Evaluation of Server Performance in VANETs in Single-Hop Scenarios

Christian Bruns
Volkswagen AG
Group Research Electronics
Wolfsburg, Germany
christian.bruns1@volkswagen.de

Klaus Jobmann
Institute of Communications Engineering
University of Hanover
Germany
jo@ant.uni-hannover.de

***Abstract* – In this paper a fundamental C2X-communication scenario is investigated via simulation in order to evaluate the influence of different serving strategies on the performance of an immobile server. For investigations the wireless LAN amendment IEEE 802.11p was implemented as communication technology in ns-2. Additionally a link adaptation algorithm has been added that exploits the multi-rate capability of IEEE 802.11. To enhance its serving performance, the server is enabled to recognize and exploit mobility parameters of requesting mobile clients. Improved serving heuristics are analyzed and compared to existing standard approaches.**

***Index terms* – Server Performance, Inter-Vehicle-Communication, IEEE 802.11p, Link Adaptation, Resource-Allocation, Neighbour-Awareness, VANETS**

I. Introduction

During the last years wireless car-to-car- and car-to-infrastructure-communication (C2X-communication) became a major research topic in the field of Intelligent Transportation Systems (ITS). Communication in vehicular ad-hoc networks (VANETS) serves as the basis for future active safety functions that are expected to significantly reduce the number of traffic accidents. Furthermore various other services addressing domains like traffic efficiency or infotainment are conceivable.

This paper focuses on a fundamental single-hop communication scenario where a stationary C2X-node asynchronously serves the requests of mobile clients which pass by on a nearby roadway. It is assumed that these nodes locally exchange information about their positions, velocity, and driving directions. A second basic characteristic of the setting is the usage of the multi-rate wireless LAN technology IEEE 802.11. Particularly the IEEE 802.11p amendment [IEE05] has been adopted. Operation of the server is investigated for varying load.

By intelligently exploiting information about the movement of mobile stations as well as the multi-rate capability it is shown how the performance of the server can be improved. The investigations are based on a concept, proposed in a previous paper [BLR+05], to enhance usage of the wireless medium, which is still considered to be the bottleneck of the communication system.

This paper is organized as follows: After the introduction, in the second paragraph a short survey of current research activities and approaches in the field of C2X-communications is given. Afterwards the investigated application scenario, the employed serving strategies, and the simulation environment are explained. Subsequently simulation results on the performance of the server are presented and discussed. Eventually the last chapter comprises an outlook on ongoing activities and further extensions of the performed investigations.

II. Technologies, Protocols, and Applications For C2X-Networks

In Europe and the US the wireles LAN ammendment IEEE 802.11p has been selected as basis for short to medium range communication services in vehicular environments. IEEE 802.11p is essentially a modified version of the IEEE 802.11a specification which operates in a different frequency band and supports lower data rates due to changes in the channel bandwidth. Like prior versions of the IEEE 802.11 standard, the new version also supports variable data rates depending on the channel conditions. [Zhu03] denotes several limitations of IEEE 802.11 in vehicular environments including the blocking problem in multi-hop scenarios and implications of mobility on the communication duration.

Recent European projects like FleetNet [Fra04] or CarTalk2000 concentrated on the evaluation and development of algorithms for position-based routing and forwarding [Mau01] [Mai04]. These protocols are based on position-awareness of each node and were intended to enable data

communication over multiple hops to a destination node or a target area, even in environments with node mobility and without need for a fixed infrastructure. For this, nodes regulary broadcast so-called beacon messages that include position information to inform neighbouring nodes about their presence and location.
The mobility of nodes implicates a continuously changing network topology. Consequently the number of neighbours of specific units usually varies over time just as the distance to these nodes. Another issue derives from partitioning of the network due to a partially low number of C2X-units. Furthermore the wireless medium is a scarce resource, particularly in networks with a high density of nodes and a high node mobility [Skl97].
Application that basically distribute messages unidirectionally and whose information comprises a single or few data packets only might cope well with these limitations. This is reasoned by the fact that often the packets can be cached by relay nodes and fowarded later in cases where no appropriate next node is available. Examples are the distribution of warning or traffic information messages [WER04]. However often there are also constraints due to maximum latencies for the information.
In contrast, many applications base on bidirectional communications between two or more nodes and usually there are numerous packets to be exchanged. This implicates a need for a certain interval of a stable link between these nodes. Sample applications are media download from a stationary server or information exchange with vehicles in the vicinity. Consequently much harder demands on the network environment are required in order to avoid communication interruption which leads to waste of the channel capacity.

III. Investigated Application And Scenario

To realize applications that are based on stable links in [BLR+05] a concept has been presented to assess neighbouring C2X-stations based on their mobility parameters and use this information to estimate the link stability to these units. In the following this idea is applied to a basic single-hop scenario illustrated in Figure 1.
An infrastructure server is installed close to a roadway and offers services that can be selected and employed by surrounding C2X-nodes. The spectrum of conceivable applications is immense. For instance the server could administrate and provide dedicated traffic information or information about local points of interest (POI) that are distributed to nodes in the surrounding. If the C2X-server is connected to an existing local- or metropolitan area network or the Internet, these information could also be retrieved from other remote servers. In the setting it is assumed that each vehicle requests its own personal information, i.e. no two clients retrieve the same data set.

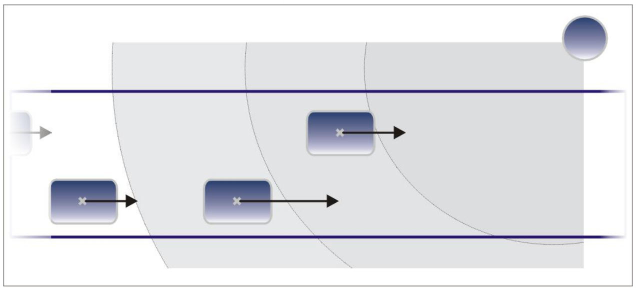

Figure 1 Immobile C2X-server serving requests of mobile clients with different data rates

The basic proceeding is described in the following. The server announces its service via service advertisments that are broadcasted regularly with frequency f_{sa}.
Furthermore all nodes periodically broadcast information about their movement by means of beacon messages that are issued with frequency f_{beacon}. The beacons include the current position of the corresponding C2X-station, as well as the mobility vector, i.e. the driving direction and velocity along with a timestamp. This timestamp indicates the point in time when the corresponding information has been determined. It is necessary for reconstruction of the real movement on the receiver’s side, since a non-deterministic delay is introduced by the lower layers of the system, for instance by the MAC protocol.
All vehicles, which move into the communication range of the server and receive an announcement, send a directed message to the server and query for downloading a certain amount of data. The server in turn parses the request of each client and prepares the data for transmission. Note that the delay, introduced by this preparation, is neglected, i.e. the data is immediately available.
In the setting it is further assumed that when a vehicle leaves the service range prior to being served completely, its request times out and might

not be finished any more. Consequently there is only a limited interval to finish specific queries.

As depicted in Figure 1 the information can be transmitted with higher data rates within certain areas around the server if the link adaptation feature of IEEE 802.11p is enabled. In the analysis is presumed, that all vehicles have the same omnidirectional antenna charateristic and that radio waves propagate like in a freespace environment. It It is regarded that this is a reasonable assumption for rural evironments.

With these determinations it is possible to describe the feasible data rate depending on the distance between sender and receiver (cp. Figure 2). Since the server receives mobility information from all C2X-nodes, he is in a position to derive a specific *serving profile* for each vehicle in its vicinity.

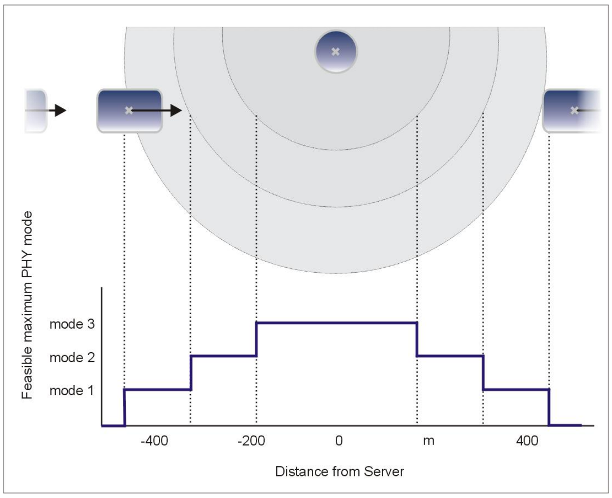

Figure 2 Principle of estimating the feasible data rate over time (profile) by the server using the position and movement information of the corresponding vehicle.

When it comes to process the requests of clients, the server needs to utilize an adequate serving strategy, especially in situations of heavy load. A possible objective for server's operation is to maximise the overall number of completed jobs. Contrariwise the the amount of uselessly transmitted data should be kept as low as possible. Useless transmissions in this context denote packets that have been transmitted to clients which eventually could not be finished in time.

IV. Terminology and Serving Algorithms

Before describing the serving policies that considered in the analysis, the following notational parameters are defined:

- A request that has been issued by a certain vehicle *i* results in a job J_i that shall be satisfied by the server. Each vehicle issues a single request only. In the following the terms job, customer, and request are used interchangeably.
- r_i denotes the *release date* (arrival time) of a job J_i, i.e. the time when the request arrives at the server.
- d_i denotes the *due date* or *deadline* of job J_i. It represents the point in time when the vehicle leaves the communication range of the server. In contrast to the release date this time is estimated by the server using information about the current position, velocity and moving direction of the corresponding vehicle. The estimation is continuously updated upon reception of new beacon messages.
- w_i denotes the *weight* also referred to as *value* which defines the importance of a job. For investigations it is defined that the weight equals the size of a request, i.e. the amount of data that has to be transmitted to the client. In the simulations the initial size of all jobs is assumed to be equal for all clients, hence their weights are equal as well.
- c_i denotes the *estimated remaining processing time* needed at the current time to successfully complete job J_i without preemption. Note that although a job is not served in between its corresponding c_i may vary over time. This variation stems from the corresponding data rate profile, i.e. the current position in this profile.
- v_i denotes the *value density* of a job J_i defined by the quotient w_i/c_i. Note that this definition for the value density takes into account jobs that have already been processed partly. Since the weight of job J_i stays constant but the remaining processing time might become very short when a job is nearly finished, the corresponding value density finally tends to infinity.

For the performance study three different serving strategies are considered. Due to the packet oriented transmission process a preemptive job processing is possible and allowed.

- First Come First Serve (FCFS): The customers are served in order of their arrival, i.e. by their release dates r_i. This standard policy is often adopted in real systems as it is easy to implement and an intrinsically fair algorithm.

- Enhanced First Come First Serve (EFCFS): The basic processing order follows FCFS but the server estimates the feasibility of successfully serving a client prior to transmitting the data. This is done by using the current data rate profile and integrating over time to get the theoretical maximum amount of data that can be transmitted to the corresponding node. The decision is based on a comparison between the remaining and the theoretically feasible value. Hence, huge jobs that are close to their deadlines prior to being served are ommitted.
- Enhanced Highest Density First (EHDF): The order of tasks is assigned according to the current value density v_i of the job J_i. However, ***enhanced*** denotes that, like in case of EFCFS, the server estimates whether it is feasible to complete the customer prior to its due date d_i. Otherwise the job is ignored.

V. Simulation Environment

The simulation environment employed for investigations consists of three logical components that emulate vehicle mobility, the communication network and protocols, and finally the application logic.

For vehicle mobility the microscopic vehicle traffic simulator Vissim is widely accepted for scientific investigations [LCS+05] [SDK+05] and has been adopted for modelling realistic traffic scenarios. With Vissim 3.70 a straight roadway scenario has been set up. There are two lanes and the vehicles have an average velocity of 50 km/h. For the analysis only one driving direction is assumed and the number of vehicle arrivals per hour was set to 700. This value allows free traffic flow, i.e. there is only little interaction between the mobile nodes. The information about the vehicle movement is generated in a time discrete manner and is in turn fed into the network simulator.

The network simulator 2 (ns-2), one of the mostly reputated network simulators for IP-based wired and wireless networks, was used for modelling the communication protocols as well as radio wave propagation. Its software architecture allows to modify existing modules or to attach own extensions.

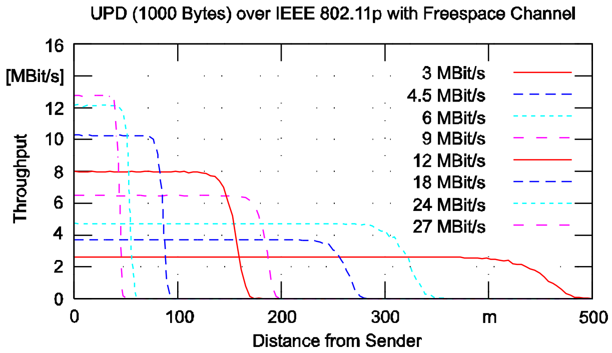

Figure 3 Throughput for all PHY modes of IEEE 802.11p for UPD (1000 Bytes) and freespace channel

Features that were added to version 2.28 of the ns-2 implementation include:

- The existing IEEE 802.11 PHY and MAC code was extended to IEEE 802.11p complicance with multi-rate support according to the draft specification [IEE05]. Figure 3 shows the saturation throughput for single station UDP traffic. See [Het01] for comparable studies on IEEE 802.11a.
- A modified version of the Auto Rate Fallback (ARF) algorithm [KM97] has been implemented for investigations which constitutes an IEEE 802.11 compliant link adaptation scheme. With this algorithm a station basically counts the number of consecutive successful and erroneous transmissions to each other node and switches to the next higher / lower PHY mode when the limit $C_{\text{limit,success}}$ / $C_{limit,fail}$ is exceed. Since a low Signal to Interference Ratio (SIR) as well as collisions may reason an erroneous transmission attempt and ARF cannot distinguish both cases, the RTS/CTS mechanism of the MAC protocol is enabled for simulations which reduces the collision probability significantly. In order to improve the average number of initial transmission attempts to a new node, a startup mode is introduced. This mode begins with the highest PHY rate, if this transmission fails the next lower mode is selected for retransmission. For the simulations it has been determined to limit the number of PHY modes m to the mandatory rates 3, 6, and 12 Mbit/s (m=3). The motivation for this decision is manifold. Firstly the number of attempts to determine the optimal PHY mode and consequently the overhead due to switching

between the data rates is reduced. Secondly lost connections will be determined ealier, at most after $C_{limit,fail} \cdot m$ unsuccessful attempts. The last reason is that the complexity of the implementation is decreased and the implementation is equal for each unit.

- Packet Error computation based on the actual signal strength has been added and the *no routing protocol* (NORT) module has been adapted.

The third component of the simulator environment, the application logic, was also implemented as new module in ns-2. This implementation comprises the client application as well as the C2X-server operation, including concept implementation and scheduling routines.

In order to draw substantiated conclusions from the investigations, interfering influences from other (intermediate) layers or applications have to be eleminated or minimized respectively. This leads the following additional configuration:

- The server is the only station which intends to occupy the wireless resources as needed. There is no traffic from concurrent applications. However, as described above the beacon messages as well as the service advertisements are exchanged. The frequencies f_{sa} and f_{beacon} are constant and have been set to 1/s.
- A plain reliable transport protocol is adopted. Since TCP inheres several drawbacks which for instance lead to misinterpretations of packet losses in wireless networks, the unreliable UDP protocol has been selected as a basis. The implementation of UDP was extended by a transport layer acknowledgement that is sent for each received packet. This modified protocol is named *Acknowledged UDP* (ACK-UDP) and constitutes an elementary means of enabling the C2X-server to verify the successful transmission of the requested application data.
- No routing protocol is utilized for simulations. Thus besides the network layer address, which is included in each packet, no additional overhead is introduced by this layer.
- On application layer a fix packet length of 1000 Bytes is used by the C2X-server for serving the clients.

VI. Simulation Results

In this section simulation results on the performances of the selected serving policies for the described scenario are presented and discussed. Different simulation runs with increasing jobsize have been performed for the described scenario.

Note that in Figure 6 and Figure 7 the values of the ordinate represent the maximum throughput on application layer for the considered setting and PHY modes. For the analytical determination of these values the influence of the beacons and the service anouncements were neglected. Due to the transport layer acknowledgements and the RTS/CTS handshakes, these values differ significantly from the values shown in Figure 3.

Regarding the performance study, at first results on the number jobs that have successfully been processed by the server are analyzed.

Figure 4 shows the rate of processed jobs in percent of the total amount of requests. When the network load is low, i.e. the requests are comparatively small and the network is not saturated, the server is in a position of processing all requests in time. As the load at the server increases, it is not possible to satisfy all requests any more. In addition to the results of the strategies a reference curve for optimal serving is depicted. This curve has been calculated for the given vehicle arrival rate with the assumptions that a.) each job is processed with the highest possible average data rate b.) upon finishing a job there is always a next job that can also be served in the same optimal manner.

The illustration clearly reveals that the naïve strategy FCFS inheres the poorest performance for higher loads. Its corresponding rate of processed jobs decreases sharply at a jobsize around 20 MBit and quickly tends to zero when the load is further increased. In contrast the enhanced strategies EFCFS and EHDF have much better performance whereat EHDF comes very close to an optimum resource allocation.

The interpretion of these simulation results starts with an analysis of FCFS. In overload the overall requested amount of data exceeds the capacity of the server which leads to queueing. In case of FCFS the server is forced to sequentially process all jobs which leads to a high average waiting time of the queued jobs. For medium loads the introduced waiting times pushes the server towards higher PHY modes so a higher throughput can be achieved. However, for jobsizes above 20 Mbit, according to Figure 5 the average waiting time also increases drastically for FCFS. With the specified average speed of 50 km/h and a maximum communication range around 450 meters (cp. Figure 3) the average communication duration calculates to approximately 65 seconds. Thus from

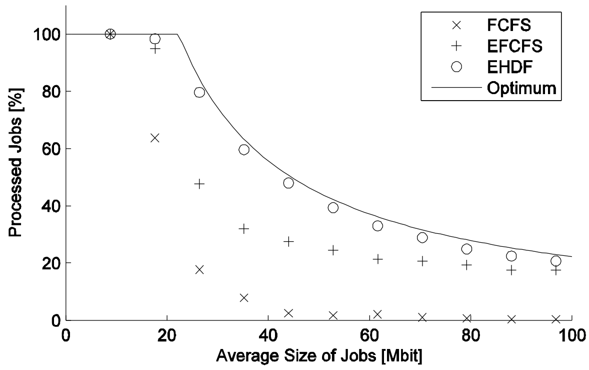

Figure 4 Number of successfully processed jobs depending on the size of jobs

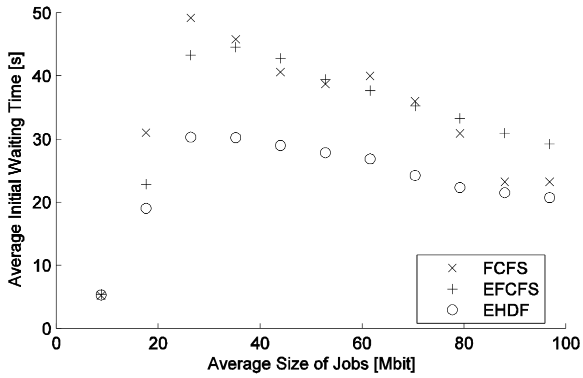

Figure 5 Average initital waiting time of jobs that are successfully processed by the server

the data rate profile it can be derived that in this stage the server cannot operate with the maximum PHY mode any more. This circumstance in turn worsens the situation as processing time increases and an even higher waiting time is inposed on the queued jobs. Finally when the processor starts processing a job, the remaining time until its due date is likely not sufficient to complete it, which leads to a high rate of dropped jobs. The situation constitutes a worst case for the network, since the server intends to occupy the wireless resources completely but wastes them due to useless transmissions. This statement can be justified in Figure 7.

The enhanced policy EFCFS overcomes this major drawback of FCFS as it intends to ignore all jobs that are expected to exceed their deadline prior to being served completely. In the considered scenario with free traffic flow the variation of the velocity of vehicles is low and consequently the estimation about the future movement of the nodes is precise.

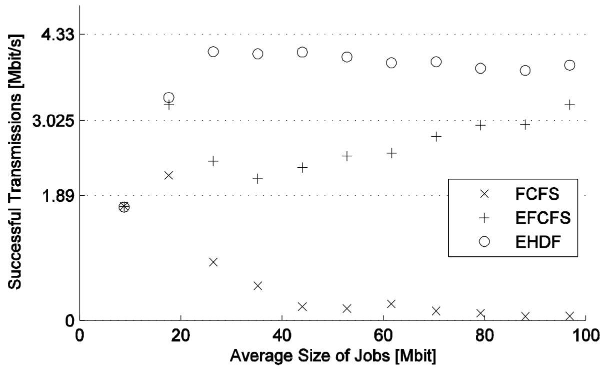

Figure 6 Average throughput on application layer for successful transmissions depending on the size of jobs.

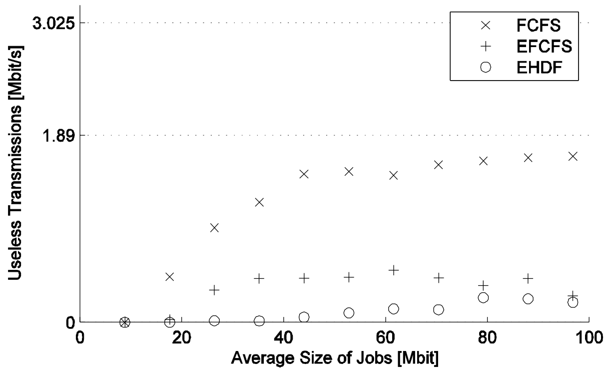

Figure 7 Average data rate on application layer that is spent on serving jobs which are not finished completely.

Furthermore the throughput estimations are good since there is nearly no concurrent communciation on the channel. According to Figure 7 the enhanced serving strategy is appropriate to reduce the wasted bandwidth significantly. However, there is still potential for further improvements, since due to its relation to FCFS this enhanced strategy also tends to long queues and processes jobs rather late close to their deadlines. Thus even small derivations of the movement lead to a wrong estimation and may cause an early dropping during processing.

For FCFS as well as EFCFS Figure 5 clearly shows how the increasing jobsize implies a lower average waiting time for processed jobs. Supposing that, there is a fix maximum average communication duration, it is necessary to start a job earlier the higher its size is in order to finish it prior to its due date.

Finally the strategy EHDF shows superior performance in this scenario as it approaches the overall capacity of the system even for high loads.

The graph in Figure 6 reveals that for low and medium jobsizes the C2X-server does not utilize the highest data rate for serving which is caused by its *greedy* nature, i.e. if there is a node that can be served, it is served instantly. However, when the load demands for queueing the node turns to the higher data rates and holds this mode as possible.
In the given setting there are approximately 12 vehicles in the communication range of the server per time, which can be assessed by Little's result [Kle76] on a system with arrival rate 700/3600s, and system time 65s. Consequently the server is likely able of serving one of the available clients with the optimal data rate. This can also be derived from Figure 5 which depicts the much lower waiting time for the processed jobs for EHDF compared to the other strategies. As the served clients usually still have a remaining time until their due date, a variation of the movement of the corresponding vehicle has only minor influence. Thus the probability that a job which is still being processed times out is very low. It is expected that there is an increasing influence for very high loads, i.e. when the jobsize approaches the capacity of the server for a single customer.

VII. Conclusion And Outlook

In this paper a comparative study among three serving policies for a stationary C2X-server in a multi-rate vehicular ad-hoc network setting has been presented. The mobility of nodes leads to a variation of the feasible data rate for serving and implies a limited communication duration. Two of the investigated policies use the deadlines or rather the deadlines and the values of jobs to improve the performance and to achieve a restrained degradation in overload situations.
One important result that can be derived from the simulation study is that scheduling by value density appears to be a very effective strategy in overload conditions. The analysed strategy EHDF shows superior performance in the simulations and reduces the average amount of wasted bandwidth significantly compared to the popular FCFS strategy.
In order to derive fundamental conclusions, the setting has been investigated under rather optimal conditions without interfering communications and in a freespace environment. Current research addresses the investigation of more sophisticated communication settings like intersection scenarios and the usage of more realistic channel models.

VIII. References

[BLR+05] C. Bruns, A. Lübke, B. Rech, K. Jobmann: *Concept for Improved Allocation of Communication Resources in Vehicular Ad hoc Networks.* 5th European Congress on Intelligent Transport Systems, Hanover, Germany, June 2005

[Fra04] W. Franz, C. Wagner, C. Maihöfer, H. Hartenstein: *FleetNet: Platform for Inter-Vehicle Communications.* WIT 2004, Hamburg, Germany, March 2004

[Het01] A. Hettich: *Leistungsbewertung der Standards HIPERLAN/2 und IEEE 802.11 für drahtlose lokale Netze.* ABMT 23, 1. Edition, Aachen, Germany, 2001

[IEE05] IEEE Standard 802.11p Draft Amendment: *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Wireless Access in Vehicular Environments (WAVE)*, 2005

[Kle76] L. Kleinrock: *Queueing Systems, Volume 2: Computer Applications.* Wiley, New York, 1976.

[KM97] A. Kamerman, L. Monteban: *WaveLAN-II: A High Performance Wireless LAN for the Unlicensed Band.* Bell Labs Technical Journal, 2(3):118–133, Summer 1997.

[LCS+05] C. Lochert, M. Caliskan, B. Scheuermann, A. Barthels, A. Cervantes, M. Mauve: *Multiple Simulator Interlinking Environment for Inter Vehicle Communication.* VANET 2005, Cologne, Germany, September 2005

[Mai04] C. Maihöfer: *A Survey on Geocast Routing Protocols*. IEEE Communications Surveys and Tutorials, Vol. 6, No. 2, 2004

[Mau01] M. Mauve, J. Widmer, H. Hartenstein: *A survey on position-based routing in mobile ad hoc networks.* IEEE Networks, Nov./Dec. 2001

[SDK+05] C. Schroth, F. Doetzer, T. Kosch, B. Ostermaier, M. Strassberger: *Simulating the traffic effects of vehicle-to-vehicle messaging systems.* In Proceedings of the 5th IEEE

International Conference on ITS Telecommunications, Brest, France, June 2005.

[Skl97] B. Sklar: *Rayleigh Fading Channels in Mobile Digital Communication Systems Part I: Characterization.* IEEE Communications Magazine, July 1997

[WER04] L. Wischhof, A. Ebner, H. Rohling: *Self-Organizing Traffic Information System based on Car-to-Car Communication: Prototype Implementation.* WIT 2004, Hamburg, Germany, March 2004.

[Zhu03] J. Zhu, S. Roy: *MAC for dedicated short range communications in intelligent transport system.* IEEE Communications Magazine, December 2003