$, or $A(a, b) = (a, b)$ if $a + b < 2$.
- An integer average function $A(.,.)$ is a mapping $\mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$ such that for two numbers $a \leq b$, $A(a, b) = (k, k)$ if $a + b = 2k \geq 2$, $A(a, b) = (k, k + 1)$ if $a + b = 2k + 1 \geq 2$, or $A(a, b) = (a, b)$ if $a + b < 2$.
- For a list $L : a_1, a_2, \dots, a_m$ of numbers, define the potential of L to be $P(L) = a_1^2 + a_2^2 + \dots + a_m^2$.
- For an average function $A(.,.)$, define $S_A(\langle a, b \rangle) = 2(b - d)(b - c)$, where $A(a, b) = (c, d)$. Number b is considered a bar of length b . $S_A(\langle a, b \rangle)$ can be considered a small piece of length $b - d$ from the bar of length b to go down by $(b - c)$. Function $S_A(\cdot)$ gives the potential change after an average operation (See Lemma 1).
- Let $A(.,.)$ be an average function. Assume that a_1, a_2, \dots, a_n is a list of numbers. It is transformed into another list a'_1, a'_2, \dots, a'_n by a series of average operations. Define its sum of the product to be $S(H) = \sum_{(a,b) \in H} S_A(a, b) = P(L) - P(L')$ (see Lemma 1), where H is the set of tuples (a, b) that take average operations to transform the first list into the second list. It is considered as the change of the potential after taking a series of average operations.

Definition 5. A stage of communication is an average operation among a set of independent edges in the connected graph, which consists of a set of nodes. Two nodes connected by one of the independent edges can communicate. It allows those pairs of nodes to exchange messages in parallel.

Definition 6. Convergence speed represents the speed that the network can enter into an ϵ -balanced status. In this paper, we use the number of stages (see Definition 5) to reflect the convergence speed.

We use the number of stages to characterize the complexity to enter into a balanced status among a set of nodes whose communication is based on their connected graph, in which each node represents a vehicle and every edge connects two vehicles within a distance no greater than their range of communication. Furthermore, the number of stages can reflect how many transmissions are needed to achieve an ϵ -balanced status. Thus, the number of stages can be utilized to denote communication overhead in the network. Through the stages, the message dissemination tasks are distributed to the number of nodes for cooperative delivery. The process of task assignment is as shown in Figure 2.

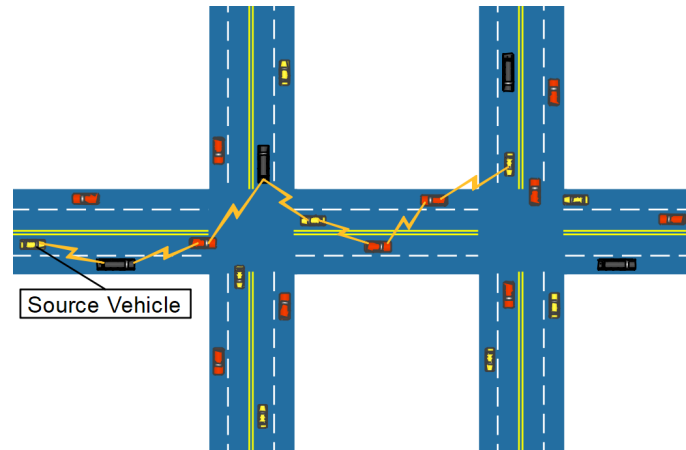


Figure 2. Task assignment process.

3.3. Algorithm

Here, to improve the performance of data dissemination, we propose an efficient data replication algorithm EDDA that can make the network system enter an ϵ -balanced status in a short time, which is presented as algorithm 1.

In the proposed data replication algorithm, a message M is carried by one specific node. We need to distribute the message to the nodes in a given area. The total number of messages M that can be replicated is controlled by a parameter n . Each node i receives a corresponding parameter $n_i \geq 1$ to control the number of copies for which M can be replicated when it receives message M , where n_i indicates the number of copies that can be distributed by node i . After a certain number of stages, those n_1, \dots, n_m will enter into a balanced status (see Definition 3). The complexity to enter into a balanced status is measured by the number of stages that are counted in algorithm 1. To give an overview, EDDA schedules with the following three steps.

- (1) First, EDDA constructs the graph and initializes the value of every vertex of graph G . In the initialization procedure (see algorithm 2), the vehicle node that carries the message will be assigned a value of n . All other nodes will be assigned a value of zero, which means they do not have the message.
- (2) Second, select independent edges from G so different pairs of nodes can communicate with each other in parallel. After the selection, replace the values of nodes with their new values by taking the average of the current values. Each stage should update the values of the nodes in the graph one time. Then, go to the next stage, and stop the average operations until the system is ϵ -balanced.
- (3) Third, EDDA outputs graph G' , with the final values of all nodes updated. If new nodes enter into the network and break the balance, the procedures will be executed again to achieve network balance.

The pseudocode of EDDA is described as follows, the initialization procedure is presented separately as algorithm 2.

3.4. Approximation

To analyze the proposed algorithm, we discuss how many nodes can receive the message when the system achieves an ϵ -balanced status.

Lemma 2. Assume that L is a list of numbers. Let H be a finite set of tuples taking average operations when the list is transformed into list L' . Then, the sum of numbers in L' is the same as the sum of numbers in L .

Algorithm 1 Data replication algorithm.

Input: bounded message graph G (see Definition 2); parameter ϵ
Output: Bounded message graph G' ; number of stages a

- 1: Call Algorithm 2;
- 2: Let $a = 0$;
- 3: **repeat**
- 4: Select independent edges (disjoint pairs of G) $(i_1, j_1), (i_2, j_2), \dots, (i_k, j_k)$ with $n_{i_t} \neq n_{j_t}$ for $t = 1, \dots, k$;
- 5: **for** $t = 1$ to k **do**
- 6: $n_{i_t} = (n_{i_t} + n_{j_t})/2$;
- 7: $n_{j_t} = (n_{i_t} + n_{j_t})/2$;
- 8: $a = a + 1$;
- 9: let n_{max} and n_{min} be the maximum and minimum value of G , respectively;
- 10: **until** $(n_{max} - n_{min} \leq \epsilon)$
- 11: **return** a and graph G' with updated values;

Algorithm 2 Initialization.

Input: Graph G ; parameter n
Output: Weighted graph G with values

- 1: Let v_i denote the nodes in graph G ;
- 2: Let v_1 ($i = 1$) denote the node that carries message M ;
- 3: Let n_{v_i} denote the assigned value of node v_i ;
- 4: $n_{v_1} = n$;
- 5: **for** $i = 2$ to m **do**
- 6: $n_{v_i} = 0$;

Proof. It follows from the definition of the average operation. \square

Theorem 1. Let m be the number of nodes that form a connected graph and n be an integer to control the number of copies of the message to be sent. Then, when the system enters into an ϵ -balanced status, there are at least $\min(\frac{n}{2+\epsilon}, m)$ nodes that have received the message.

Proof. When the system is ϵ -balanced, we have $|n_i - n_j| \leq \epsilon$ for any two nodes that can communicate. We also have $n_i \geq 1$ for each $n_i \neq 0$.

Case 1. Every $n_i \neq 0$. In this case, we have $n_i \geq 1$, and all m nodes have received the message. It is trivial to see $m \geq \min(\frac{n}{2+\epsilon}, m)$.

Case 2. There is at least one n_i with $n_i = 0$. In this case, each positive n_i satisfies $n_i \leq 2 + \epsilon$. Otherwise, there is a pair of neighbors n_s and n_t such that $n_s = 0$ and $n_t \geq 2$ since m nodes are connected. They can take average operations and bring a contradiction. Therefore, the number of nodes that have received the message is at least $\frac{n}{2+\epsilon}$ by Lemma 2. \square

4. Speed of Convergence on Arbitrary Graph and Linear Graph

In this section, we show that the system will converge to an ϵ -balanced status after a finite number of stages. We first derive the convergence speed of arbitrary graph, then present detailed analysis of the upper bound and lower bound of linear graph.

4.1. Arbitrary Graph

In this part, we discuss the convergence speed of arbitrary graph. Lemma 1 shows that the potential of a list goes down after each average operation.

Definition 7. Let real number $w \geq 0$. Let $G(V, E)$ be a graph and $V' \subseteq V$. If every node $v \in V'$ has a weight at least w , then V' is a w -region of G . A w -spread of G satisfies the conditions:

1. A node with a weight of at least $2w$ can take the average with a node with weight zero, and
2. two nodes with a weight of at least w can take the average.

Lemma 3. Assume that every node in the list of nodes in an arbitrary graph $G(V, E)$ has a weight of at least w . A w -spread takes at most $\frac{4n^4}{\epsilon}$ stages of real average operations to reach the ϵ -balanced state.

Proof. Let L be the list of nodes with weights in the beginning. Clearly, we have $P(L) \leq n^2$.

Assume that there are two nodes with $n_u - n_v > \epsilon$. There is a path from u to v in the graph such that each node has a value of at least one. Thus, the path length has at most n nodes. There are two nodes with values n_i and n_j such that $n_j - n_i \geq \frac{\epsilon}{n}$. Thus, $S_A(n_i, n_j) \geq (\frac{\epsilon}{2n})^2$.

Therefore, the total number of stages is at most $\frac{P(L)}{(\frac{\epsilon}{2n})^2} \leq \frac{n^2}{(\frac{\epsilon}{2n})^2} = \frac{4n^4}{\epsilon^2}$. \square

Theorem 2. Let w be greater than zero. Assume that the list of nodes in an arbitrary graph $G(V, E)$ has a total sum of weights at most n . It takes at most $\frac{4n^4 \cdot \min(m, \frac{n^2}{w^2})}{\epsilon}$ stages of w -spread operations to reach the ϵ -balanced state, where m is the number of nodes in the graph.

Proof. Let L be the list of nodes with weights in the beginning. We have $P(L) = n^2$. The number of average operations between a node with $(n_i \geq 2w) \wedge (n_i > 0)$ and a node with $n_j = 0$ is at most m . The number of the average number of average operations between a node with $(n_i \geq 2w) \wedge (n_i > 0)$ and a node with $n_j = 0$ is at most $\frac{n^2}{w^2}$ since an average operation has $S(n_i, n_j) \geq w^2$. This follows from Lemma 3. \square

4.2. Upper Bound for Linear Graph

Now, we discuss the convergence speed of linear graph that generally describes the traffic on a highway. We believe linear graph is an important special case that should be studied.

Definition 8. A linear graph $G(V, E)$ consists a list of nodes v_1, v_2, \dots, v_n , and the set of edges $E = \{v_1v_2, v_2v_3, \dots, v_iv_{i+1}, \dots, v_{n-1}v_n\}$. In other words, each edge connects two consecutive nodes in the list v_1, v_2, \dots, v_n , and every two consecutive nodes has an edge connecting them.

Theorem 3. Assume that the list of nodes in a linear graph $G(V, E)$ starts with the initial values $L : n, 0, 0, \dots, 0$. Then, it takes at most $O(\frac{n^2}{\epsilon})$ stages of real average operations to reach the stage of ϵ -balance for a linear graph.

Proof. The total number of average operations $A(n_i, 0)$ with $n_i \geq 2$ is at most n . We focus on the number of average operations $A(n_i, n_{i+1})$ with $n_i, n_{i+1} \geq 1$.

Let $n_1 \geq n_2 \geq n_3 \dots \geq n_m$ with $m \leq n$ since there are at most m nodes with a value of at least one. Let $d_i = n_i - n_{i+1}$.

Let each node i take an average operation with one of the neighbors $j \in \{i - 1, i + 1\}$ such that i and j do not take the average in the last stage. In two consecutive stages, we have made at least contribution $\sum_{i=1}^{m-1} (\frac{d_i}{2})^2 \geq \frac{(\sum_{i=1}^{m-1} d_i/2)^2}{m-1} = \frac{(\sum_{i=1}^{m-1} d_i)^2}{4(m-1)}$ to $S(H)$ by the Cauchy–Schwarz inequality, where H is the set of all average operations.

Let k be the minimum integer with $\frac{n}{2^k} \leq \epsilon$. Divide $[\frac{n}{2^k}, n]$ into $O(\log n)$ intervals: $I_1 = (\frac{n}{2}, n], I_2 = (\frac{n}{2^2}, \frac{n}{2}], \dots, I_k = (\frac{n}{2^k}, \frac{n}{2^{k-1}}]$. The region t represents the interval I_t such that $n_1 - n_m \in (\frac{n}{2^{t+1}}, \frac{n}{2^t}]$. When the system enters into the balanced status, we have $n_1 - n_m \leq \frac{n}{2^t} \leq \epsilon$. To achieve the maximum potential, we assume there are 2^t numbers, which are equal to $\frac{n}{2^t}$ (which satisfies $2^t \cdot \frac{n}{2^t} \geq n$). According to the definition of potential $P(L) = a_1^2 + a_2^2 + \dots + a_m^2$, here

$a_1 = \dots = a_m = \frac{n}{2^t}$, $m = 2^t$, we can get that the maximum potential of a list in region t is at most $2^t \left(\frac{n}{2^t}\right)^2$ (see Part 2 of Lemma 1). Therefore, it takes at most $2^t \left(\frac{n}{2^t}\right)^2 \cdot \frac{1}{\frac{(\frac{n}{2^t})^2}{4(m-1)}} \leq 4(m-1)2^t$ stages in the region t .

The total number of stages is bounded by $n + \sum_{t=1}^k 4m2^t = O\left(\frac{n^2}{\epsilon}\right)$. \square

Theorem 4. Let w be a parameter greater than zero. Assume that the list of nodes in a linear graph $G(V, E)$ starts with the initial values $L : n, 0, 0, \dots, 0$. Then, it takes at most $O\left(\frac{n^2}{\epsilon} \cdot \min\left(m, \frac{n}{w}\right)\right)$ stages of w -spread operations to reach the stage of ϵ -balance for a linear graph.

Proof. Since the total number of nodes is m , there are at most m times that the average operation involves a node with zero value. There are at most $\frac{n}{w}$ nodes to have a value of at least w . This follows from Theorem 3. \square

According to the above theorems, the list of nodes in a linear graph $G(V, E)$ with initial values $L : n, 0, 0, \dots, 0$ reaches the balance after certain stages. Assume a new node with value zero joins the list; it is obvious that the balance will be broken. As the values of the new list are smaller than the initial ones, it should take fewer real average operations for the system to enter into a new balanced status, which is relatively quicker than the previous balance process.

4.3. Lower Bound for Linear Graph

Following the result of the upper bound, we derive a nontrivial super linear lower bound for linear graph. It is derived based on some bottleneck properties for messages passing over a linear graph.

We consider the linear graph for which Node 1 connects to Node 2, node i connects to both node $i - 1$ and node $i + 1$ for $1 < i < m$ and node m connects to node $m - 1$. For $m = \frac{n}{8}$, it is trivial to see a linear lower bound to enter into ϵ -balance since it takes $\Omega(m)$ stages for a positive value to reach node m from Node 1 in a linear graph.

Definition 9. Assume that, Node 1 has value n , and all of the other nodes have zero in the beginning. A normalized process is that at stage j with odd j , node $2i - 1$ takes the average with $2i$ for $i = 1, 2, \dots$; and, at stage j with even j , node $2i$ takes the average with $2i + 1$ for $i = 1, 2, \dots$.

Definition 10. Assume that m nodes have values $n_1 \geq n_2 \geq \dots \geq n_m$, respectively. The gap at node i is defined by $n_i - n_{i+1}$.

For a node i with a value of nine and its neighbor node $i + 1$ with a value of three, both nodes have a value of $\frac{9+3}{2} = 6$ after taking the average operation. Node i sends the $9 - 6 = 3$ value to node $i + 1$, and node $i + 1$ receives three from this average operation. Each node can only take one average operation each moment. Average operations do not change the total values among the nodes.

Definition 11. For each node i , define $\text{Rec}(i, j)$ to be the amount that node i has received after j stages and $\text{Snd}(i, j)$ to be the amount that node i has sent after j stages.

Definition 12. For each node i , define $\text{MaxRec}(i, j)$ to be the largest amount that node i can receive after j stages and $\text{MaxSnd}(i, j)$ to be the amount that node i can send after j stages. Define $\text{Current}(i, j)$ to be the current value at node i right after stage j .

Lemma 4. Let $g_{1,j}, g_{2,j}, \dots, g_{m',j}$ be the positive gaps (see Definition 10) between two consecutive nodes in the linear graph from Node 1 to node m after stage j . Using normalized processes, we always get $g_{1,j} \geq g_{2,j} \geq \dots \geq g_{m',j}$.

Proof. This follows a simple induction. It is trivial at Stage 0 since we only have Node 1 with a value of n , and all of the other nodes have a value of zero after Stage 0. Assume that this is true after stage j .

Consider stage $j + 1$. After stage $j + 1$ in the normalized process, every positive gap after stage $j + 1$ is equal to $\frac{g_{i,j} + g_{i+1,j}}{2}$. For two consecutive gaps $\frac{g_{i,j} + g_{i+1,j}}{2}$ and $\frac{g_{i+1,j} + g_{i+2,j}}{2}$, we have $\frac{g_{i,j} + g_{i+1,j}}{2} \geq \frac{g_{i+1,j} + g_{i+2,j}}{2}$ because $g_{i,j} \geq g_{i+2,j}$ by the inductive hypothesis. Therefore, it is nondecreasing. \square

Lemma 5. Using the normalized stages, each node i has $\text{Rec}(i, j) = \text{MaxRec}(i, j)$ and $\text{Snd}(i, j) = \text{MaxSnd}(i, j)$ for all j .

Proof. We prove this by contradiction. Assume that $\text{Rec}(i, j) < \text{MaxRec}(i, j)$ or $\text{Snd}(i, j) < \text{MaxSnd}(i, j)$ for some i, j . Let (i, j) be the tuple such that it satisfies the following conditions:

- Either $\text{Rec}(i, j) < \text{MaxRec}(i, j)$ or $\text{Snd}(i, j) < \text{MaxSnd}(i, j)$,
- Integer j is the least, and $(\text{Rec}(i, j) < \text{MaxRec}(i, j)$ or $\text{Snd}(i, j) < \text{MaxSnd}(i, j))$.

(A) We assume $\text{Rec}(i, j) < \text{MaxRec}(i, j)$. Let P^* be another process that has $\text{Rec}^*(\cdot), \text{Current}^*(\cdot), \text{Snd}^*(\cdot)$, and it has $\text{Rec}^*(i, j) = \text{MaxRec}(i, j)$.

Case 1. $i = 1$. It is trivial that $\text{Rec}(i, j) = \text{MaxRec}(i, j) = n$ for all j . Therefore, we have a contradiction.

Case 2. $i > 1$ and node i take the average with node $i - 1$ at stage j in the normalized process. Node i receives flow from $i - 1$; node $i - 1$ has $\text{Rec}(i - 1, j - 1) = \text{MaxRec}(i - 1, j - 1)$; and node i has $\text{Snd}(i, j - 1) = \text{MaxSnd}(i, j - 1)$. Thus, we have $\text{Current}(i, j) = \frac{\text{MaxRec}(i - 1, j - 1) - \text{MaxSnd}(i, j - 1)}{2}$.

In this case, we have $\text{Current}(i - 1, j - 1) > \text{Current}(i, j - 1)$, and there is an average operation between node $i - 1$ and node i at stage j .

We have:

$$\begin{aligned} & \text{Current}(i - 1, j) = \text{Current}(i, j) \\ = & \frac{\text{Current}(i - 1, j - 1) + \text{Current}(i, j - 1)}{2} \\ = & \frac{(\text{MaxRec}(i - 1, j - 1) - \text{MaxSnd}(i - 1, j - 1))}{2} \\ & + \frac{(\text{MaxRec}(i, j - 1) - \text{MaxSnd}(i, j - 1))}{2} \\ = & \frac{\text{MaxRec}(i - 1, j - 1) - \text{MaxSnd}(i, j - 1)}{2}. \end{aligned}$$

Thus, we have $\text{Current}(i - 1, j) = \text{Current}(i, j) = \frac{\text{MaxRec}(i - 1, j - 1) - \text{MaxSnd}(i, j - 1)}{2}$.

Thus,

$$\begin{aligned} & \text{Rec}(i, j) \\ = & \text{Current}(i, j) + \text{Snd}(i, j - 1) \\ = & \frac{\text{MaxRec}(i - 1, j - 1) - \text{MaxSnd}(i, j - 1)}{2} \\ & + \text{MaxSnd}(i, j - 1) \\ = & \frac{\text{MaxRec}(i - 1, j - 1) + \text{MaxSnd}(i, j - 1)}{2}. \end{aligned}$$

Now, consider the process P^* . We have:

$$\begin{aligned}
 & \text{Rec}^*(i, j) \\
 = & \text{Current}^*(i, j) + \text{Snd}^*(i, j - 1) \\
 = & \frac{(\text{Rec}^*(i - 1, j - 1) - \text{Snd}^*(i - 1, j - 1))}{2} \\
 & + \frac{(\text{Rec}^*(i, j - 1) - \text{Snd}^*(i, j - 1))}{2} \\
 & + \text{Snd}^*(i, j - 1) \\
 = & \frac{\text{Rec}^*(i - 1, j - 1) + \text{Snd}^*(i, j - 1)}{2}.
 \end{aligned}$$

On the other hand, we have $\text{Rec}^*(i - 1, j - 1) \leq \text{MaxRec}(i - 1, j - 1)$ and $\text{Snd}^*(i, j - 1) \leq \text{MaxSnd}(i - 1, j - 1)$. Thus, $\text{Rec}^*(i, j) \leq \text{Rec}(i, j)$. A contradiction.

Thus, $\text{Rec}(i, j) = \text{MaxRec}(i, j)$. It is easy to see that $\text{MaxSnd}(i - 1, j) = \text{MaxRec}(i, j)$.

Case 3. $i > 1$, and node i does not take the average with node $i - 1$ at stage j in the normalized process.

We have $\text{Current}(i - 1, j - 1) = \text{Current}(i, j - 1)$ since there is no average operation between node $i - 1$ and node i .

There is an average operation between node i and node $i + 1$.

In this case, we have $\text{Rec}(i, j) = \text{Rec}(i, j - 1) = \text{MaxRec}(i, j - 1)$ by the inductive hypothesis.

Now, consider process P^* . If node i does not take the average with node $i - 1$ at stage j in P^* , by the inductive hypothesis, $\text{Rec}(i, j) = \text{MaxRec}(i, j - 1) \geq \text{Rec}^*(i, j - 1) = \text{Rec}^*(i, j)$. A contradiction.

If node i does take the average with node $i - 1$ at stage j , we still have that $\text{Rec}^*(i, j - 1) = \text{Rec}^*(i, j - 2)$. Clearly, node i does not take the average with node $i - 1$ at stage $j - 1$ (otherwise, node i does not take the average with node $i - 1$ at stage j since their values after stage $j - 1$ are equal ($\text{Current}(i - 1, j - 1) = \text{Current}(i, j - 1)$)).

At the normalized process, node i takes the average with node $i - 1$ at stage $j - 1$. Thus, $\text{Rec}(i - 1, j - 1) \leq \text{MaxRec}(i - 1, j - 2)$ and $\text{Snd}(i, j - 1) \leq \text{MaxSnd}(i, j - 2)$.

We have:

$$\begin{aligned}
 & \text{Rec}^*(i, j) \\
 = & \text{Current}^*(i, j) + \text{Snd}^*(i, j - 1) \\
 = & \frac{(\text{Rec}^*(i - 1, j - 1) - \text{Snd}^*(i - 1, j - 1))}{2} \\
 & + \frac{(\text{Rec}^*(i, j - 1) - \text{Snd}^*(i, j - 1))}{2} \\
 & + \text{Snd}^*(i, j - 1) \\
 = & \frac{\text{Rec}^*(i - 1, j - 1) + \text{Snd}^*(i, j - 1)}{2} \\
 \leq & \frac{\text{MaxRec}(i - 1, j - 1) + \text{MaxSnd}(i, j - 1)}{2} \\
 = & \frac{\text{MaxRec}(i - 1, j - 2) + \text{MaxSnd}(i, j - 2)}{2}.
 \end{aligned}$$

On the other hand, node i takes the average with node $i - 1$ at stage $j - 1$. Thus,

$$\begin{aligned}
 & \text{Rec}(i, j - 1) \\
 = & \text{Current}(i, j - 1) + \text{Snd}(i, j - 1) \\
 = & \frac{(\text{Rec}(i - 1, j - 1) - \text{Snd}(i - 1, j - 1))}{2} \\
 & + \frac{(\text{Rec}(i, j - 1) - \text{Snd}(i, j - 1))}{2} \\
 & + \text{Snd}(i, j - 1) \\
 = & \frac{\text{Rec}(i - 1, j - 1) + \text{Snd}(i, j - 1)}{2} \\
 = & \frac{\text{Rec}(i - 1, j - 2) + \text{Snd}(i, j - 2)}{2} \\
 = & \frac{\text{MaxRec}(i - 1, j - 2) + \text{MaxSnd}^*(i, j - 2)}{2}.
 \end{aligned}$$

Thus, $\text{Rec}(i, j - 1) \geq \text{Rec}^*(i, j)$. A contradiction.

(B) We assume $\text{Snd}(i, j) < \text{MaxSnd}(i, j)$.

Since $\text{Snd}(i, j) = \text{Rec}(i + 1, j)$ and $\text{MaxSnd}(i, j) = \text{MaxRec}(i + 1, j)$, it transforms into the problem $\text{Rec}(i + 1, j) < \text{MaxRec}(i + 1, j)$, which is Case A, and a contradiction has been derived.

□

Lemma 6. For any constant $\delta > 0$, after n^δ normalized stages, Node 1 has a value of at most $\frac{n}{\sqrt{c \log n}}$ for some fixed $c > 0$.

Proof. Let $f(n) = \sqrt{c \log n}$, where c will be determined later. After k stages, let $n_1 \geq n_2 \geq n_3 \cdots \geq n_{k'}$ be all values of nodes greater than zero. Clearly, $k' \leq k$ since each stage can only make one node have a value of zero for a value greater than zero in the linear graph. Let $d_i = n_i - n_{i+1}$. Assume that $\sum_{i=1}^k d_i \geq \frac{n}{cf(n)}$.

Let each node i take an average operation with one of the neighbors $j \in \{i - 1, i + 1\}$ such that i and j do not take the average in the last stage. In two consecutive normalized stages $k - 1$ and k , the potential of the nodes is reduced by at least $\sum_{i=1}^k (\frac{d_i}{2})^2 \geq \frac{(\sum_{i=1}^k d_i/2)^2}{k} \geq \frac{n^2}{4kf(n)^2}$ by the Cauchy-Schwarz inequality.

After $h = n^\delta$ stages, the potential is reduced by at least $\sum_{k=1}^h \frac{n^2}{4kf(n)^2} \geq \frac{n^2}{4f(n)^2} \cdot (\sum_{k=1}^h \frac{1}{k}) \geq \frac{n^2 \log h}{4df(n)^2}$ for some fixed positive real d since $\sum_{k=1}^h \frac{1}{k} = \Phi(\log h)$. Therefore, for $h = n^\delta$, $\frac{\log h}{4df(n)^2} \geq 1$ by selecting a fixed c to be small enough. Since the maximum potential is at most n^2 , which is the beginning of the system for $n_1 = n$ and $n_i = 0$ for all $i > 1$, after at most $h = n^\delta$ normalized stages, the value at Node 1 is $n_1 = \sum_{i=1}^k d_i < \frac{n}{f(n)}$. □

Theorem 5. It requires $\Omega(\max(n, m) \sqrt{\log \max(n, m)})$ stages to enter an ϵ -balance in a linear graph model.

Proof. Assume that Node 1 starts with the largest value n . Let $m = n/8$. Let $f(n) = \sqrt{c \log n}$, where c is the same as that in Lemma 6. By Lemma 6, after $n^\delta < m/2$ normalized stages, we have $n_1 \leq \frac{n}{f(n)}$. We still have at least $m/2$ nodes with zero values. We can assume that all stages are normalized stages by Lemma 5. This is because the normalized stages give the maximum flow from Node 1 to others. By Lemma 4, $d_1 \geq d_2 \geq \cdots$. In the rest of the stages, each stage sends at most $\frac{n_1}{m/2} \leq \frac{16}{f(n)}$ values to the area that contains the last $\frac{m}{2}$ nodes with zero values. When the system enters ϵ -balance

for $\epsilon < 1$, each node must have a value of at least four. Thus, the total number of stages is at least $4 \cdot \frac{m}{2} \cdot \frac{1}{f(n)} = \frac{1}{8} \cdot mf(n) = \Omega(n\sqrt{\log n}) = \Omega(\max(n, m)\sqrt{\log \max(n, m)})$.
 \square

5. Performance Evaluation

In this section, we first introduce the simulation environment, then present performance metrics and finally give a demonstration of the simulation results.

5.1. Simulation Setup

To evaluate the performance of the proposed replication algorithm, we have conducted extensive simulations. In the simulations, the following default settings are used.

Compromised by the complexity of simulations, we select a bounded area on the map of Los Angeles for the simulations. We extract a 2000 m \times 2000 m rectangle street area through OpenStreetMap [57], the satellite map of which is presented in Figure 3a. Then, we use Simulation of Urban Mobility (SUMO) [58] to convert the extracted area to the road networks. Figure 3b shows the converted road network of the selected area in Los Angeles. The realistic mobility trace of vehicles is generated by the open-source microscopic space-continuous and time-discrete vehicular traffic generator package SUMO. SUMO uses a collision-free car-following model to determine the speeds and the positions of the vehicles. The output from SUMO is converted into input files for the movement of nodes in the NS-3 simulator. As for the V2V communication, the Nakagami- m channel model is applied. We set the value of m according to the communication distance. When the distance is less than 50 m, m is set as one. When the distance is in the range of [50, 150], m is set as 1.5. When the distance is larger than 150 m, m is set as two.

The coverage of V2V communications is set as 300 m. The transmission frame duration is set as 1 ms. The average encounter duration is related to the vehicle's velocity and density. The MAC layer protocol follows 802.11p, with the distributed coordination function enabled. We refer to [38] and set the simulation time as one hour.

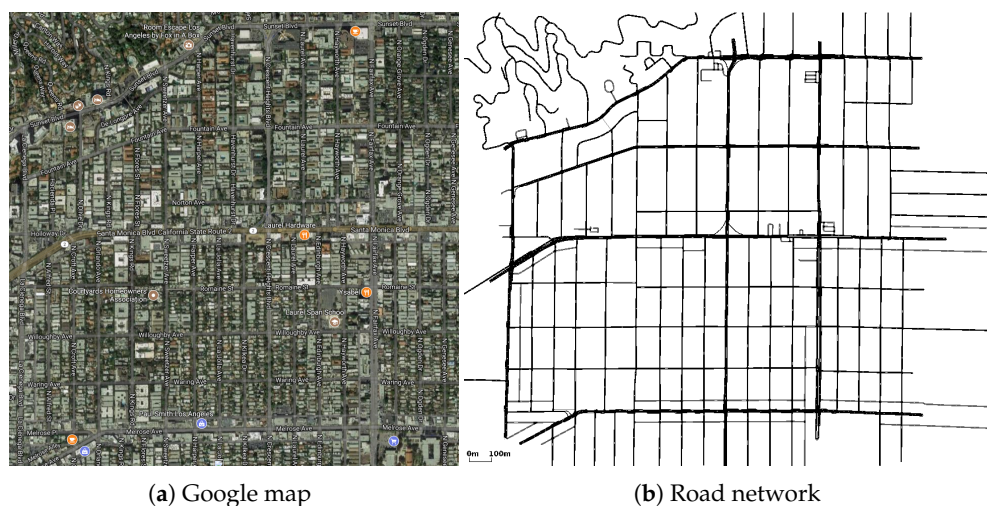


Figure 3. Los Angeles.

We conduct two sets of simulations, which are implemented in the urban and highway environment. We set up different numbers of vehicles to indicate normal traffic conditions and sparse traffic conditions, rather than a static number of vehicles. In each set, the replication limit of message M is indicated by parameter n ; in other words, the number of copies that can be spread is n . As for the value of parameter n , we apply the method in [11] for calculation. We also vary n to see

the effect of different numbers of message copies on network performance. Note that in the tradeoff analysis, all metrics are plotted as functions of the number of copies and the number of vehicles, respectively. We repeat the simulations and provide the results plotted with the confidence interval. The main experimental parameters are shown in Table 1.

Table 1. Parameters used in the simulations.

Parameter	Value
Simulation area	2000 m × 2000 m
Simulation time	1 h
Vehicle communication range	300 m
Vehicle velocity	[30, 120] km/h
Number of Vehicles	100, 300, 500
Number of message copies	[50, 500]

5.2. Performance Metrics

In the simulation, we vary two parameters including the number of message copies and the number of vehicles in the VANET. The range of these parameters will be elaborated along with the performance analysis.

The performance is measured using the following criteria: data delivery ratio, transmissions and dissemination delay. The data delivery ratio is the ratio of the number of vehicles that receive the message to the total number of copies. We use the data delivery ratio to evaluate the dissemination performance of the compared algorithms. Transmissions are denoted by the number of stages (see Definition 5), which characterizes the complexity for the system to enter into a balanced status among a set of nodes. As the number of stages can reflect how many transmissions are needed to achieve an ϵ -balanced status, it is utilized as a measurement of communication overhead to indicate how many average operations the system would perform when it reaches network consensus. We choose data dissemination delay as a measurement of effectiveness, which indicates the time interval from the beginning of data dissemination to the moment that the network enters into a balanced status.

Here, it is worth noting that in the compared schemes, to mitigate the simulation complexity, the number of message copies n represents the maximum message copies that can be spread in the process of data dissemination. In replication-based algorithms, such as epidemic routing, SW [35] and capacity-constrained replication algorithm (CCR) [11], n is set as the replication limit of the message. As for greedy perimeter stateless routing (GPSR), a single-copy routing algorithm, we use n to denote the the maximum number of nodes that can receive the message, which increases when the node carries a message meeting a node without the message.

5.3. Data Delivery Ratio

In this section, we compare the performance of the proposed algorithm EDDA with other data dissemination schemes, such as GPSR, epidemic routing, SW and CCR, in terms of data delivery ratio. We also show how the data delivery ratio is affected by the number of maximum message copies and the number of vehicles. Here, we use the number of vehicles to indicate the network size.

Figures 4 and 5 show the delivery ratio as a function of the number of message copies when the number of vehicles is fixed as 300. It compares data dissemination performance under different numbers of message replicas in the network. Figure 4 shows the results in a regular urban environment, and Figure 5 describes the case of a highway. It is easy to see that when the number of message copies is low, the epidemic algorithm achieves the best delivery ratio. However, as the number of message copies increases in the network, the network traffic dramatically grows in epidemic routing, as well as the number of data collisions increases and the data delivery ratio reduces. As a result, the performance of epidemic routing decreases and gradually underperforms the other algorithms. On the contrary, EDDA outperforms the other compared algorithms when more messages are spread in the network.

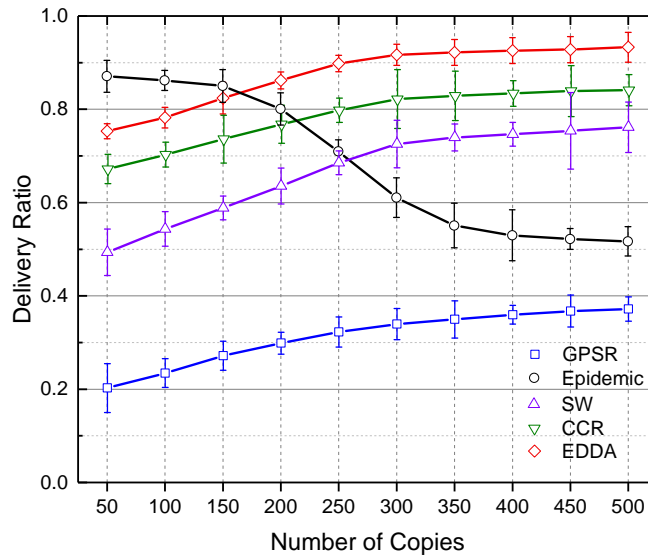


Figure 4. Delivery ratio vs. number of message copies in urban VANET.

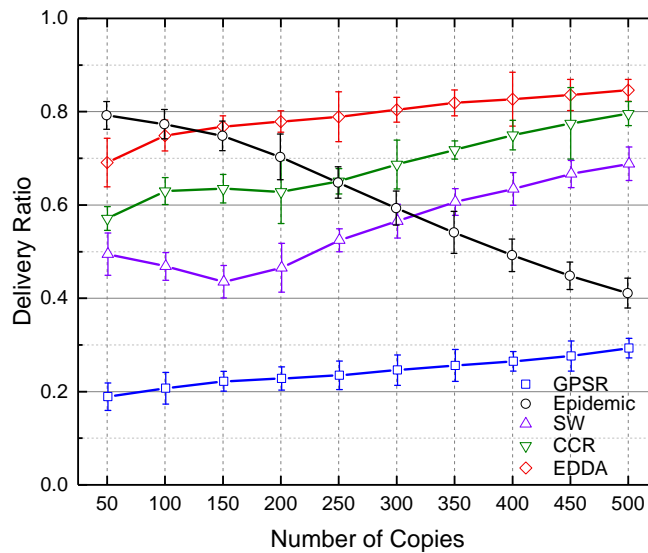


Figure 5. Delivery ratio vs. number of message copies on the highway.

Figures 6 and 7 show the delivery ratio as a function of the number of vehicles. In the simulations, we fix the maximum message copies that can be replicated as 300, and the number of vehicles is set as 100, 300 and 500. Figure 6 shows the results in urban VANET, while Figure 7 describes the results of highway. We can see that the delivery ratio increases when there are more vehicles in the area. This is because when there are very few vehicles on the roads, it may be hard to find the nearby forwarding vehicles, and hence, messages are dropped after the wait time. When there are more vehicles on the roads, the network connectivity becomes better as the vehicle density is higher, and the frequent node mobility will help carry and forward the packets that temporarily reach the sparse area.

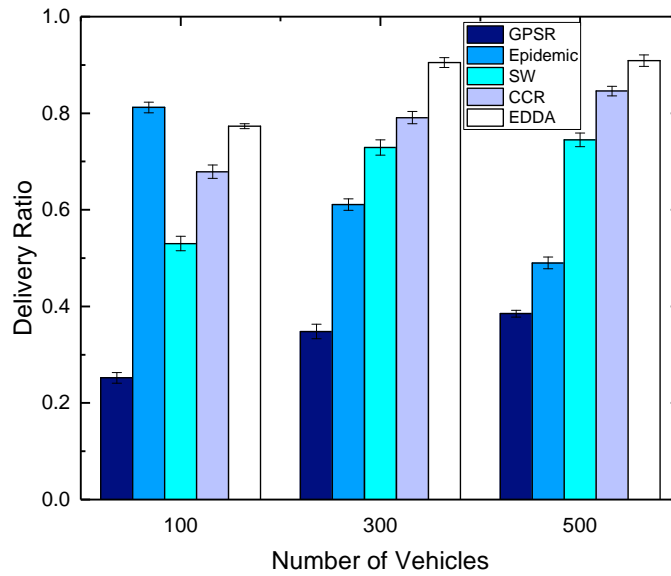


Figure 6. Delivery ratio vs. number of vehicles in urban VANET.

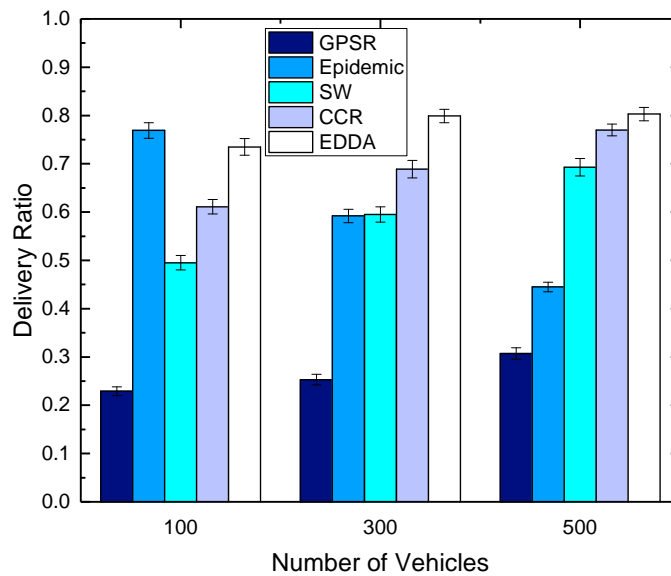


Figure 7. Delivery ratio vs. number of vehicles on the highway.

5.4. Transmissions

Figures 8 and 9 respectively depict the total number of transmissions for all compared algorithms when the number of message copies varies from 100–500 under the circumstances of urban and highway. As is evident from these two figures, EDDA performs significantly fewer transmissions than other algorithms. As the number of message copies increases, data transmissions will increase rapidly until the network converges, thus more redundant transmissions would occur. Especially, GPSR and epidemic need more transmissions to achieve ideal data dissemination performance. SW sprays first and then waits for another spray, which results in fewer transmissions. Likewise, CCR can also achieve better performance by taking advantage of the available network capacity. Benefiting from the bounded number of message copies and distributed averaging operations, EDDA outperforms the other compared schemes.

Figures 10 and 11 show the impact of the number of vehicles on transmissions for the compared schemes in urban and highway, respectively. It is obvious to see that the number of transmissions increases greatly when the vehicle density is larger. This is because when more vehicles participate in data dissemination, more communications happen among vehicles. However, the trends of the transmission amount for all the compared algorithms are similar. EDDA always consumes fewer transmissions when compared with other algorithms. We can get such a result that the proposed algorithm can achieve better results even in a traffic scenario with lower vehicle density.

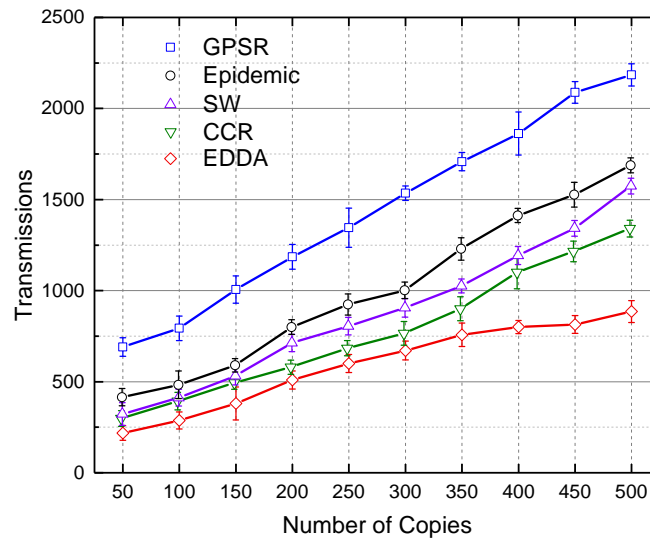


Figure 8. Transmissions vs. number of message copies in urban VANET.

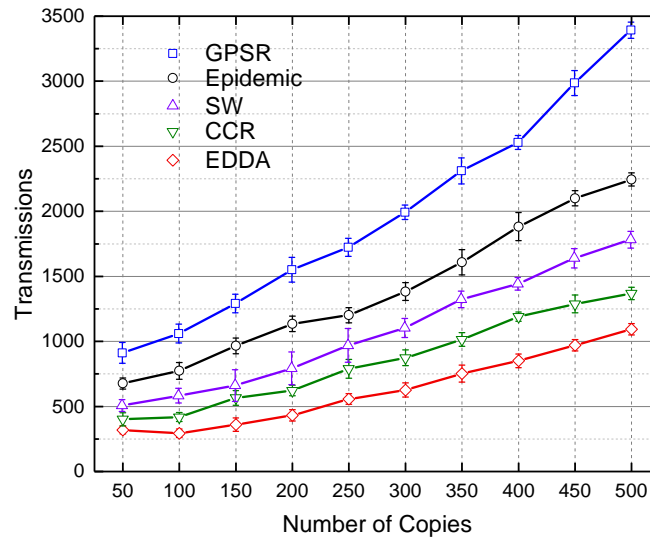


Figure 9. Transmissions vs. number of message copies on the highway.

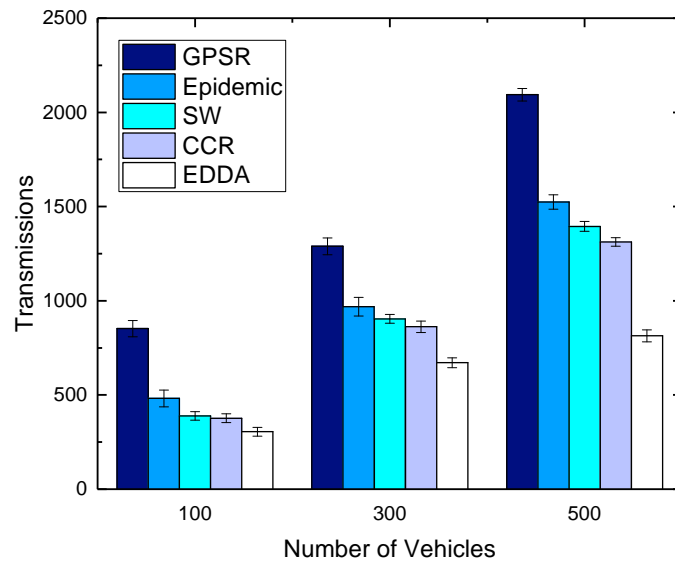


Figure 10. Transmissions vs. number of vehicles in urban VANET.

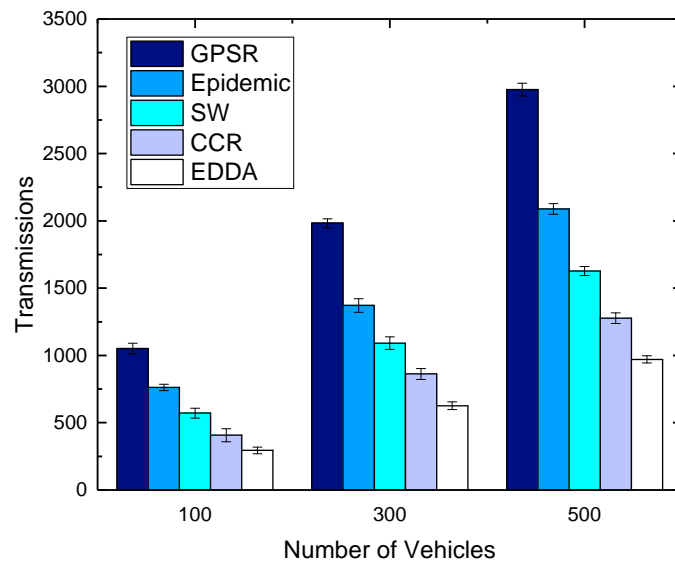


Figure 11. Transmissions vs. number of vehicles on the highway.

5.5. Data Dissemination Delay

In this section, we compare the data dissemination delay of the compared algorithms. The compared algorithms are GPSR, the epidemic routing scheme, SW and CCR. We will show how data dissemination delay changes when the number of message copies varies from 100–500 in Figures 12 and 13. Figure 12 shows the results in the regular urban environment, and Figure 13 describes the situation of the highway. When the number of message copies increases, the dissemination time increases. The possible reason is that the more messages to be delivered, the more data transmissions occur until the network converges, hence causing the increasing dissemination delay. We again observe that our proposed EDDA achieves better dissemination delay.

Figures 14 and 15 compare the effect of different vehicle densities when the number of message copies is equal to 300. From the differences in the figures, we can see that data dissemination delay increases with the increasing vehicle nodes. In the scenarios when the number of vehicles increases, the source vehicle typically needs to make multiple attempts to disseminate the messages to other

vehicles in the destination area, which increases the overall dissemination time. Epidemic routing and the rest of the schemes manage to achieve good delays for low vehicle density, but perform poorly for most values. GPSR has a relatively low dissemination delay in the sparse environment, and the delay will increase when the network becomes dense. On the other hand, due to the limitation that the number of message copies that can be spread is bounded, SW, CCR and EDDA exhibit greater stability. They perform fewer transmissions in all scenarios, while achieving slightly increased dissemination delays as the level of network connectivity increases. As is evident from the figures, in terms of dissemination delay, if traffic loads are low, it is known that most of the schemes can obtain good delays under these conditions. Nevertheless, if traffic starts increasing, EDDA actually outperforms all schemes in terms of delay.

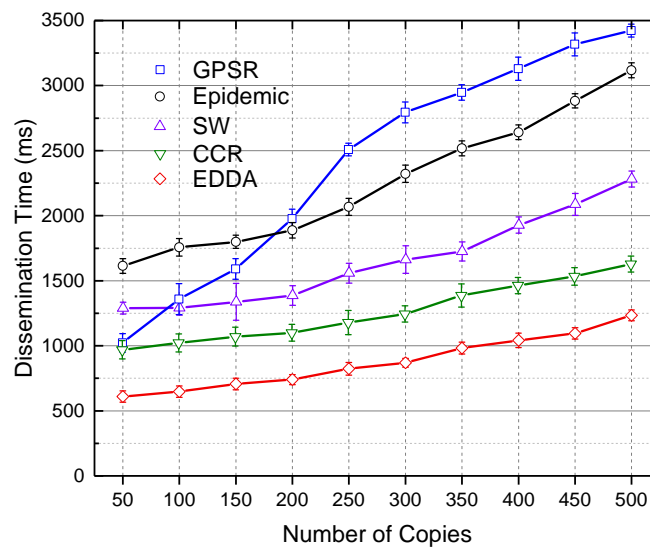


Figure 12. Dissemination delay vs. number of message copies in Urban VANET.

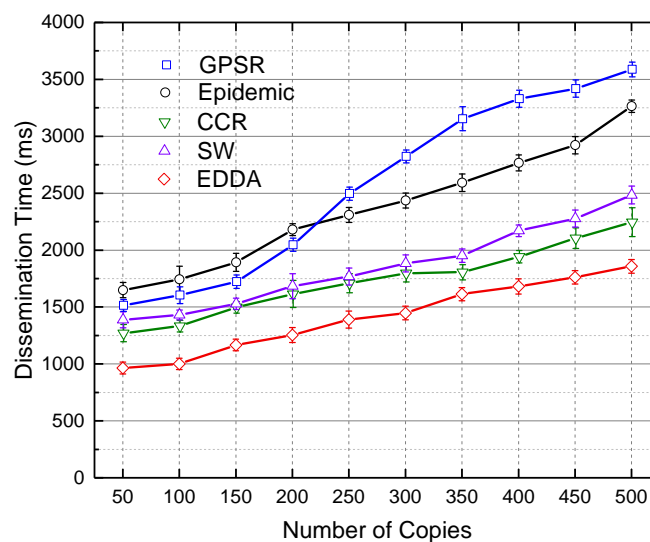


Figure 13. Dissemination delay vs. number of message copies on the highway.

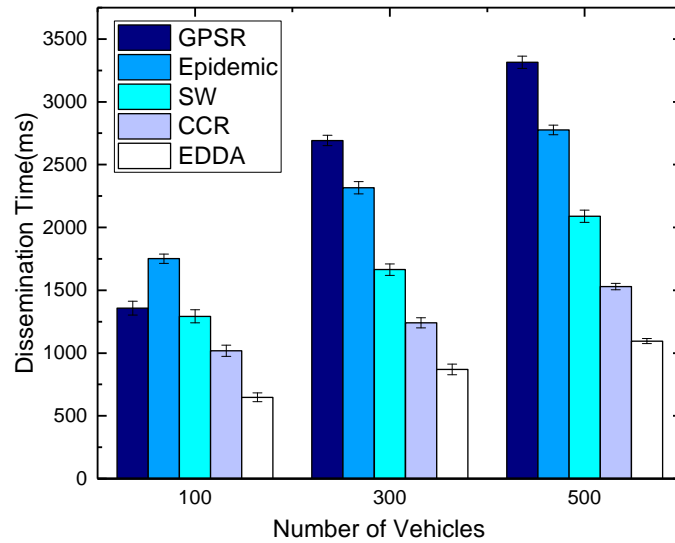


Figure 14. Dissemination delay vs. number of vehicles in urban VANET.

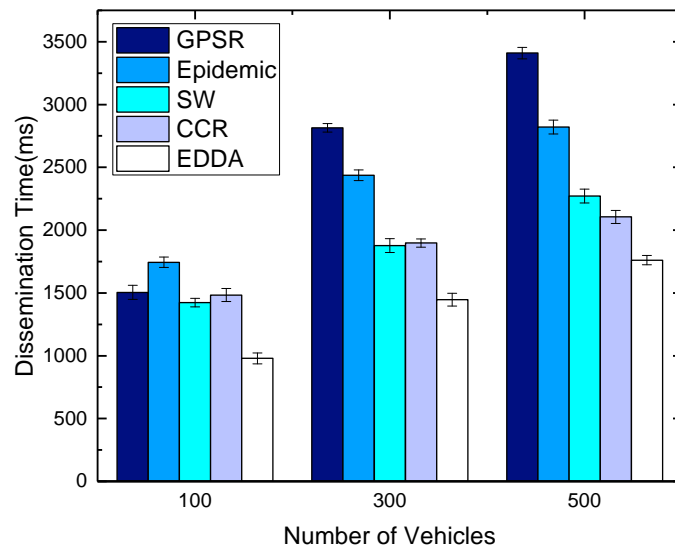


Figure 15. Dissemination delay vs. number of vehicles on the highway.

5.6. Evaluation of Network Balance on the Highway

In this section, we particularly discuss the situation when the algorithm is applied to the highway environment. We will see the differences between the delay for the network to enter into a balanced status and the delay for the network to rebalance if a new node enters the network and breaks the previous balance. Note that we assume vehicles do not take over each other on the highway in order to maintain an order of vehicles and simplify the problem.

Here, we use two parameters t_1 and t_2 to indicate the delays. t_1 indicates the time for the system to enter into balance. Then, we consider the situation that a new vehicle node joins the balanced system. Obviously, the system would need to do additional average operations to be rebalanced. t_2 is used to indicate the time for the system to be rebalanced.

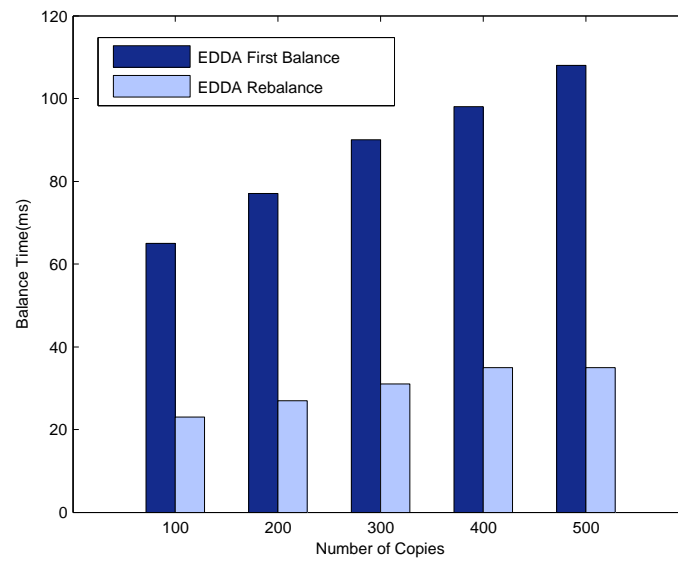


Figure 16. Time vs. number of message copies when the number of vehicle nodes is 50.

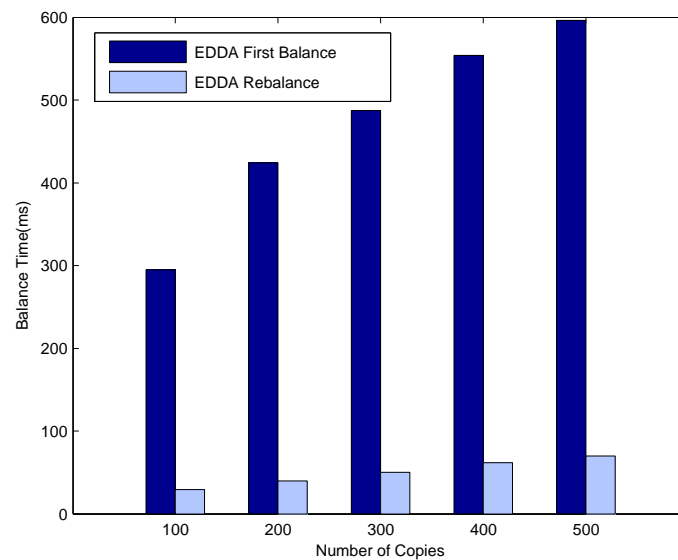


Figure 17. Time vs. number of message copies when the number of vehicle nodes is 100.

Figures 16–18 respectively describe the delays for the system to enter into a balanced status when the number of maximum allowed message copies varies from 100–500. The number of vehicle nodes is 50 in Figure 16, 100 in Figure 17 and 150 in Figure 18. As is evident from the three figures, both t_1 and t_2 increase when the maximum number of message copies increases from 100–500. Besides, we can easily see that t_2 is much smaller than t_1 . This is because when new nodes join the network, they will break the system balance, and the system needs to do additional average operations to be balanced again. Nevertheless, it needs fewer average operations to obtain system rebalance than the initial balance status.

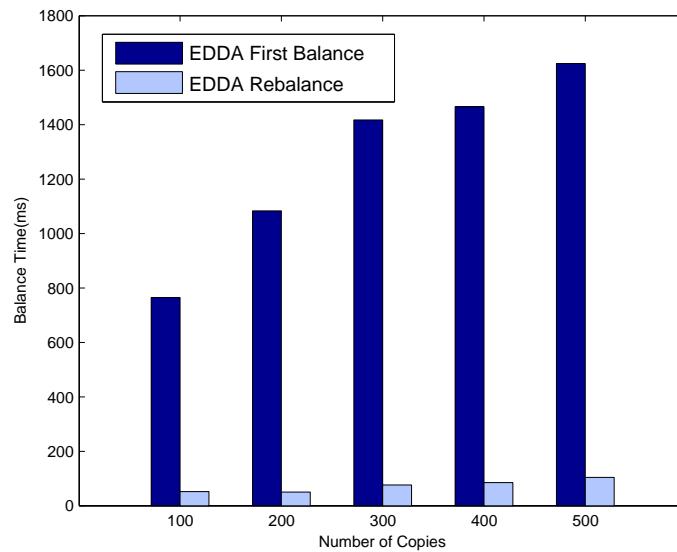


Figure 18. Time vs. number of message copies when the number of vehicle nodes is 150.

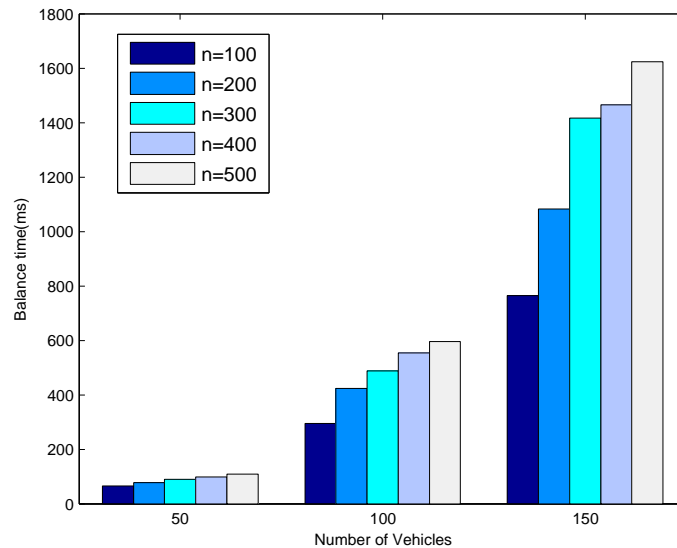


Figure 19. t_1 VS. number of vehicles.

Figure 19 shows the delays for the system to enter into a balanced status when the number of vehicle nodes varies from 50, to 100, to 150. The figure also provides a comparison of the delays when the number of message copies is 100, 200, 300, 400 and 500. As can be seen from the figure, when the number of message copies is fixed, the time to balance the system increases as the number of vehicle nodes increases since it needs to do more average operations.

6. Conclusions

To facilitate data dissemination in VANETs, we investigate distributed data replication algorithms. We use graph theory to describe the network topology and discuss the cases of arbitrary graph and linear graph. Then, we propose a distributed data replication algorithm with a bounded number of message copies disseminated to an area of interest. Moreover, we measure the complexity of network convergence by the number of communication stages in a distributed computing environment. We prove that there are at least $\min(\frac{n}{2+\epsilon}, m)$ nodes that can receive the message when the system enters into an ϵ -balanced status. Finally, detailed analysis of the convergence speed of the proposed algorithm,

as well as the upper bound and lower bound are given. It shows that through pairwise average operations, the network can enter into a balanced status in a quick manner. Simulations driven by real traces in Los Angeles have been done to demonstrate the effectiveness and the superior performance of the proposed solution, which results in a substantial dissemination delay reduction and decreasing communication overhead compared with the state-of-the-art solutions.

Acknowledgments: This work is supported by the National Science Foundation of China (No. 61772385, No. 61373040, No. 61572370), by National Science Foundation Early Career Award 0845376 and the Bensten Fellowship of the University of Texas-Rio Grande Valley.

Author Contributions: This paper presents part of Junyu Zhu's Ph.D. study research. Junyu Zhu originated this work and drafted the manuscript, performed the experiment and analyzed the experiment results. Chuanhe Huang and Bin Fu contributed to the theory studies. Sipei Guo and Xiyang Fan helped refine the algorithm and improve the quality of this work. All authors read and approved the final manuscript.

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

Abbreviations

The following abbreviations are used in this manuscript:

VANETs	Vehicular ad-hoc networks
V2V	Vehicle-to-vehicle
RSU	Road side unit
EDDA	Efficient distributed data replication algorithm

References

1. Cunha, F.; Villas, L.; Boukerche, A.; Maia, G.; Viana, A.; Mini, R.A.F.; Loureiro, A.A.F. Data communication in VANETs: Protocols, applications and challenges. *Ad Hoc Netw.* **2016**, *44*, 104–117, doi:10.1016/j.adhoc.2016.02.017.
2. Mukherjee, J.C.; Gupta, A.; Sreenivas, R.C. Event Notification in VANET With Capacitated Roadside Units. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 1867–1879, doi:10.1109/TITS.2015.2511145.
3. Said, A.M.; Marot, M.; Ibrahim, A.W.; Afifi, H. Modeling interactive real-time applications in VANETs with performance evaluation. *Comput. Netw.* **2016**, *104*, 66–78, doi:10.1016/j.comnet.2016.04.021.
4. Hu, M.; Zhong, Z.; Ni, M.; Wang, Z.; Xie, W.; Qiao, X. Integrity-oriented Content Offloading in Vehicular Sensor Network. *IEEE Access.* **2017**, *5*, 4140–4153, doi:10.1109/ACCESS.2017.2683640.
5. Kosch, T. Local danger warning based on vehicle ad-hoc networks: Prototype and simulation. In Proceedings of the First International Workshop on Intelligent Transportation (WIT), Paderborn, Germany, 5–8 July 2016.
6. Nandan, A.; Das, S.; Zhou, B.; Pau, G.; Gerla, M. Adtorrent: Digital billboards for vehicular networks. In Proceedings of the IEEE/ACM International Workshop on Vehicle-to-Vehicle Communications(V2VCOM), San Diego, CA, July, 2005; pp. 286–294.
7. Kangasharju, J.; Heinemann, A. Incentives for electronic coupon systems. In Proceedings of the International Workshop on Decentralized Resource Sharing in Mobile Computing and Networking, Los Angeles, CA, USA, 25 July 2006; pp. 60–62, doi:10.1145/1161252.1161268.
8. Lee, U.; Park, J.S.; Amir, E.; Gerla, M. FleaNet: A Virtual Market Place on Vehicular Networks. *IEEE Trans. Veh. Technol.* **2010**, *59*, 344–355, doi:10.1109/MOBIOQ.2006.340433.
9. Liu, F.; Chen, Z.; Xia, B. Data Dissemination With Network Coding in Two-Way Vehicle-to-Vehicle Networks. *IEEE Trans. Veh. Technol.* **2016**, *65*, 2445–2456, doi:10.1109/TVT.2015.2424915.
10. He, J.; Cai, L.; Cheng, P.; Pan, J. Delay Minimization for Data Dissemination in Large-Scale VANETs with Buses and Taxis. *IEEE Trans. Mob. Comput.* **2016**, *15*, 1939–1950, doi:10.1109/TMC.2015.2480062.
11. Wu, Y.C.; Zhu, Y.; Zhu, H.; Li, B. CCR: Capacity-constrained replication for data delivery in vehicular networks. In Proceedings of the 2013 IEEE INFOCOM, Turin, Italy, 14–19 April 2013; pp. 2580–2588.

12. Janech, J.; Lieskovsky, A.; Krsak, E. Comparison of strategies for data replication in VANET environment. In Proceedings of the International Conference on Advanced Information Networking and Applications Workshops, Fukuoka, Japan, 26–29 March 2012; Volume 26, pp. 575–580, doi:10.1109/WAINA.2012.179.
13. Rajagopalan, S.; Shah, D. Distributed averaging in dynamic networks. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 845–854, doi:10.1109/JSTSP.2011.2114635.
14. Hadjicostis, C.N.; Vaidya, N.H.; Domínguez-García, A.D. Robust distributed average consensus via exchange of running sums. *IEEE Trans. Autom. Control.* **2016**, *61*, 1492–1507, doi:10.1109/TAC.2015.2471695.
15. Dimakis, A.G.; Sarwate, A.D.; Wainwright, M.J. Geographic gossip: Efficient aggregation for sensor networks. In Proceedings of the ACM International Conference on Information Processing in Sensor Networks, Nashville, TN, USA, 19–21 April 2006; pp. 69–76, doi:10.1145/1127777.1127791.
16. Xing, M.; He, J.; Cai, L. Utility Maximization for Multimedia Data Dissemination in Large-Scale VANETs. *IEEE Trans. Mob. Comput.* **2017**, *16*, 1188–1198, doi:10.1109/TMC.2016.2582482.
17. Hajlaoui, R.; Guyennet, H.; Moulahi, T. A Survey on Heuristic-Based Routing Methods in Vehicular Ad-Hoc Network: Technical Challenges and Future Trends. *IEEE Sens. J.* **2016**, *16*, 6782–6792, doi:10.1109/JSEN.2016.2583382.
18. Ye, F.; Roy, S.; Wang, H. Efficient Data Dissemination in Vehicular Ad Hoc Networks. *IEEE J. Sel. Areas Commun.* **2012**, *30*, 769–779, doi:10.1109/JSAC.2012.120511.
19. Wu, C.; Chen, X.; Ji, Y.; Ohzahata, S.; Ji, Y. Efficient Broadcasting in VANETs Using Dynamic Backbone and Network Coding. *IEEE Trans Wirel. Commun.* **2015**, *14*, 6057–6071, doi:10.1109/TWC.2015.2447812.
20. Torres, A.; Calafate, C.T.; Cano, J.C.; Manzoni, P.; Ji, Y. Evaluation of flooding schemes for real-time video transmission in VANETs. *Ad Hoc Netw.* **2015**, *24*, 3–20, doi:10.1016/j.adhoc.2014.07.030.
21. Yang, Y.; Zhao, C.; Yao, S.; Zhang, W.X.; Ge, X.H.; Mao, G.Q. Delay Performance of Network-Coding-Based Epidemic Routing. *IEEE Trans. Veh. Technol.* **2016**, *65*, 3676–3684, doi:10.1109/TVT.2015.2433263.
22. Taherkhani, N.; Pierre, S. Prioritizing and scheduling messages for congestion control in vehicular ad hoc networks. *Comput. Netw.* **2016**, *108*, 15–28, doi:10.1016/j.comnet.2016.06.027.
23. Goonewardene, R.T.; Ali, F.H.; Stipidis, E. Robust mobility adaptive clustering scheme with support for geographic routing for vehicular ad hoc networks. *IET Intell. Transp. Syst.* **2009**, *3*, 148–158, doi:10.1049/iet-its:20070052.
24. Luo, Y.Y.; Zhang, W.; Hu, Y.Q. A New Cluster Based Routing Protocol for VANET. In Proceedings of the Second International Conference on Networks Security, Wireless Communications and Trusted Computing, Wuhan, China, 24–25 April 2010; pp. 176–180, doi:10.1109/NSWCTC.2010.48.
25. Ohta, Y.; Ohta, T.; Kakuda, Y. An autonomous clustering based data transfer scheme using positions and moving direction of vehicles for VANETs. In Proceedings of the 2012 IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, China, 1–4 April 2012; pp. 2900–2904, doi:10.1109/WCNC.2012.6214299.
26. Lin, D.; Kang, J.; Squicciarini, A.; Wu, Y.J.; Gurung, S.; Tonguz, O. MoZo: A Moving Zone Based Routing Protocol Using Pure V2V Communication in VANETs. *IEEE Trans. Mob. Comput.* **2017**, *16*, 1357–1370, doi:10.1109/TMC.2016.2592915.
27. Kumer, N.; Rodrigues, J.; Lloret, J. Replication-aware data dissemination for vehicular ad hoc networks using location determine. *Mob. Netw. Appl.* **2015**, *20*, 251–267, doi:10.1007/s11036-015-0572-9.
28. Bi, Y.G.; Shan, H.G.; Shen, X.M.; Wang, N.; Zhao, H. A multi-hop broadcast protocol for emergency message dissemination in urban vehicular ad hoc networks. *IEEE Trans. Intell. Transport. Syst.* **2016**, *17*, 736–750, doi:10.1109/TITS.2015.2481486.
29. Rivoirard, L.; Wahl, M.; Sonni, P.; Berbineau, M.; Gruyer, D. Chain-Branch-Leaf: A Clustering Scheme for Vehicular Networks Using Only V2V Communications. *Ad Hoc Netw.* **2017**, *68*, 70–84, doi:10.1016/j.adhoc.2017.10.007.
30. Nilsson, M.; Gustafson, C.; Abbas, T.; Tufvesson, F. A Measurement-Based Multilink Shadowing Model for V2V Network Simulations of Highway Scenarios. *IEEE Trans. Veh. Technol.* **2017**, *66*, 8632–8643, doi:10.1109/TVT.2017.2709258.
31. Basheer, H.S.; Bassil, C. A review of broadcasting safety data in V2V: Weaknesses and requirements. *Ad Hoc Netw.* **2017**, *65*, 13–25, doi:10.1016/j.adhoc.2017.07.004.

32. Vahdat, A.; Becker, D. Epidemic routing for partially connected ad hoc networks. Master's Thesis, University of Kansas, Lawrence, KS, USA, 2000; pp. 1–14.
33. Lindgren, A.; Doria, A.; Scheln, O. Probabilistic routing in intermittently connected networks. In Proceedings of the Fourth ACM International Symposium on Mobile Ad Hoc Networking and Computing, New York, NY, USA, 2003; Volume 7, pp. 19–20, doi:10.1145/961268.961272.
34. Burgess, J.; Gallagher, B.; Jensen, D.; Levine, N.B. MaxProp: Routing for vehicle-based disruption-tolerant networks. In Proceedings of the 25TH IEEE International Conference on Computer Communications, IEEE INFOCOM 2006, Barcelona, Spain, 23–29 April 2006, doi:10.1109/INFOCOM.2006.228.
35. Spyropoulos, T.; Psounis, K.; Raghavendra, C.S. Efficient routing in intermittently connected mobile networks: The multiple-copy case. *IEEE/ACM Trans. Netw.* **2008**, *16*, 77–90, doi:10.1109/TNET.2007.897964.
36. Balasubramanian, A.; Levine, B.N.; Venkataramani, A. Replication routing in DTNs: A resource allocation approach. *IEEE/ACM Trans. Netw.* **2010**, *18*, 596–609, doi:10.1109/TNET.2009.2036365.
37. Akabane, A.T.; Villas, L.A.; Madeira, E.R.M. An adaptive solution for data dissemination under diverse road traffic conditions in urban scenarios. In Proceedings of the IEEE Wireless Communications and Networking Conference, New Orleans, LA, USA, 9–12 March 2015; pp. 1654–1659, doi:10.1109/WCNC.2015.7127716.
38. Yan, T.; Zhang, W.S.; Wang, G.L. DOVE: Data dissemination to a desired number of receivers in VANET. *IEEE Trans. Veh. Technol.* **2014**, *63*, 1903–1916, doi:10.1109/TVT.2013.2287692.
39. Takahashi, A.; Nishiyama, H.; Kato, N.; Nakahira, K.; Sugiyama, T. Replication control for ensuring reliability of convergecast message delivery in infrastructure-aided dtns. *IEEE Trans. Veh. Technol.* **2014**, *63*, 3223–3231, doi:10.1109/TVT.2014.2299288.
40. Li, W.F.; Ma, X.M.; Wu, J.; Trivedi, K.S.; Huang, X.L.; Liu, Q.W. Analytical Model and Performance evaluation of Long Term Evolution for vehicle Safety Services. *IEEE Trans. Veh. Technol.* **2017**, *66*, 1926–1939, doi:10.1109/TVT.2016.2580571.
41. Li, P.; Huang, C.; Liu, Q. BCDP: Budget constrained and delay-bounded placement for hybrid roadside units in vehicular ad hoc networks. *Sensors* **2014**, *14*, 22564–22594, doi:10.3390/s141222564.
42. Chen, J.; Mao, G.; Li, C.; Liang, W.; Zhang, D. Capacity of Cooperative Vehicular Networks with Infrastructure Support: Multi-user Case. *IEEE Trans. Veh. Technol.* **2017**, doi:10.1109/TVT.2017.2753772.
43. Li, G.; Boukhatem, L.; Wu, J. Adaptive Quality-of-Service-Based Routing for Vehicular Ad Hoc Networks With Ant Colony Optimization. *IEEE Trans. Veh. Technol.* **2016**, *66*, 3249–3264, doi:10.1109/TVT.2016.2586382.
44. Tsitsiklis, J.N. Problems in decentralized decision making and computation. Ph.D. Thesis, Massachusetts Institute of Technology, Boston, MA, USA 1984.
45. Carli, R.; Fagnanib, F. Communication constraints in the average consensus problem. *Automatica* **2008**, *44*, 671–684.
46. Hadjicostis, C.N.; Charalambous, T. Average consensus in the presence of delays in directed graph topologies. *IEEE Trans. Autom. Control.* **2014**, *59*, 763–768, doi:10.1109/TAC.2013.2275669.
47. Nedic, A.; Olshevsky, A.; Ozdaglar, A.; Tsitsiklis, J.N. On Distributed Averaging Algorithms and Quantization Effects. *IEEE Trans. Autom. Control.* **2009**, *54*, 2506–2517, doi:10.1109/TAC.2009.2031203.
48. Khosravi, A.; Kaviani, Y.S. Broadcast Gossip Ratio Consensus: Asynchronous Distributed Averaging in Strongly Connected Networks. *IEEE Trans. Signal Process.* **2017**, *65*, 119–129, doi:10.1109/TSP.2016.2614790.
49. Nedić, A.; Ozdaglar, A. Convergence rate for consensus with delays. *J. Glob. Optim.* **2010**, *47*, 437–456, doi:10.1109/CDC.2008.4738941.
50. Fagnani, F.; Zampieri, S. Randomized consensus algorithms over large scale networks. *IEEE J. Sel. Areas Commun.* **2008**, *26*, 634–649, doi:10.1109/JSAC.2008.080506.
51. Aysal, T.C.; Yildiz, M.E.; Sarwate, A.D.; Scaglione, A. Broadcast Gossip Algorithms for Consensus. *IEEE Trans. Signal Process.* **2009**, *57*, 2748–2761, doi:10.1109/TSP.2009.2016247.
52. Boyd, S.; Ghosh, A.; Prabhakar, B.; Shah, D. Randomized gossip algorithms. *IEEE Trans. Inf. Theory* **2006**, *52*, 2508–2530, doi:10.1109/TIT.2006.874516.
53. Wu, S.C.; Rabbat, M.G. Broadcast Gossip Algorithms for Consensus on Strongly Connected Digraphs. *IEEE Trans. Signal Process.* **2013**, *61*, 3959–3971, doi:10.1109/TSP.2013.2264056.
54. Nedić, A.; Ozdaglar, A.; Parrilo, P.A. Constrained Consensus and Optimization in Multi-Agent Networks. *IEEE Trans. Autom. Control.* **2010**, *55*, 922–938, doi:10.1109/TAC.2010.2041686.
55. Nedić, A.; Liu, J. On Convergence Rate of Weighted-Averaging Dynamics for Consensus Problems. *IEEE Trans. Autom. Control.* **2017**, *62*, 766–781, doi:10.1109/TAC.2016.2572004.

56. Shi, G.D.; Li, B.; Johansson, M.; Johansson, K.H. Finite-Time Convergent Gossiping. *IEEE/ACM Trans. Netw.* **2016**, *24*, 2782–2794, doi:10.1109/TNET.2015.2484345.
57. Haklay, M.; Weber, P. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Comput.* **2008**, *7*, 12–18, doi:10.1109/MPRV.2008.80.
58. SUMO-Simulation of Urban Mobility. Available Online: <http://sumo.sourceforge.net> (accessed on 1 January 2001).



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).