Analysis of Utility-Based Data Dissemination Approaches in VANETs

Ramon S. Schwartz*, Anthony E. Ohazulike†, Hylke W. van Dijk$^{\psi}$ and Hans Scholten*

Abstract— **By disseminating data through Vehicular Ad-hoc Networks (VANETs), vehicles are able to share relevant sensor data to acquire information about accidents, traffic, and even pollution. Data relevance is measured by a utility function which considers the contextual information that vehicles currently have about their environment. To be effective, data dissemination protocols must cope with intermittent connectivity due to the high speeds of vehicles. Problems arise when not all data can be exchanged due to the limited time available. In this paper, we explore and compare two fundamentally distinct approaches to tackling this problem. The first aims to maximize the system *efficiency*. In contrast, the second trades efficiency by a *fair* data distribution over vehicles by means of Nash Bargaining as used in game theory. By means of an extensive simulation campaign, an approach relying on fairness is shown to outperform efficiency in terms of delivery ratio, Jain's fairness index, sum of utility gains, number of hops and number of files downloaded.**

Keywords: Vehicular Ad-hoc Networks (VANETs), Vehicular Sensor Networks (VSNs), Data Dissemination

I. INTRODUCTION

The recent effort put into the use of wireless transmitters to enable vehicle-to-vehicle communication has resulted in a new type of network referred to as Vehicular Ad-hoc Network (VANET). In particular, Vehicular Sensor Networks (VSNs) rely on diverse sensors being built into modern vehicles in order for vehicles to continuously gather, process, and disseminate relevant sensor data. Examples of such sensors are chemical spill detectors, still/video cameras, and acoustic detectors. Possible applications that will use such networks range from traffic flow and pollution monitoring to safety warning systems.

Due to the high speeds of vehicles, the connectivity duration time for data exchange can be seriously limited, e.g., when vehicles are moving in opposite directions. In addition, the continuous collection of data can result in large amounts of data carried by each vehicle, specially considering sensor data of audio and video. In such a scenario, data dissemination protocols must incorporate efficient mechanisms to select the most relevant data to maximize the utility (benefit) gain of vehicles. Data utility is measured on the basis of available contextual information. In particular, we consider that every piece of data (file) is related to a certain geographical region and relevant to vehicles eventually approaching that region. For example, an accident is only relevant to vehicles approaching the area of hazard. In this context, two fundamentally distinct approaches can be identified in the literature. The first aims to maximize the system *efficiency*. Vehicles select data with the goal of maximizing the total utility gained by all vehicles in the vicinity [1]. The second focuses on a *fair* distribution of utility among vehicles [2].

While there is a clear trade-off between efficiency and fairness, the comparison between both approaches is still lacking. We will fill this gap by means of an extensive simulation campaign. We explore this trade-off in terms of various performance metrics, paying special attention to Jain's fairness index and the sum of utility gains.

The remainder of this paper is organized as follows. In Section II, we elaborate on the two data dissemination approaches that are to be compared. More specifically, we define a utility function which assigns a value to each file; propose one data selection model for each approach; and present a basic protocol to be used in our evaluation. Next, Section III analyzes the results obtained from our performance evaluation. Finally, Section IV concludes this paper with a general discussion and recommendations.

II. DATA DISSEMINATION APPROACHES

A data dissemination process can be organized in three alternating tasks, as we describe in detail in [3]. Opportunities for data exchange are constantly *discovered*, *assessed*, and chosen (*seized*). In particular, the assessment task consists of choosing the optimum vehicle to exchange data with and selecting the data (files) which maximizes the vehicles' utility gain. Here, our focus is on the data selection process. In time-limited data exchanges, the selection of data is crucial since this will determine the dissemination efficiency of the data distribution among vehicles.

A. Utility Function

We define utility as the benefit a vehicle can gain by exchanging a certain file. This can be either downloading or uploading. The utility is calculated based on the current level of contextual knowledge the vehicle has about the data available in its vicinity. For example, when a vehicle intends to send data to a specific destination, it must leverage the mobility of other vehicles to store, carry, and forward the data to its final destination. Similarly, when a vehicle becomes aware of the data being carried by other vehicles it evaluates the amount of interest it has in such data. We classify this contextual knowledge into the following two categories:

- ***Mobility context***: ranges from the complete route of a vehicle (e.g., a bus schedule) to only the vehicle direction, speed, mobility history, etc.

This work is supported by iLAND Project, ARTEMIS Joint Undertaking Call for proposals ARTEMIS-2008-1, Project contract no. 100026

Authors are with the faculties of Engineering Technology (CTW)† and Electrical Engineering, Mathematics and Computer Science (EWI)*, University of Twente, The Netherlands; and NHL Hogeschool$^{\psi}$, The Netherlands. Email: {`r.s.schwartz, a.e.ohazulike, hans.scholten`}`@utwente.nl; h.w.vandijk@nhl.nl`

- ***Data context***: includes the priority of the data, data size, age, geographical region, etc.

The relevance of this contextual information can be generalized in a so-called *utility* function which attributes a value $u_{i,j}$ to every data file, f_j, being carried by vehicle v_i and by others before its dissemination. The normalized utility function is given by:

$$u_{i,j} = \alpha_1 z_1^i(f_j) + \alpha_k z_2^i(f_j) +, ..., \alpha_l z_l^i(f_j). \qquad (1)$$

where $z_k^i \in [0,1]$ $(k = 1, 2, ..., l)$ is the function for each type of contextual information for vehicle i weighted by their respective parameters α_k.

In this paper, we assume that a file contains the following contextual information: geographical coordinates, generation time, and priority. The utility is calculated only when downloading files and considering three aspects:

Priority **(**α_1, $z_1^i(f_j)$**)**: we define three levels of priority for f_j related to its data, namely, emergency, traffic information, infotainment, returning respective z_1^i values of 1.0, 0.8, and 0.6. α_1 is set to 0.8.

Closest distance to a file's region **(**α_2, $z_2^i(f_j)$**)**:

$$z_2^i(f_j) = 1 - d^i(c_{f_j})/5000 \qquad (2)$$

where $d^i(c_{f_j}))$ is a function which calculates the shortest distance in meters to which the vehicle i approaches the file's geographical coordinates c_{fj}. The accuracy of this measure will depend on the current information the vehicle has about its route, e.g., complete route if it is set in a navigation system. Only distances within an area of 5 km^2 ($d(c_{f_j})) \leq 5000$) are considered, and α_2 is set to 0.15.

Data age **(**α_3, $z_3^i(f_j)$**)**:

$$z_3^i(f_j) = 0.999^{t_{fj}} \qquad (3)$$

where t_{fj} is the time elapsed since the file's generation time and α_3 is set to 0.05.

With the weights applied, most importance is given to the priority of a file. Clearly, different results can be expected when different contextual information and parameters are considered. This difference will be dependent on the application considered. Here, we define basic functions and parameters to assess the data selection approaches when a disparity of utility values exists between vehicles.

B. Data Selection Models

We evaluate two fundamentally distinct approaches for data selection which we refer to as *Total Sum Optimization (TSO)* and *Fair Sum Optimization (FSO)*. TSO aims to maximize the system *efficiency*. In [1] this is referred to as being an altruistic approach, since vehicles select data with only the goal of maximizing the total sum of utilities gained by all vehicles regardless of how much they profit individually. In contrast, FSO focuses on a *fair* distribution of utility among all vehicles. To achieve fairness, we rely on the Nash Bargaining [4] solution from game theory, which has been widely used in fields such as network bandwidth allocation and was recently proposed for use in vehicular networks for data exchange in [2].

Let U be the utility matrix for m vehicles and n files,

$$U = \begin{array}{c c} & \begin{array}{cccc} f_1 & f_2 & \cdots & f_n \end{array} \\ \begin{array}{c} v_1 \\ v_2 \\ \vdots \\ v_m \end{array} & \begin{pmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,n} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m,1} & u_{m,2} & \cdots & u_{m,n} \end{pmatrix} \end{array}. \qquad (4)$$

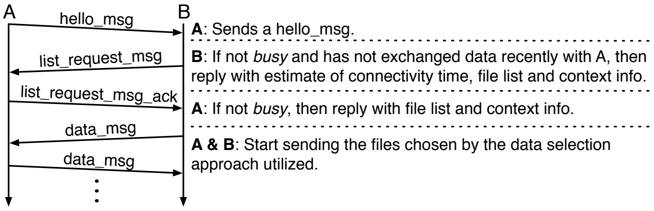
where $u_{i,j}$ is given by (1). TSO and FSO are defined respectively by (5) and (6). The binary vector $x = x_1, ..., x_n$ selects the files f_j which will be transmitted. Given the total data exchange duration time, the maximum number of files which can be transmitted is k_{max}.

$$\max \sum_{i=1}^{m} \sum_{j=1}^{n} [u_{ij} x_j] \quad (5) \qquad \max \prod_{i=1}^{m} \sum_{j=1}^{n} [u_{ij} x_j] \quad (6)$$

$$\text{with} \quad x_j \in \{0, 1\} \quad \text{s.t.} \quad \sum_{j=1}^{n} x_j \leq k_{max}.$$

The *Nash Bargaining* solution maximizes the product of the utility sum of each vehicle. In [4] it is proved that in a convex, closed and bounded set the solution is unique for the axioms: Pareto optimality, symmetry, scale covariance, and independence of irrelevant alternatives.

C. Basic Protocol



A: Sends a hello_msg.

B: If not *busy* and has not exchanged data recently with A, then reply with estimate of connectivity time, file list and context info.

A: If not *busy*, then reply with file list and context info.

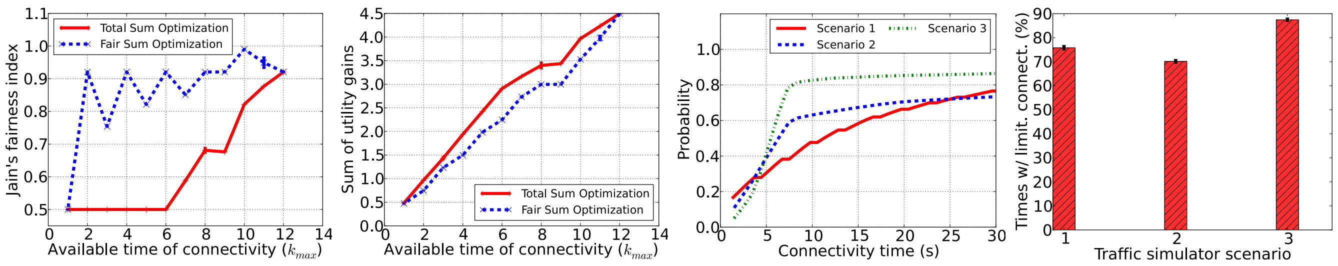
A & B: Start sending the files chosen by the data selection approach utilized.

Fig. 1. Basic protocol for data exchange between every pair of vehicle

To compare the data selection models, we define a basic protocol for the data exchange of every pair of vehicle. The decision to restrict the protocol to only two vehicles is because of the extra robustness achieved when using RTS/CTS of the 802.11p MAC layer standard for VANETs [5]. The basic protocol works as shown in Fig. 1. Vehicles which are not currently exchanging data (because they are busy) and which have not recently exchanged data with each other start a simple handshaking process after a periodic *hello* message is received. This involves the exchange of the list of files they possess and the contextual information used by utility function $u_{i,j}$. This guarantees that both vehicles generate the same utility matrix U. The connectivity time is estimated by vehicle *B* and included in *list_request_msg*. This estimation is based on the current route information available about both vehicles. Finally, the data selection model is run by both vehicles individually and files are exchanged by means of data messages.

III. PERFORMANCE EVALUATION

The performance evaluation of both TSO and FSO is carried out by means of simulations. Our goal is to understand



(a) Jain's fairness index x k_{max} (b) Sum of utility gains x k_{max} (c) CDF of the connectivity time (d) Times with limited connectivity

Fig. 2. Results for two-vehicle scenario in (a) and (b) with 95% confidence intervals; connectivity profile of the traffic simulator scenarios in (c) and (d)

the fundamental differences between the two approaches in various performance metrics and scenarios.

We utilize OMNeT++ 4.1 [6] with MiXiM v2.0.1. We adjust the available implementation of IEEE 802.11b to comply with basic specifications of the 802.11p version. In the MAC layer, we set the bit rate to 6 Mbit/s, the Contention Window (CW) to values between 15 and 1023, the slot time to 13 μs, the SIFS to 32 μs, and the DIFS to 58 μs. In the physical layer, we operate in the 5.9 GHz frequency band, using 10 MHz of bandwidth. Based on estimates, we set the transmission power to 168.98 mW to achieve approximately 500 meters of interference range and 250 m of transmission range, assuming the Friis Free Space Path Loss (FSPL) propagation model with an exponent α equal to 3.5. The signal-to-noise threshold is set to 0.1259 mW, receiver sensitivity to -119.5dBm, and thermal noise to -110dBm.

For the data dissemination approaches, beacons are of 24 bytes and sent at 1 Hz and data messages are 2312 bytes (the maximum allowed by 802.11p). Every vehicle begins the simulation with 10 files. Each file's geographical coordinates are set to one of the two diagonal extremes of the simulation map: either (0,0) or (x_{max},y_{max}). The decision of choosing either coordinates is made at the beginning of the simulation based on the vehicles's direction. More specifically, if a vehicle is traveling towards a region near (0,0), the file's coordinate will be assigned to (x_{max},y_{max}), or (0,0) otherwise. In this manner, we simulate vehicles which have already passed by the file's geographical region before and now carry the file to other regions, for example, when they acquire a file in one city and then travel to another. In addition, files' generation times are simply defined as the beginning of the simulation, i.e., time zero. To create disparity between the utility that each vehicle may gain in a data exchange, we assign different priorities following a similar condition to the one applied for the file's geographical coordinate: 1.0 if it is traveling towards (0,0); and 0.8 or 0.6 (with 50% each) otherwise. Finally, whenever two vehicles exchange data the optimization problems defined for each data selection approach (Section II-B) are solved externally by the General Algebraic Modeling System (GAMS) [7].

Since the approaches being compared depend on an accurate connectivity duration time, we take the following two measures to improve this estimation. First, in place of sending multiple data messages for each file, we simulate large files by decreasing the rate at which data messages are sent to the MAC layer. Each data message is sent every 1.5 second, which simulates files of 1.125 Mbyte. Second, we assume that every vehicle is aware of its complete route followed in the simulation. This could be thought of as if every driver had set the route in his navigation system. Although these measures are certainly not realistic, we are able to focus more closely on the data selection approaches and more accurate conclusions can be drawn about which approach an actual protocol should be based on.

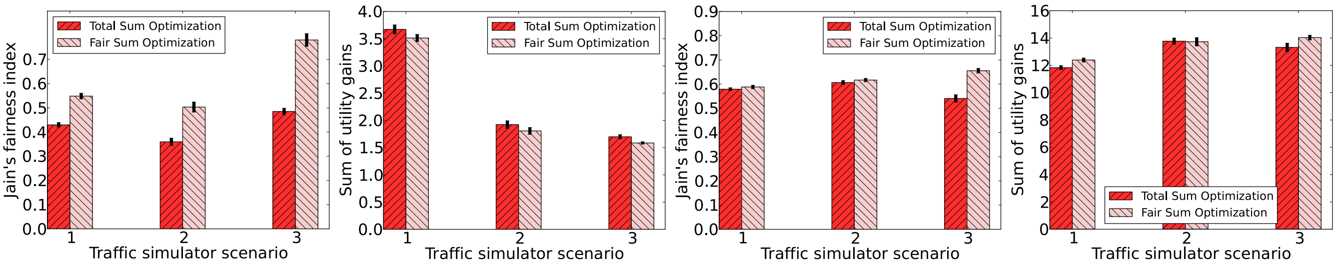
Our evaluation considers the following metrics:

- ***Jain's fairness index***: defined as $\left(\sum_i^n x_i\right)^2 / \left[n \sum_i^n x_i^2\right]$ (see [8]), where n is the total number of vehicles and x_i is the utility sum gained by vehicle i. It indicates how well utility gains are distributed among vehicles. $1/n$ and 1 are the worst and best cases, respectively.
- ***Sum of utility gains***: the average total utility sum gained by an arbitrary vehicle. It measures the overall performance of the dissemination approach.
- ***Percentage of nodes which gained utility***: also indicates how well utility gains are distributed but specifically in terms of which vehicles gained what level of utility.
- ***Number of files downloaded***: average number of files downloaded by an arbitrary vehicle.
- ***Number of hops***: average number of hops that a file travels.
- ***Delivery ratio***: the percentage of files received by vehicles which eventually approach a 1 km radius surrounding the file's geographical coordinates. It measures the approach's efficiency in terms of how often files are distributed to vehicles actually interested in the files.
- ***Delay***: the average amount of time taken from the file's generation until it is received by vehicles that will be traveling to the area to which the file relates.

A. *Two-Vehicle Scenario*

We first consider a simple scenario in which two vehicles drive in opposite directions at an average speed of 120 km/h. A space of 10 meters separates each road direction and in the simulation vehicles move at steps of 1 second. This simulation consists of 50 runs. Our goal is to assess the performance of each approach in a basic scenario with increasing available connectivity time (k_{max}).

Fig. 2(a) presents the results of applying the Jain's fairness index. FSO seeks symmetry in the distribution. Therefore, the index is above 0.9 for even numbers of k_{max} and above 0.75 for odd numbers, since with even numbers the probability of



(a) Jain's index (lim. data exchanges) (b) Utility sum (lim. data exchanges) (c) Jain's index (all data exchanges) (d) Utility sum (all data exchanges)

Fig. 3. Results with 95% confidence intervals considering data exchanges with limited connectivity time in (a) and (b); all data exchanges in (c) and (d)

achieving symmetry is higher. On the other hand, when k_{max} is low TSO tends to increase the utility of only one vehicle (index of 0.5). It is worth noticing that with the defined start-up configuration for each file's priority, generation time, and coordinate, we provoke a utility disparity between vehicles. Therefore, with TSO one vehicle will always be able to obtain a higher utility gain when compared to the other in this scenario. However, the two approaches present similar values when k_{max} is higher. This is expected since with increasing time, more files will be sent. Thus, fewer differences in the results are present.

The high level of fairness achieved with FSO comes at the price of a lower performance in terms of the total sum of utility gains, as shown in Fig. 2(b). However, this loss in terms of the maximum difference is just 20% compared to the gain in Jain's index at 40%.

Overall, FSO presents more advantageous results when compared with TSO, since it provides a fairer distribution of utility among the vehicles while not compromising the overall performance achieved by all vehicles taken together.

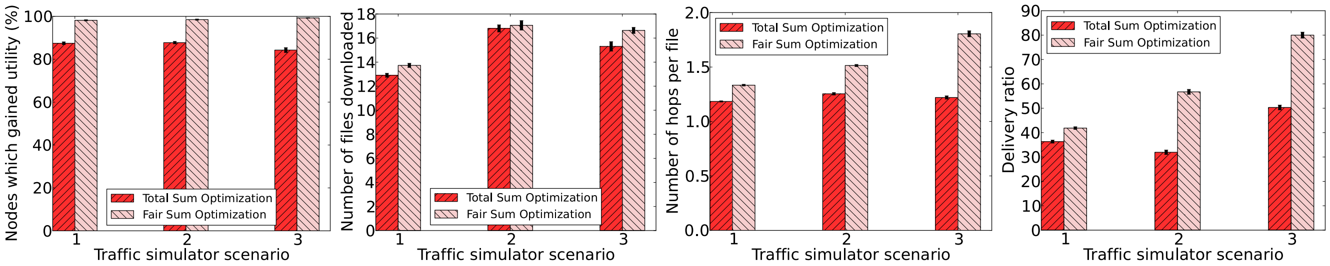
B. Traffic Simulator Scenarios

In this section, we consider more realistic scenarios created with SUMO [9]. These scenarios include vehicle overtaking, lane changing, and rely on well-known car-following mobility models such as Krauß and Intelligent Driver Model (IDM). Three scenarios with an area of 6 km^2 are considered: an OpenStreetMaps [10] map fragment (2 x 3 km^2) from the urban TAPAS Cologne traffic model [11] combining both sparse and dense networks (scenario 1), a well-connected network (scenario 2), and a sparse network (scenario 3). Scenario 1 describes the traffic within the city of Cologne (Germany) with traffic demands generated by TAPAS – a system that computes mobility demands based on the population's traveling habits and the city's infrastructure. In this scenario, the number of vehicles simultaneously moving increases linearly with time from 10 to 470, with a total of 709 generated. Vehicles' speeds vary from 0 to 100 km/h. Scenarios 2 and 3 contain a 10-kilometer straight highway with two lanes in each road direction. Lanes are 4 meters wide with a 10-meter space between each direction. A moderate traffic flow generated for scenario 2 which leads to a density of 7.5 vehicles/km/lane and for scenario 3 a low traffic flow which leads to a density of 2.5 vehicles/km/lane. In both scenarios, vehicles' speeds vary from 80 km/h to 120 km/h. Simulations consist of 25 runs of 300 seconds and with vehicles moving at intervals of 1 second.

We first study the connectivity profile of each scenario in Figures 2(c) and 2(d). Fig. 2(c) shows the cumulative distribution function (CDF) for the connectivity time of any interaction between a pair of vehicles with both approaches. We consider a restricted version of the figure limiting the x-axis to just 30 seconds to focus on a limited period (changes little after 30 s). In the sparse topology present in scenario 3, 80% of interactions are performed under 10 seconds. In the remaining scenarios, 60% of interactions are within 10 seconds in scenario 2 while 50% are within 10 seconds in scenario 1. Fig. 2(d) shows the percentage of times a data exchange is performed with limited connectivity time, i.e., not all files that are available can be exchanged. While these percentages depend significantly on the amount of data initially carried by each vehicle, in our configuration with only 10 files of 1.125 Mbyte (11.25 Mbyte) values are equal or greater than 70% for all scenarios. This verifies that vehicles often have to deal with limited connectivity time.

Fig. 3 shows the results of applying the Jain's fairness index and the sum of utility gains. We first narrow down the results to only cases with time-limited data exchanges in Figures 3(a) and 3(b). As with the results in the two-vehicle scenario, FSO achieves a Jain's index which is 20% to 65% higher than TSO. The difference is most noticeable in scenario 3 because of the presence of a sparse network in which vehicles can mostly communicate with other vehicles traveling in the opposite direction. This reflects the results shown in Fig. 2(a) where with a short connectivity time results between the two approaches diverge in a great extent. In addition, despite the greater sum of utility gains achieved by TSO, the difference compared with FSO is less than 7%. When considering all data exchanges in each scenario (Figures 3(c) and 3(d)), the Jain's index is still higher with FSO; however, it has a lower difference compared to TSO. Especially in data exchanges with abundant time, vehicles are likely to gain more in terms of utility, which directly influences the average value of the Jain's index.

Interestingly, with all data exchanges considered, FSO outperforms TSO in terms of the sum of utility gains. To understand the reasons for this result, we evaluate four other metrics, as shown in Fig. 4. The first metric is the percentage of vehicles which gained utility during the simulation. As shown in Fig. 4(a), while FSO seeks symmetry in the utility



(a) Vehicles which gained any utility (b) Number of files downloaded (c) Number of hops per file (d) Delivery ratio

Fig. 4. Results for various metrics for the traffic simulator scenarios with 95% confidence intervals

distribution and almost guarantees that every vehicle receives some utility, with TSO more than 15% of the vehicles finish the simulation without any file having been downloaded. This is also reflected in the total number of files downloaded (Fig. 4(b)) with more files being downloaded when FSO is utilized. FSO also has advantages in terms of the average number of hops a file travels and especially in terms of the delivery ratio (Figures 4(c) and 4(d)). This is a direct consequence of the previous results: with more files being downloaded by more vehicles, files are more likely to spread throughout the network and reach interested vehicles. Finally, given that files are better distributed among vehicles a shorter delays is expected with FSO. However, due to the much lower delivery ratio achieved by TSO, the results are not comparable. Therefore, the results for both approaches are almost identical with small variations. For scenarios 2 and 3, in which vehicles move at high speeds the average delay is within the range of 110 to 120 seconds. For scenario 1, the delay is higher (170 to 180 seconds) since vehicles are moving through an urban area and thus at lower speeds.

Overall, improving fairness while disseminating data can lead to a better overall Jain's fairness index, increased sum of utility gains, number of hops and number of files downloaded. Most importantly, as a direct consequence of these results, a fair approach has the potential to significantly improve the delivery ratio.

IV. Conclusion and Recommendation

This paper has presented a study of the trade-off between efficiency and fairness when disseminating data in vehicular environments. By means of simulation, we have evaluated both approaches using a variety of performance metrics and scenarios. Overall, the use of a data selection strategy which seeks fairness was shown to increase the delivery ratio, Jain's fairness index, the sum of utility gains, the number of hops and the number of files downloaded.

We concentrated on comparing *optimal* results of both efficient and fair strategies. Therefore, a few simplifications were proposed which deserve further attention in future work. First, the basic protocol used in our comparison prioritizes communication robustness with the use of RTS/CTS of 802.11p. Hence, it allows only the communication between every pair of vehicles. However, in various situations multiple vehicles will be simultaneously available for exchanging data. Although more unreliable, broadcast communication may effectively improve the protocol's performance by making better use of the available bandwidth, and consequently of time. Second, the data selection model defined to achieve fairness relies on the Nash Bargaining solution for binary variables. This solution falls in the class of Mixed-Integer Non-Linear Programming (MINLP) which are NP-hard problems. Finally, we considered that the connectivity time is calculated before the actual data exchange. Such calculation can be difficult due to unpredictable variations of busy radio medium and incomplete information about vehicle routes.

In conclusion, any protocol should rely on a fair data distribution among vehicles, considering both unicast and broadcast communication. Furthermore, it should adapt data selection decisions during the data exchange as new context knowledge about the vicinity is acquired.

References

[1] T. Kosch, C. Adler, S. Eichler, C. Schroth, and M. Strassberger, "The scalability problem of vehicular ad hoc networks and how to solve it," *IEEE Wireless Communications*, vol. 13, no. 5, pp. 22–28, Oct. 2006.

[2] E. Shrestha, B. and Niyato, D. and Zhu Han and Hossain, "Wireless Access in Vehicular Environments Using Bit Torrent and Bargaining," in *IEEE Global Telecommunications Conference, IEEE GLOBECOM*. Boise State University, 2008, pp. 1–5.

[3] R. S. Schwartz, H. W. van Dijk, and H. Scholten, "Towards Opportunistic Sensed Data Dissemination in Vehicular Environments," in *IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, Seattle, USA, 2011.

[4] J. Nash Jr, "The bargaining problem," *Econometrica: Journal of the Econometric Society*, vol. 18, no. 2, pp. 155–162.

[5] "IEEE Draft Amendment to Standard for Information Technology - Telecommunications and Information exchange between systems - Local and Metropolitan networks - specific requirements - Part 11: Wireless LAN Medium Access Control," *IEEE 802.11p Draft 2.0*.

[6] A. Varga, "The OMNeT++ discrete event simulation system," in *Proceedings of the European Simulation Multiconference (ESM2001)*, 2001, pp. 319–324.

[7] M. Bussieck and A. Meeraus, "General algebraic modeling system (GAMS)," *APPLIED OPTIMIZATION*, vol. 88, pp. 137–158, 2004.

[8] W. Jain, R. and Chiu, D.M. and Hawe, *A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Systems*. Digital Equipment Corp., 1984.

[9] D. Krajzewicz, G. Hertkorn, C. Rössel, and P. Wagner, "SUMO (Simulation of Urban MObility)-an open-source traffic simulation," in *Proceedings of the 4th Middle East Symposium on Simulation and Modelling (MESM20002)*, 2002, pp. 183–187.

[10] M. M. Haklay and P. Weber, "OpenStreetMap: User-Generated Street Maps," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, Oct. 2008.

[11] C. Varschen and P. Wagner, "Mikroskopische Modellierung der Personenverkehrsnachfrage auf Basis von Zeitverwendungstagebüchern," *Integrierte Mikro-Simulation von Raum-und Verkehrsentwicklung. Theorie, Konzepte, Modelle, Praxis*, vol. 81, pp. 63–69, 2006.