Universität Karlsruhe (TH)
Fakultät für Informatik
76128 Karlsruhe

Ad-Hoc Netzwerke für die Kommunikation zwischen Fahrzeugen

Seminar – WS 2004/05

Herausgeber:
Prof. Dr. Hannes Hartenstein
Marc Torrent-Moreno

Universität Karlsruhe (TH)
Institut für Telematik
Lehr- u. Forschungsbereich Dezentrale Systeme und Netzdienste

Interner Bericht 2005-2
ISSN 1432–7864

Abstract

This Technical Report includes the written works performed by the students based on the papers grouped as seminar of "Ad-Hoc Networks for Inter-Vehicle Communication" (VANETs). Although it was the first time this seminar was offered, it attracted a large number of motivated and diligent students, proving the high interest that VANETs have at the moment.

Contents

Preface

The seminar *'Ad-Hoc Netzwerke für die Kommunikation zwischen Fahrzeugen'* (Ad-Hoc Networks for Inter-Vehicle Communication) has explored the innovative field of vehicular networks, also known as VANETs. Recently, VANETs have gained the attention of all, academia, industry and people in general, thanks to some governments' participation in projects worldwide. This projects have as a main goal to improve safety on the roads and traffic flow. In Germany, the 'Network on Wheels' (NoW) project started last year, June 2004, following the successful Fleetnet project. NoW is founded by the German Ministry of Education and Research (BMB+F) and it is composed by several companies: DaimlerChrysler, BMW, Volkswagen, Siemens, NEC, Fokus, as well as the universities of Mannheim, München and Karlsruhe. The scope of this project includes a) solving technical challenges related to robustness, performance, scalability, security and safety for inter-vehicle communications (where our seminar is focused), b) design a base architecture and platform for VANET technology and c) explore applications for market introduction. This type of networks present a perfect subject of study for a seminar considering its research challenges as well as its real life applicability. Thus the seminar has attracted the interest of a large number of highly motivated students. The 8 selected papers have given the participants a very good state of the art of the on-going research in the following fields: Simulation and Modeling, Safety Applications Performance, Positioning, Security and Data Dissemination.

The intended goal of the seminar has been to develop a technical basis in an innovative field, to understand VANETs' challenges and to recognize the variety of fields of activity related to VANETs. Furthermore, special attention was addressed with every student to practice and improve their problem statement, solution study, discussion and presentation skills.

We thank all participant students for their written contributions and strong commitment.

Review of "Broadcast Reception Rates an Effects of Priority Access in 802.11-Based Vehicular Ad-Hoc Networks"

Daniel Schmidt

Abstract

The following report is based on the paper "Broadcast reception Rates and Effects of Priority Access in 802.11-Based Vehicular Ad-Hoc Networks" by Marc Torrent-Moreno, Daniel Jiang and Hannes Hartenstein [TMJH04]. It addresses the problem of probability of reception of broadcast messages and how to improve it by adding priority access mechanisms to the used protocol. Therefore, two different scenarios, a static and a dynamic one, were set-up to simulate the effects of the various protocol parameters and of the improvements that have been made by reducing channel access time and improving reception rates. To quantify the results, the simulations were done using two different radio propagation models: the deterministic two-ray-ground model an the Nakagami distribution [Naka60]. The paper shows that a well designed protocol is needed to achieve the improvements which are necessary to improve the reliability of VANETs.

1 Introduction

The most important aspect of vehicular ad-hoc networks (VANETs) is without a doubt the transmission of safety related data between the cars. Since the data has to be delivered to all intended receivers, a broadcast message system is used to send these valuable data to all cars within a certain communication range within fractions of seconds. The probably life-saving importance requires the chances of receiving the broadcast message by other cars to be as high as possible. In this paper, a priority access mechanism has been used as suggested by the 802.11e draft [802], utilizing a 5.9 Ghz spectrum with 10 MHz channels and offering a total data rate of 6 Mbps. As a real world experiment would be too complex and costly at this stage of development, the analysis of this priority mechanism is done using computer simulation. First, the influence of some parameters is evaluated in a static scenario. The insights gained from this simulation are then used in a dynamic model to observe their impact on a vehicular network.

2 Radio propagation models

If we want to study the effects of a priority mechanism, we need to know the quality of reception at different distances from the sender. To describe this propagation of radio waves, various models have been developed. The simplest of them is a free space model without any reflection, diffraction or scattering, which is highly unrealistic. To increase the degree of realism, a second "ray" that reflects off the ground on its way to the receiver is added in the two-ray-ground model. Later, a far more sophisticated model developed by M. Nakagami is used, which adds a fading parameter.

2.1 Two-ray-ground model

The two-ray-ground (TRG) model is the simpler radio propagation model of the both used. To study the propagation of radio waves the model assumes two rays between the sender and the receiver. One is the direct line of sight, the second is reflected by the ground as shown in Figure 1.

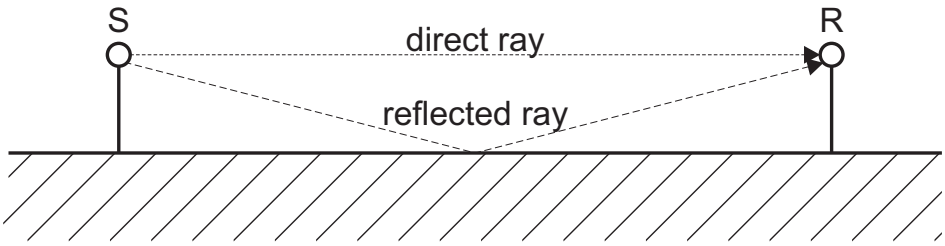


Figure 1: Simple explanation of the two-ray-ground model

2.2 Nakagami distribution

In contrast to the very simple two-ray-ground model, the radio propagation model developed by M. Nakagami is far more sophisticated and realistic, as it adds a fading parameter m to represent the channel fluctuations.

$f(x;\Omega,m) = \frac{2m^m x^{2m-1}}{\Gamma(m)\Omega^m} \exp\left[-\frac{mx^2}{\Omega}\right]$

$x \geq 0, \Omega > 0, m \geq 1/2$

x: signal amplitude
m: fading parameter depending on d
Ω: average power depending on d

The most likely values for m and Ω for the work on the paper were determined by measuring the packets received by moving cars on a highway following a vehicle, which sent 200 Byte packets every 100 ms. The recorded data was then used to estimate the values with the help of a maximum-likelihood estimation. Ω was estimated to a certain value of average power, decreasing with d^{-2}. The fading parameter m has been fixed to the values 3, 1.5 and 1 for distances lower than 50, between 50 and 150, and higher than 150 meters, respectively.

3 ns-2

ns-2 [ns2] is widely used to simulate wired and wireless networks and their respective ISO/OSI layers. The program has been developed by the University of California, Berkeley and is now maintained by the Information Science Institute of the University of Southern California. To create a simulation you can either choose one of the several protocol implementations already included in the package or start programming your own. The simulation setup is then written into a TCL script, where you can i.e. define your nodes, their parameters and the traffic between them. The results are stored in so called trace files and can be analyzed

with tools like TraceGraph[1] which allow a graphical display of the information. There are even tools to animate the results gained from the simulation[2].

4 Channel access

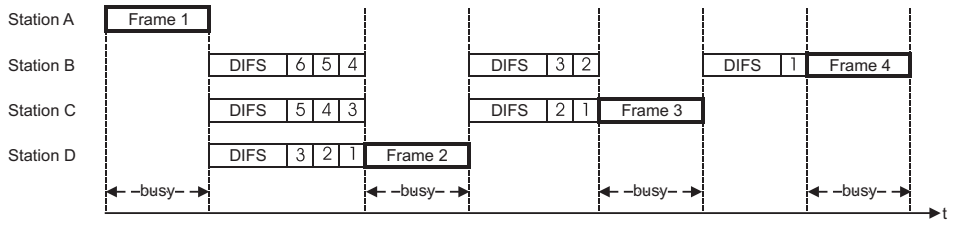


Figure 2: Channel access in 802.11

To understand how priority access works we first take a look at the distributed coordination function (DCF) channel access method used in "classic" IEEE 802.11. If a node wants to send a packet to the channel it senses the medium. If it's busy, the node waits until the medium becomes available again pauses until the DCF interframe space (DIFS) time is over and the backoff timer (BT), a randomly assigned number limited by the contention window (CW), is decremented to 0. If another node accesses the channel in the mean time, the backoff procedure is suspendend and can only be resumed after the node currently accessing the channel has finished its transmission and one more DIFS time has passed. If BT value reaches 0 - if a collision occurs with another node, a new BT value is assigned and the whole procedure starts all over.

In the example given in Figure 2, there are four stations. Station A has already accessed the channel and transmits its data. Stations B, C and D have BT values of 6, 5 and 3, respectively. When Station A has finished its transmission, the three other nodes pause for the given DIFS time and then start decrementing their backoff timers. As station D is the first to reach BT=0 it starts transmitting its data immediately when it reaches BT=0. During that time, the other backoff processes are paused, until the medium becomes available again. Then, after the DIFS time and two more SlotTimes, Station C reaches BT=0 and sends its packets, causing Station B to pause its backoff procedure.

5 Implementation of priority access

To create a priority access method the Enhanced Distributed Channel Access (EDCA) has been implemented as a ns-2 module as proposed in the IEEE 802.11e draft with respect to broadcast messages. The implementation consists of four different access categories (AC) which define the level of priority. The higher the access category the higher the priority of the node actually is. Instead of using DIFS and a fixed contention window (CW_{min}), as normal network nodes do, the prioritized nodes use interframe spaces (AIFSD[AC])and different contention windows (CW_{min}[AC]) depending on their respective access category - Table 1. The interframe space is determined by the following function:

$$\text{AIFSD[AC]} = \text{SIFS} + \text{AIFS[AC]} * \text{SlotTime}$$

[1]TraceGraph, http://www.geocities.com/tracegraph/
[2]NAM: Network Animator, http://www.isi.edu/nsnam/nam/

AC	CW_{min}	AIFS
0	aCW_{min}	2
1	aCW_{min}	1
2	(aCW_{min}+1)/2-1	1
3	(aCW_{min}+1)/4-1	1

Table 1: 802.11e priority parameters

AIFS	CW_{min}	AC
2	15	0
2	7	-
1	7	-
1	3	3

Table 2: Scenario 1 priority parameters

The values of Short Interframe Space (SIFS) and SlotTime are given by the simulation setup - 32 μs and 13 μs, respectively.

Prioritized nodes can now access the channel earlier than non-prioritized nodes, therefore resulting in shorter channel access times.

6 Simulation scenarios

Both simulations were run on a 3.2Ghz Pentium 4 with 1 GB RAM. Each pass took up to nine hours of computation time and resulted in five gigabytes of data. The analysis of this data lasted for over three hours.

6.1 Basic scenario

The basic scenario uses a static topology without any moving nodes. As this setup has no direct relevance to the addressed topic of vehicular networks it is used to study the different values for AIFS and CW and their effect on channel access time and probability of reception. The chosen values can be found in table 2.

6.1.1 Setup

The topology of the simulation setup looks like this: 600 cars are placed on an eight-lane highway. Every lane is 4 meters wide and the distance between two cars in one lane is 20 meters. In order to simulate the radio propagation, the deterministic two-ray-ground model is used. Each car sends 500 bytes sized UDP packet every 100 ms with 10% jitter. One car, placed in the middle of the highway, is acting as a prioritized node, all other have access category AC=0.

6.1.2 Results

Figure 3 shows that the AIFS value has a much bigger impact on probability of reception than a tweaked contention window. A lower AIFS results in an almost doubled probability whilst a smaller CW increases it only minimally.

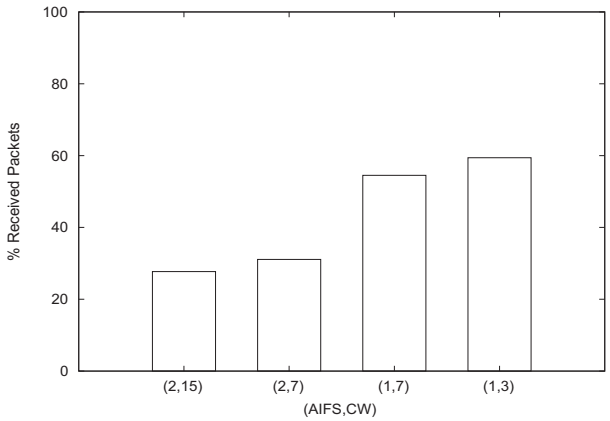


Figure 3: Probability of reception for some AIFS- and CW-values

AIFS/CW	RcvPkts	SentBT1	RcvBT1	SentBT0	RcvBT0
2/15	27.7%	69.3%	22.6%	5.8%	67.2%
2/7	31.1%	66.8%	22.1%	12.0%	66.6%
1/7	54.5%	46.4%	71.1%	11.9%	76.4%
1/3	59.4%	45.9%	71.0%	26.6%	78.9%

Table 3: AIFS and CW effect

To explain this behavior, we have to understand how this is happening. Let's assume a busy saturated medium and lots of nodes waiting for a free channel to transmit data, so their backoff procedures are paused. If all the nodes are non-prioritized, some of them will eventually collide when the channel becomes available, if the had the same BT value. However, if one of the nodes is prioritized, it could access the channel one SlotTime earlier than all its neighbors since it only had an AIFS=1 whereas the non-prioritized had AIFS=2.

Two cases can lead to this situation: First, during a busy period, the prioritized node creates a packet an chooses a BT > 0. When possible, it decrements the backoff timer and pauses with a BT=1. With the channel becoming available again, it can only collide with nodes having generated a packet during the last busy period and a backoff timer value of 0, which is very unlikely. Second, again assuming a busy channel, the prioritized node generates a packet and gets a BT=0 setting. This node now cannot collide with a non-prioritized one because the latter would have to wait at least one more SlotTime before decrementing its BT value. To determine which of these two situation has more impact, the results of the simulation have been further investigated - see Table 3.

- The number of successfully received packets is specified by the parameter **RcvPakts**
- When a packet is sent after the backoff timer has been paused with BT=1, **SentBT1** ist increased by 1
- **RcvBT1** indicates the number of successfully received SentBT1 packets
- **SentBT0:** This is the number of sent packets after new backoff timer has been initialized with a starting value of BT=0
- As before, **RcvBT0** states the number of received SentBT0 packets

The analysis shows that most of the packets have been sent after a paused BT=1 - see columns SentBT1 and SentBT0. Column RcvBT1 shows very well the Influence of a low AIFSD value.

With priority access, chances for a successful reception are three times higher than without. If we take a look on the RcvBT0 values, the chances for a colliding SentBT0 packet are very low, as mentioned before. One could expect there cannot occur any collision, but the hidden terminal problem exists. In this case, the reception of data is compromised by a node close enough to the receiver but outside the sender's carrier sense range so it cannot be detected by the sending node.

6.2 Dynamic scenario

So let's have a look on the more realistic and more important scenario with added mobility. In contrast to the static scenario, now, all the cars are moving on a circular highway.

6.2.1 Setup

Again, we have an eight-lane highway, but now, the road is circular and there are four lanes in each direction, as can be seen in Figure 4. The circle has an inner radius of 350 meters

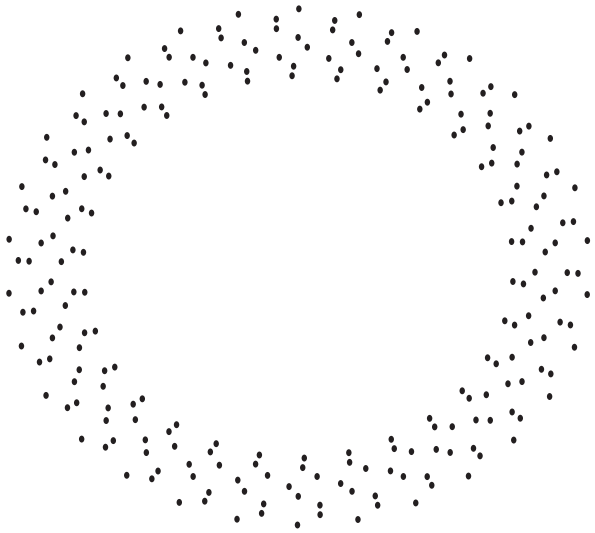


Figure 4: Scenario 2 setup

to avoid interferences caused by cars on the opposite side of it. The cars are distributed uniformly along the circle and drive with a constant speed of 55 km/h to 120 km/h which has been randomly assigned to all the cars of one lane. So every car in each lane travels at the same velocity and no crashes can occur. Every car except one has access category AC=0. The high- prioritized car is placed on the 6th lane counted from the middle of the circle. Four sub-scenarios were created differing in intended communication range and packet size:

1. 100 m communication range, 200 Bytes packet size
2. 100 m communication range, 500 Bytes packet size
3. 200 m communication range, 200 Bytes packet size
4. 200 m communication range, 500 Bytes packet size

All these scenarios use the two-ray-ground model. So, the last one was re-run using the Nakagami distribution.

Scenario	Channel load	Priority CAT	Non-Priority CAT
100m/200B	1.28 Mbps	0.4 ms	0.9 ms
100m/500B	3.2 Mbps	1.6 ms	4.8 ms
200m/200B	2.56 Mbps	1.2 ms	3.9 ms
200m/500B	6.4 Mbps	3.6 ms	16.4 ms
200m/500B (Nak)	6.4 Mbps	9.0 ms	26.5 ms

Table 4: Dynamic senario: channel access time (CAT)

6.2.2 Results

The most interesting metrics in this scenario are channel access time, a standard metric defined by the time from packet creation to sending it to the channel an the already mentioned probability of reception, the percentage of successfully received packets by the cars at a certain distance from the sender, ± 2.5 m. The latter has been chosen because one of the main goals of a VANET is, in a case of emergency, to inform as many other cars as possible, which are intended to receive that data.

Table 4 contains the measured channel access times for both prioritized and non-prioritized nodes in different environments. As we can see, the channel access time is much shorter for priority nodes. Figures 5, 6 and 7 visualizes the probability of reception for the 100m/500B-, 200m/200B and 200m/500B-scenarios using the non-deterministic two-ray-ground radio propagation model. The figures show, that on a saturated channel, using a priory access mechanism pays off more (about 150% improvement) than on a less loaded channel (only about 10%). If you take a closer look at the figures, you will probably see the decrease of probability of reception at about two thirds of the intended communication range. This is one of the consequences of the hidden terminal problem, where nodes outside the sender's carrier sense range are interfering the reception of the packets by its designated receivers.

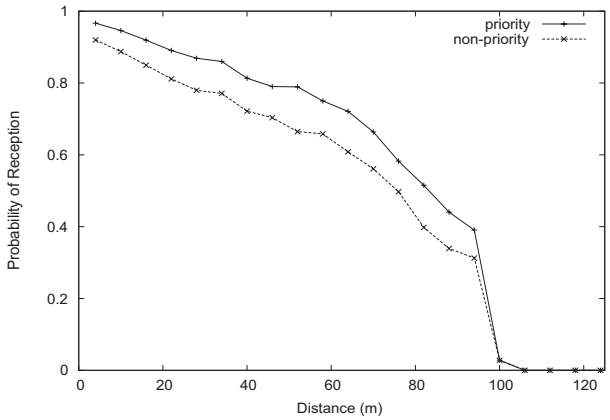


Figure 5: Probability of reception for 100 m communication range and 500 Bytes packet size

As the previous results of this scenario were run using the two-ray-ground model, one more pass with the Nakagami distribution model was done. As viewable in figure 8, probability of reception drops as channel access lasts longer. This shows very well, how big the impacts of using another model can be. But still, the reason for this collapse remains unknown. That's why another investigation was undertaken to study the sensed packets received per second (Sens. Pkts/s) by a node and the ratio of the channel idle time (Ch. Idle Time).

- **Sens. Pkts/s** is the average number of received packets per second with power geater or equal than the Carrier Sense Threshold, which is the defined power level the node can sense.

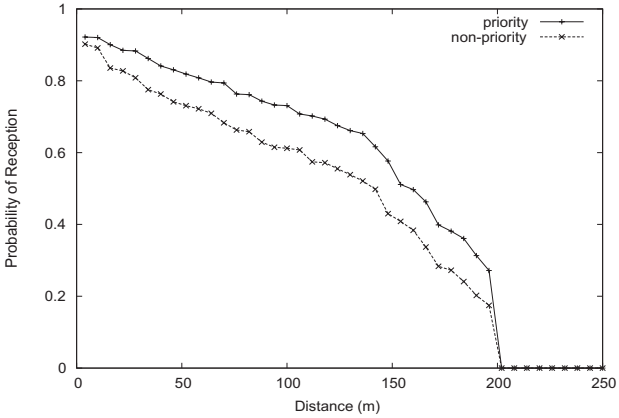


Figure 6: Probability of reception for 200 m communication range and 200 Bytes packet size

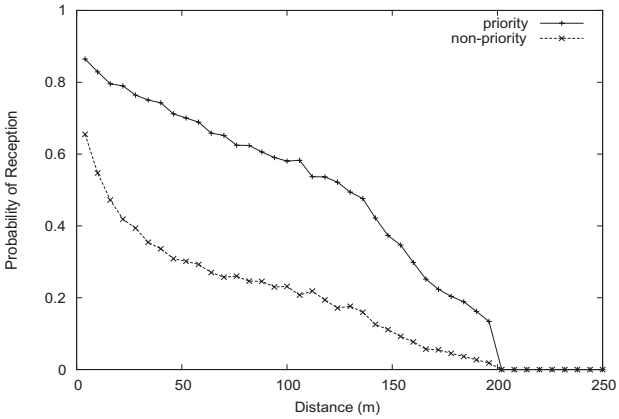


Figure 7: Probability of reception for 200 m communication range and 500 Bytes packet size

- **Ch. Idle Time** is defined by the time ratio, that the node senses the channel as idle. This value is calculated for the whole simulation time.

Scenario (200m,500B)	Priority		Non-Priority	
	TRG	Nak	TRG	Nak
Ch. Acc. Time	3.6ms	9.0ms	16.4ms	25.5ms
Sens . Pkts/s	3325.2/s	3093.2/s	3324.6/s	3096.8/s
Ch. Idle Time	10.8%	4.4%	10.6%	4.4%

Table 5: Comparison of channel access time, Sens. Pkts/s and Ch. Idle Time

The results in Table 5 show that the node in the two-ray-ground scenario senses more packets, but the node in the Nakagami scenario actually senses a busier channel. The explanation to this could be a bad distribution of the packets over the time resulting in longer busy periods. In the Nakagami scenario collisions mainly occur because of the existing hidden terminal problem and not only for nodes having the same BT value. Taking this additional amount of collisions into consideration, the performance loss in the Nakagami scenario is not such a big surprise.

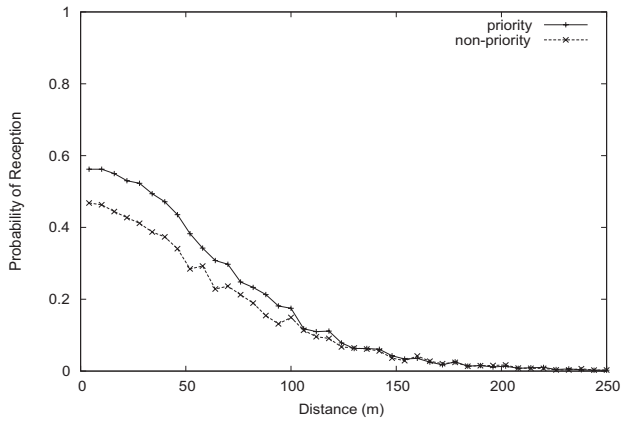


Figure 8: Probability of reception for 200 m communication range and 500 Bytes packet size using Nakagami distribution

7 Conclusion, future work

As we could see, the two very important metrics in vehicular ad hoc networks, probability of reception and especially channel access time, can be improved by adding priority access mechanisms to the protocol. The efforts only pay off when used in a saturated environment with a high channel load - a situation that can be expected if VANETs become popular. On the other hand, a good radio propagation model has to be chosen. The partly tremendous differences between two-ray-ground and Nakagami model show, that the performance fluctuation caused by a non-deterministic propagation model can be enormous. To further improve the performance of VANETs, there a still some factors which haven't been investigated in this paper. First, the service quality could not only been enhanced by tuning the protocol, but also by using further developed hardware technologies like better antennas, or enhancing the communication range by better retransmission and multi-hop relay strategies. Another thing not taken into consideration is the temporal and spational correlation of the power strengths. In this paper, every node has a individual power level not being influenced by its neighbors.

The paper shows in a good and understandable way the importance of a priority mechanism. This mechanism is implemented by tweaking only two parameters, the AIFS value and the size of the contention window. But unfortunately, we can not be sure of the significance of the improvements. As we have seen, the utilization of another model can have a tremendous effect on the yielded results. After all, only a real world experiment could prove or disprove the results of this work.

References

[802] IEEE 802.11e/D4.4, Draft Supplement to Part 11: Wireless Medium Access Control (MAC) and physical layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS), June 2003.

[Naka60] M. Nakagami. The m-distribution, a General Formula of Intensity Distrbution of the Rapid Fading. *Statistical Methods in Radio Wave Propagation*, 1960.

[ns2] The Network Simulator - ns-2, http://www.isi.edu/nsnam/ns/.

[TMJH04] Marc Torrent-Moreno, Daniel Jiang and Hannes Hartenstein. Broadcast reception rates and effects of priority access in 802.11-based vehicular ad-hoc networks. In *VANET '04: Proceedings of the first ACM workshop on Vehicular ad hoc networks*. ACM Press, 2004, S. 10–18.

[wir] IT-Infothek - Wireless LAN, http://www.it-infothek.de/fhtw/semester_2/re_od_08.html.

List of Figures

List of Tables

Review of "Vehicle-to-Vehicle Safety Messaging in DSRC"

Philipp Woock

Abstract

This paper discusses the proposals of Xu, Mak, Ko and Sengupta on how to design a MAC protocol for vehicular ad-hoc networks that is suited for transmitting safety messages in the DSRC frequency band. Their protocols are based on packet repetition. They explore the influences of transmission power, number of vehicles and resulting interferer number, message range and data rate on the feasibility of transferring safety messages. To be able to make propositions about the different protocol variants, they define their Quality of Service measures: The probability of reception failure (PRF) and the channel busy time (CBT). All protocols are evaluated and compared in terms of PRF and CBT to find out which of them with which parameter settings are usable in vehicular ad-hoc networks. Xu et al. show that their research is a promising starting point for further development.

1 Introduction

1.1 Motivation

As there is an ever increasing number of cars on the streets, car crashes and traffic jams increase as well. To enhance overall security, cars somehow should warn each other about dangers. To do this, the cars form a vehicular ad-hoc network. Then the driver (perhaps assisted by an ESP-like device) can react to such a warning prior to seeing the danger with his own eyes. How these messages between the cars are sent is topic of the paper.

1.2 Safety Messages in general

Safety Messages cover information regarding crashes, traffic jams, hard braking manoeuvres, fishtailing cars, slippery ground and many other situations where other road users are in danger and should be notified. Each vehicle will be sending them continuously. Vehicles that recognise a hazard or become a hazard themselves create messages describing type and position of the hazard and send them to all cars within reach.

Of course the real world situation is not that easy. Sending and receiving messages to and from lots of cars need elaborate mechanisms. Otherwise messages get lost or just come too late. Questions arise: Will there be too much messages when there are lots of cars? Will too much safety messages hinder other uses of the system? How big should such a message be? And there are even more questions. The possible solutions to those problems are presented in the paper of Xu et al.

As stated above, a safety message should contain mainly type and GPS position information. Therefore it can be kept relatively small sized. For we want the messages to be transmitted quickly, it is helpful to have them as small as possible. Furthermore, smaller messages require smaller bandwidth.

Furthermore these messages have only a short lifetime because they are only useful as long as they are recent.

1.3 Broadcast vs. unicast

The vehicular ad-hoc networks are different from networks one is usually familiar with. In most internet situations, communication is unicast. That means that two agents communicate with each other. In our vehicular network one vehicle should deliver the safety message not to just one other vehicle but to *all* vehicles that can be reached because the safety messages are important for all vehicles in the sending range. This means we need broadcast communication.

Assume we were using unicast communication: Each vehicle would need to establish a separate connection to every other vehicle in reach. This alone squares the number of connections. In addition there is the overhead of setting up these connections and closing them. In vehicular ad-hoc networks there are many changes in the neighbourhood which would cause continuous handshaking and termination of connections. This is extremely complicated and not suited to get a first idea of how vehicular ad-hoc networks will look like.

Moreover they assume that no roadside radio stations for synchronisation purposes are installed, as this is very costly and would cause huge efforts to install them. A pure ad-hoc network is the goal.

1.4 Some notions

Not every reader may be familiar with basic notions in this paper. I want to clear up those notions first:

What mainly hinders the unfailing reception of safety messages in a wireless ad-hoc network is interference and attenuation.

1.4.1 Interference / SINR

Interference arises when two senders A and B concurrently send a signal and a receiver within the range of both of them tries to decode the signal. He receives different signal power from A and B depending on where he exactly is. If the received signal power of A and B is roughly equal, neither one of the signals can be decoded correctly. For successful decoding, one signal must be stronger than the other by a certain margin. Just imagine how hard it is listening to two persons talking to you simultaneously and equally loud.

A more formal definition:
"'The interference range is the range within which transmission from an interferer makes the signal-to-interference-and-noise-ratio (SINR) of the legitimate receiver smaller than a certain threshold, so that the legitimate receiver cannot correctly receive the message from the legitimate transmitter"' [Wu03]. See also Appendix A of [XuMS04].

1.4.2 Attenuation

Attenuation describes that every radio signal is damped as the distance to the sender increases and due to obstacles and the surrounding air. At a certain distance the attenuation is so high that the signal has become indistinguishable from thermal noise. To estimate the real attenuation there are several so-called path loss models. The paper uses the two following models.

1.4.3 Friis and two-ray model

To estimate the real transmission power the destination radio receives, one makes some simplifications. These models are two very basic models that certainly don't fully match the situation on the roads. But to get an idea if the system will be feasible at all they are sufficiently accurate.

The Friis Free-Space model is used for short distances, where source and destination radio have a line of sight between them. To calculate it, it takes nothing more than wavelength, antenna gains, transmission power and system loss into account. The received power decreases with d^2, d being the distance in-between.

The two-ray ground reflection model is used for longer distances where reflections from the ground have to be considered additional to the line of sight. It is not well suited for short distances. Depending on the length difference of the two paths and the wavelength, the signal partially cancels out at the receiver or doubles in received power. The signal power at the receiver decreases with d^4. [Ye00]

1.4.4 Hidden terminals

Hidden terminals are nodes that are not aware that another node wants to receive data. It is possible that they start their own data transmission and interfere not-knowingly with that other communication. With a RTS/CTS protocol this problem is solved. Sometimes hidden terminals are referred to as hidden nodes. (See Figure 1)

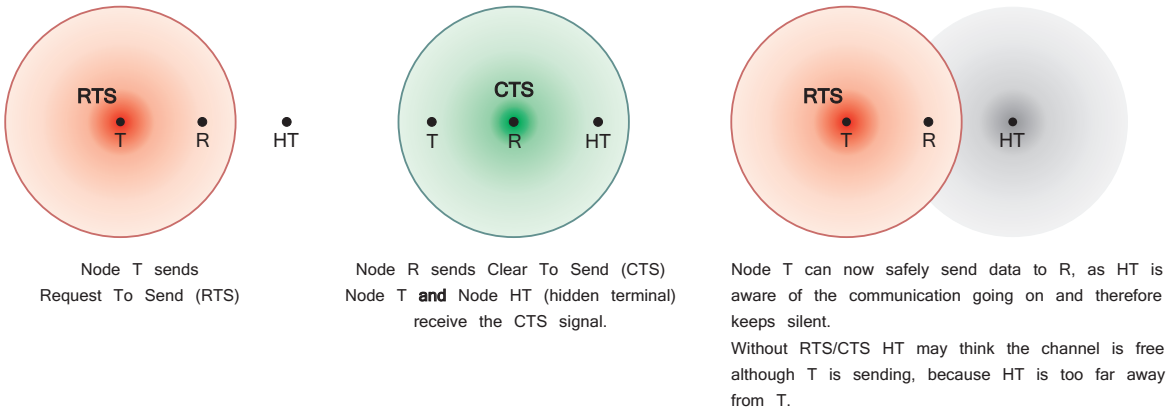


Figure 1: Hidden terminal problem

1.4.5 DSRC - Dedicated Short Range Communications

The FCC granted 75 MHz at 5.9 GHz for uses that "'increase traveller safety, reduce fuel consumption and pollution and continue to advance the nation's economy"'. This DSRC band has got seven channels. Six service channels and one control channel. All vehicles are monitoring the control channel regularly. Messages have one of three priorities: "'safety of life"', "'safety"' or "'non-safety"'. The paper assumes that safety messages take place only in the control channel. (It doesn't matter in which channel the messages are actually transported as long as all vehicles monitor this channel.) [XuMS04]

DSRC is based on 802.11a technology, due to the used frequencies are very similar. 802.11 technology is well-known, widely available and comparably cheap. 802.11 radios can transmit

messages safely over distances of 200 to 300 metres. The data rates that 802.11a provides are 6, 9, 12, 18, 24, 36, 48 and 54 Mbps.

The DSRC service channels are planned to be used for additional services as toll collection, updating the navigation system, entertaining services and so on.

1.4.6 Simple collision model

The simple collision model is fundamental for the considerations by Xu et al. That means if two messages overlap in time and are within interference range of each other, neither one can be decoded correctly and both are lost.

2 Why something new?

There are lots of MAC protocols around. Is none of them suitable for vehicular ad-hoc networks? A summarisation of different known MAC protocols and how applicable they are for ad-hoc networks can be found at [Peng04]

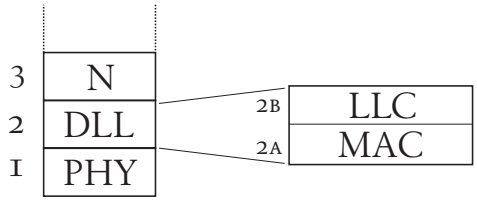


Figure 2: Where the MAC layer is located in the ISO/OSI layer model

Xu et al. rule out the known protocols for different reasons:

2.1 TDMA, FDMA, CDMA, SDMA

The Time / Frequency / Code Division Multiple Access protocols are not useful, because they need a central station that allocates time slots, frequency bands or codes. In our ad-hoc network there is no central station. Communication should be self-organizing. Furthermore we assume random access which the xDMA protocols do not do natively. They could emulate it with a very quick reassignment of slots. But again, a centralized coordination station would be needed. The authors don't mention the SDMA protocol whereas there is interesting research on how SDMA seems to be quite well-suited for vehicular ad-hoc networks. It is inattentive to not mention it and why it is possibly not suited. SDMA for mobile ad-hoc networks is discussed in [BaVa01].

2.2 MACA, MACAW, FAMA

Those not so well-known protocols are based on RTS/CTS and are therefore not suitable, because our communication is broadcast where sender and receiver do not know each other. Further reading about MACA, MACAW and FAMA: [Nara02] and improvements on them: [Wu04]

2.3 Other protocols

HIPER-LAN/1, Black Burst and the Enhanced Distributed Coordination Function (EDCF) of the 802.11e standard have some Quality of Service (QoS) features but they are not suitable anyway: HIPER-LAN/1 and Black Burst can't combat hidden terminals. EDCF gets into trouble when the number of contending packets (of the same priority) is high, which is undoubtedly the case for safety messages in vehicular ad-hoc networks.

2.4 Cellular networks

Cellular networks can manage the communication between fast moving vehicles without that much of a time delay, but they need base stations. Furthermore, cellular communication doesn't allow for two mobile nodes to directly communicate with each other. Each node first has to communicate to the base station (two hop communication). Those cellular base stations are much more expensive than simple 802.11 units that DSRC uses.

3 The scenario

3.1 Estimations about road traffic amount

How many cars take part in that ad-hoc communication? How fast do they travel? These questions are important to estimate needed bandwidth, transmission power and Quality of Service (QoS). Xu et al. assume roads with a maximum of 8 lanes and smallest vehicle distance of 10 m (per lane) which is reasonable. The full parameter list is in Table 1.

Moreover they assume the maximum speed to be 90 mph (ca. 145 km/h) and a message creation interval of 50 ms which leads to a movement of 2 metres between two messages. Assume German motorways where two cars in opposite directions may have relative speeds up to 500 km/h: Even then there are only about 7 metres between two messages. Xu et al. consider message generation intervals up to 500 ms. This is too long when travelling at high speeds. On that aforementioned German motorway the cars may send as little as two to four safety messages to each other while they are within the message range. So in Germany it may be necessary to tune the parameters according to the travelling speed. Ideally, the message creation rate increases as travelling speed increases and vice versa.

Message generation interval (ms)	50, 100, 200	
Packet Payload Size (Bytes)	100, 250, 400	
Data Rate (Mbps)	6, 9, 12, 18, 24, 36, 48, 54	
Average Vehicle Distance (m)	10 (jammed)	30 (smooth)
Message Range (m)	10-100	30-300
Lane Number	4,8	

Table 1: Traffic parameter ranges

We saw that what we need is broadcast communication without roadside stations. There is no need to relay safety messages from one to another vehicle due to the 200 m – 300 m range of 802.11 is enough by far to react to dangers. That range achievable with single-hop communication is sufficient for most vehicles to come to halt.

3.2 The used QoS measures

Xu et al. introduce two measures to describe the QoS. The PRF (probability of reception failure) and the CBT (channel busy time).

3.2.1 PRF - Probability of Reception Failure

The PRF describes how likely it is that a message can't be received properly. It is written by

$$PRF(L, \tau)$$

where L denotes the distance between sender and receiver and τ denotes the message lifetime.

The PRF is the probability that a randomly chosen message by a randomly chosen sender will not be received by a randomly chosen receiver at distance L within time τ. (Definition from [XuMS04])

To keep track of all the surrounding vehicles, each vehicle needs to have a real-time estimator that reconstructs the traffic situation from the safety messages. The estimator should be able to interpolate some missing (faulty transmitted) messages. Subsequent missing messages however should wishfully not occur. So PRFs in the range from $\frac{1}{1000}$ to $\frac{1}{100}$ or lower should be appropriate. These numbers are just a guess, since there are no real estimators yet. But these PRF values ensure that an estimator shouldn't get into trouble.

The further away the receiver is from the sender the higher is the possibility that interference occurs. Hence the worst case to cover is PRF(*Message_Range* , *Message_Life_Time*). When the paper speaks of PRF values, these worst case values are meant.

3.2.2 CBT - Channel Busy Time

The CBT describes what fraction of time is left over in the control channel for other uses than safety messages.

$$CBT := \frac{T_{safety}}{T}$$

It can only be expressed as a fraction because there is no control channel protocol yet, and it therefore cannot be ascertained how much remaining time would be sufficient for other uses. Everything one can do at the moment is to keep the CBT as low as possible to allow for the highest possible amount of other traffic as well.

4 MAC layer design considerations

The MAC layer has to decide when to send a messages and at which transmission power. The goal is that messages can be received best possible.

As we use broadcast communication policy we have no receiver feedback. This doesn't allow us to get knowledge about the receivers and thus they cannot give us information with which we could achieve higher reception reliability. We need to find other ways to increase it.

Xu et al. evaluate several variants of protocols that simply send repetitions of the message. Some of the protocols they developed do additional Carrier Sensing (CSMA). Repetition takes place in the lifetime of the safety message i.e. while the message is useful. (See Figure 3)

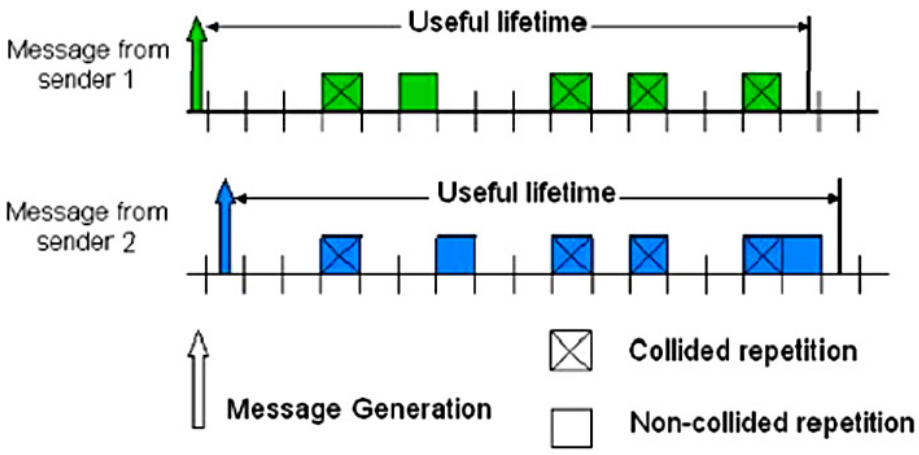


Figure 3: Repetitive transmission and collision

Figure 4: The MAC extension layer

The network layer 2 consists of two sub-layers. The LLC (logical link control, IEEE 802.2) and the MAC (medium access control) layer. Xu et al. add another extension layer between them. (Figure 4)

When a message is to be sent, the extension layer changes to repetition generation state. Here the Packet Event Queue (PEQ) is formed and time slots are assigned to the events (repetitions). Then it transits back to idle state. When the time has come for such a repetition to be sent (i. e. the timestamp belonging to the event in the PEQ is reached), the extension layer switches to the dispatch state, sends the packet down to the MAC layer and returns to idle state. (Figure 5)

On receiving a message from the MAC layer the message ID is inspected and the extension layer turns to repetition removal state. If the ID is new, the message is new and hence handed over to the LLC layer. If it is already known the message is simply discarded.

5 Protocol proposals

Xu et al. were not satisfied with any of the existing protocols (Section 2), so they designed some new protocols. In fact these are quite simple protocols basing on ALOHA respectively slotted ALOHA that try to guarantee a successful transmission by repeating the message. The basic concept is to send the same message so often that at least one of these repetitions will reach the destination. The lifetime of a message is divided into time slots with every time slot allowing one repetition of the message to be sent with actual message size and transmission rate.

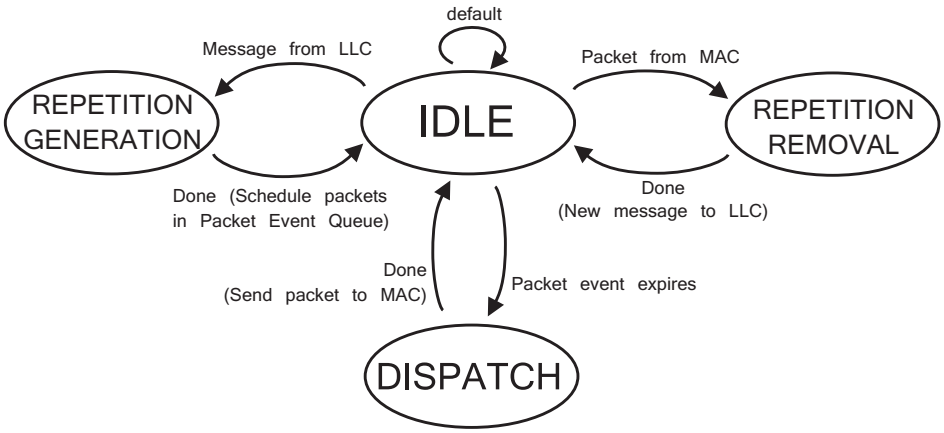


Figure 5: MAC extension layer state machine

5.1 AFR – Asynchronous Fixed Repetition

AFR is configured with k being the fixed number of repetitions. The protocol randomly selects k time slots out of the maximum of n slots the lifetime of the message provides. So each message is repeated k times. AFR is without carrier sensing.

5.2 APR – Asynchronous p-persistent Repetition

In APR each of the n time slots gets a message with probability $p = \frac{k}{n}$. Thus the expectation value of how many slots are filled is k (as in AFR) but not every message is repeated exactly k times. APR like AFR does no carrier sensing.

5.3 SFR – Synchronous Fixed Repetition

SFR is nearly the same as AFR. The only difference is that in SFR the time slots in all nodes are synchronised to a global clock. That means that repetitions overlap either completely or not at all. This reduces the number of collisions and hence increases throughput. This technique is known from slotted ALOHA.

5.4 SPR – Synchronous p-persistent Repetition

The same as APR but with synchronised slots as in SFR.

5.5 AFR-CS – Asynchronous Fixed Repetition with Carrier Sensing

AFR-CS is standard AFR improved with carrier sensing. How the modified MAC layer accomplishes carrier sensing is shown in Figure 6. Whenever there is a packet handed over from MAC EXT layer, the MAC switches to carrier sensing state. If the channel is busy, the packet is dropped. If it is idle, the system turns to MAC TX (transfer) state and delivers the packet to the PHY layer. On receiving a packet from the PHY layer, the packet is checked for integrity and handed over to the MAC EXT layer if error-free. Carrier sensing helps that packets from other senders are not destroyed through interference.

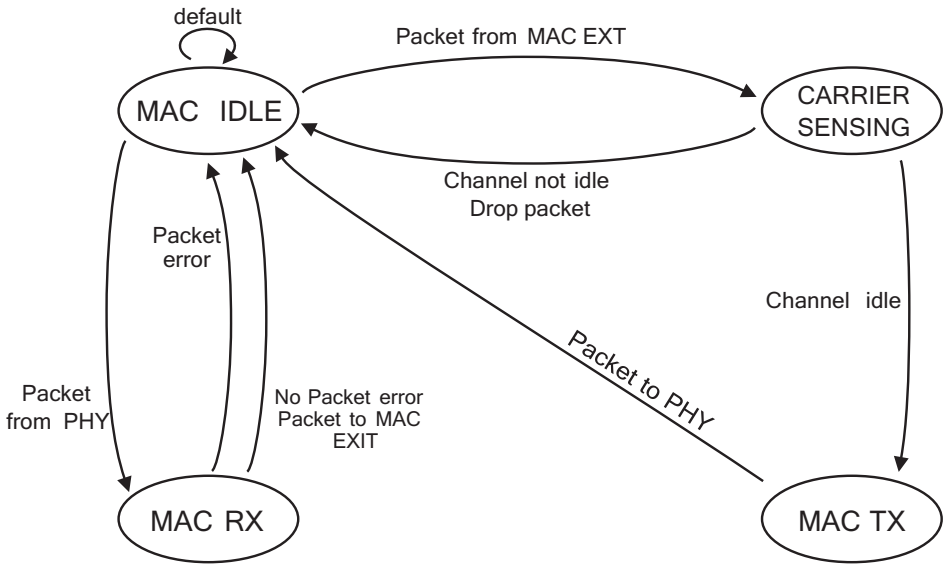


Figure 6: MAC layer with carrier sensing

5.6 APR-CS – Asynchronous p-persistent Repetition with Carrier Sensing

The same as AFR-CS but with p-persistent slot allocation as in APR.

6 Analysis

6.1 Simplifications

Of course some simplifications are made:

The analysis is not done for all possible parameter combinations as this would be a huge amount of work with possibly only little more worth. Xu et al. chose to do only simulations for a certain representative parameter set and look further only where it seems to be promising.

They mainly use these nominal parameters: Messages are generated every 100 ms and therefore have a lifetime of 100 ms. They set the message range to be 80 m. A four-lane road is assumed with 30 m distance between vehicles on each lane. The formula

$$\text{Interferer number} = \frac{2 \cdot \text{Interference Range}}{\text{Metres per vehicle}} \cdot \text{Lane number}$$

tells us that 75 interferers are involved. How to calculate the interference range can be found in Appendix A of [XuMS04]. It depends on message range, sender-receiver distance and data rate.

Another simplification is that only homogeneous traffic is studied. The transmission power, repetition number k, data rate and packet size are the same for all nodes. This is a not a heavy limitation because road traffic characteristics do not change that much between two Safety Messages. In my opinion, some adaption is needed though. Assume the change from travelling at motorway speeds to a traffic jam. The estimator must be aware of the new situation and take care of it (e.g. adjust message range). This is not subject of the paper but reminds that this is still uncovered.

6.2 Mathematical assumptions and simulation tools

Xu et al. assume the message generation process being Poisson, because Poisson is "'an approximation of the superposition of large number of independent periodic processes with various interval and starting time"'. They determine very close upper and lower bounds for the event that all repetitions of the message fail.

In addition to the mathematical analysis they perform simulations on a newly composed DSRC simulator. First they use the SHIFT simulator to simulate the vehicles driving subject to validated models on realistic roads. The trajectories of the vehicles (which are also the trajectories of the radios) are passed on to the NS-2, an open source network simulator that simulates the data transmissions using an 802.11a radio model and the proposed protocols. The data from NS-2 is post-processed to obtain CBT, PRF and the probability of long bursts of reception failures. Besides, they recoded the data structure of NS-2 to have linear execution time. That made it possible to simulate up to thousand participating vehicles.

6.3 Findings

In qualitative terms, if you increase the number of repetitions, the chance that one of them will be transmitted successfully should increase, too. Repeating very often however leads to more collisions and higher channel usage and is therefore not helping but harming. So there is an optimum number of repetitions. Of course this optimum depends on message generation rate, message range, number of participating radios (traffic density), message size and so on. It is clear that more repetitions may help to reduce PRF while increasing CBT at the same time, and vice versa.

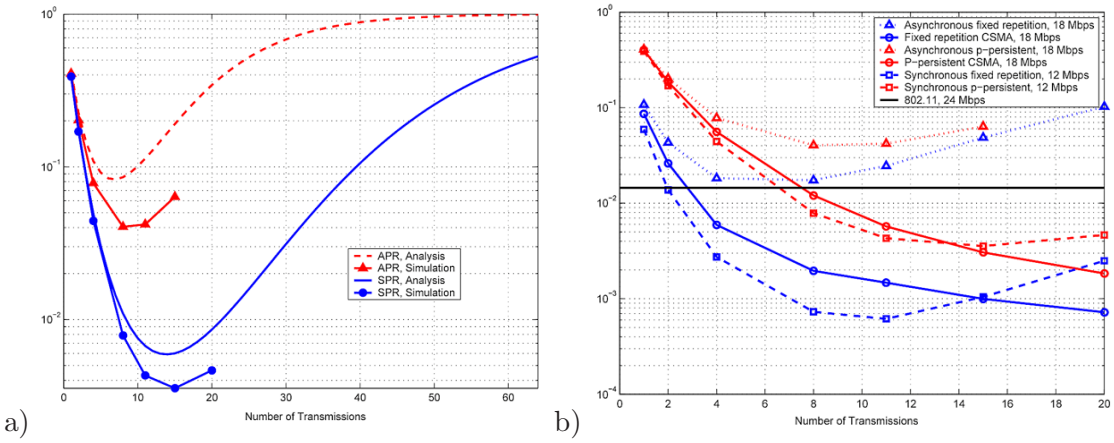
Raising the transmission power increases the probability that a message is received properly because in the near field the SINR is increased, but on the other hand the interference range is extended, too. If everyone was sending with maximum power, most of the messages would be received as noise. Imagine humans that tell stories to each other simultaneously: With everybody shouting as loud as he can, only few pieces can be understood correctly. (You understand more the nearer you are to the speaker). With everybody only whispering it's the same. That means there exists an optimum transmission power depending on the characteristics of the vehicular neighbourhood.

In their results the mathematically calculated PRF values are higher than the values from the simulation across the board. They say this is because the Poisson based model allows it that a message may collide with the previous message from the same node because their lifetimes could overlap. In simulation, two messages cannot overlap at the same node, which is just reasonable. It doesn't make sense that one node sends two different safety messages at a time, so there is no collision. Thus the PRF values from the simulation are smaller than those from mathematical analysis. Except for this, the simulation curves show the same behaviour as the mathematical calculation does. We see that simulation behaves nearly as the mathematical model predicted. (Fig. 7a)

As one can expect, the synchronous protocols and those with carrier sensing perform better than the pure asynchronous ones. The protocols with p-persistent repetition do worse than those with fixed repetition number (Fig. 7b). Xu et al. presume that the p-persistent protocols gain less in repeating more often than what they lose in doing less repetitions. They don't investigate further and there is no need to do so. Simulation showed that there is just no advantage in transmitting in that p-persistent manner. As for the synchronised protocols, they reached the lowest PRF but need a global synchronisation mechanism that is not there. As stated earlier we want to avoid any additional roadside devices.

Concerning the CBT, the AFR-CS protocol is much more "'channel friendly"' as it shows significantly lower CBT than both AFR and SFR which perform nearly identical.

The analysis of the PRF values of the different protocols shows that synchronous protocols and those with CSMA are notedly better than AFR or APR (Fig. 7b). In their model, the more the repetition rate increases the more it became apparent that the CSMA protocols have an advantage, even over the synchronous ones. Obviously, AFR-CS is the protocol of choice and that's why Xu et al. focus on AFR-CS in the following steps.



a) b)

Figure 7: a) Comparison of simulation and mathematical PRF results b) PRFs for discussed protocols in nominal setting

In their model it makes little to no difference how a number of interferers is distributed on the road. The single number tells about which PRF can be achieved at which CBT. If there are 150 interferers and a message range of 80 m it nearly doesn't matter if they are at 30 m distance on 4 lanes or if they are at 15 m distance on 8 lanes (Fig. 8).

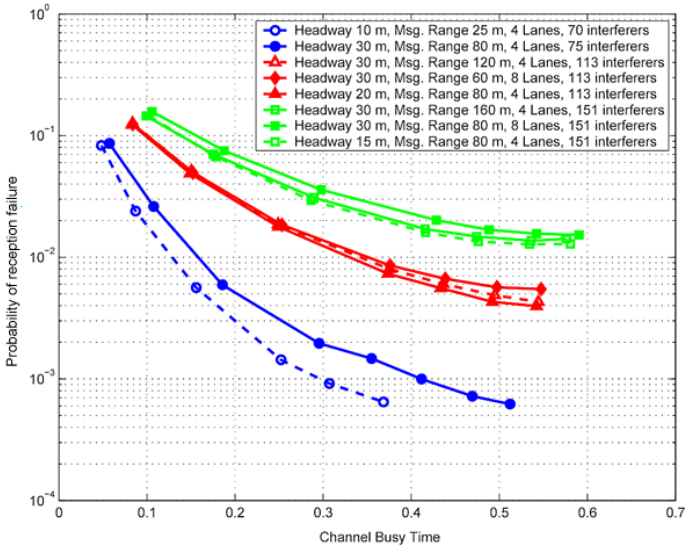


Figure 8: Independence of interferer distribution

The influence of data rate on PRF and CBT is as follows: If the data rate is high, the message needs less time to be transmitted and therefore lowering collision probability. But to achieve high data rates you also need high transmission power. That in turn leads to a higher interferer number. Considering the possible 802.11 data rates, the lowest PRF values are at 18 Mbps for the AFR-CS protocol with 24 Mbps being very close. This optimum is

different between the protocols. Most have it at 18 Mbps, the synchronous ones at 12 Mbps and pure 802.11 has it at 24 Mbps.

It is shown that the probability of message failure bursts is negligible low, in fact one order of magnitude lower than PRF values for single messages.

6.4 Feasibility

The feasibility depends on the desired values of PRF and CBT. Completely occupying the channel is not an option and PRFs don't get arbitrarily low. Xu et al. evaluate only AFR-CS as it performed best. They tune the parameters for optimum repetition number with optimal transmission data rate and minimized power while still covering the message range. They consider PRF below $\frac{1}{100}$ and CBT lower than 50%. They evaluate message sizes of 100, 250 and 400 bytes. For example they regard message size 250 bytes with message generation interval 200 ms feasible up to 110 interferers (Fig. 9). This for example equals a four lane road with inter-vehicle distance of 20 m and a message range of 80 m. As PRF and CBT depend nearly only on interferer number, this situation is the same as an eight lane road, inter-vehicle distance 30 m and 60 m message range.

These traffic situations are common, but in more extreme traffic situations with 150 interferers it is impossible to achieve a PRF of $\frac{1}{100}$ with keeping CBT below 60% (Fig. 8). If this is feasible in the end cannot be told at the moment because is not known yet how other traffic on the control channel will look like. Should 25% of the channel be enough for all other duties one can have more interferers while having low PRF. If it turns out, that safety messages can have only a maximum of, say, 35% of the channel, then the AFR-CS protocol will not work in situations with higher interferer number than about 113 as you see in (Fig. 8).

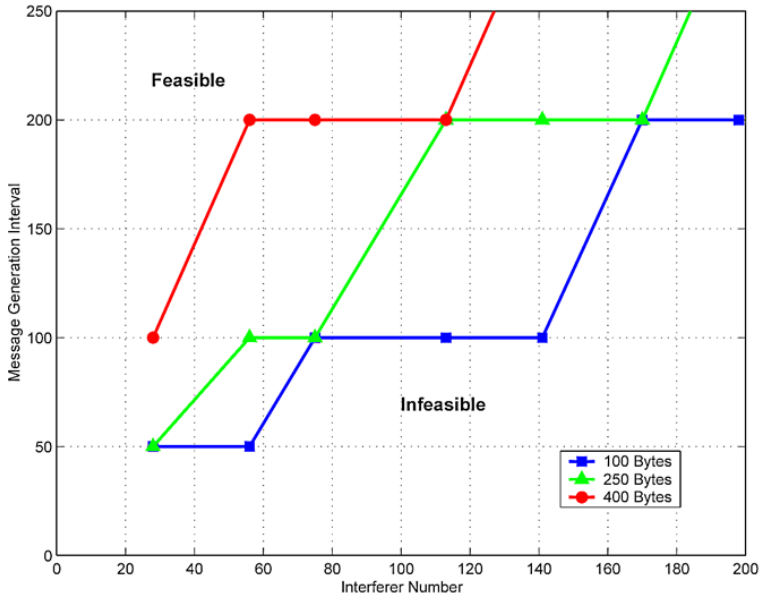


Figure 9: Feasibility region division lines for AFR-CS with $PRF < \frac{1}{100}$ and $CBT < \frac{1}{2}$. Feasible and infeasible labels have to be read separately for each line.

Xu et al. say that message generation intervals of 200 ms are feasible combined with message sizes of 250 bytes and interferer number greater of 140 if the network and application designers work together. They claim that message generation intervals faster than 200 ms are not needed as this is still faster than a human driver can react and that GPS updates position information only at 5 Hz. They aver that with 140 interferes one is able to cover a message

range of about 150 m which should allow most vehicles to come to halt. In my opinion the message generation interval should not get longer than 200 ms because the value of the safety information decreases quickly when it arrives late. This is difficult with the proposed protocols as either message sizes must be small or PRF/CBT values are not as desired.

6.5 Discussion and open issues

This paper is well thought out and Xu et al. use elaborate methods to describe and simulate their goals. Of course, as there is only very little information on how other services in the control channel will look like, it is somehow a shot in the dark. The researchers were forced to use the CBT which is a relatively simple measure. It is a similar situation with PRF values. As there is no estimator yet they have to make assumptions what an estimator might be capable of in the future. They wish PRFs to be lower than one failed message out of hundred. Perhaps a fully developed estimator can handle one failed message in twenty which would allow PRFs up to $\frac{1}{20}$. Then feasibility studied basing on PRF of $\frac{1}{100}$ is not beneficial. Look at the CBT measure: Perhaps one of these six service channels in DSRC will be relabelled as a security channel that carries only safety messages and will be mandatory for all vehicles to be monitored (in addition to the control channel). Then the CBT would be meaningless as there was no other traffic to consider in that channel. All those things depend heavily on future development. But for the time being the assumptions in [XuMS04] are reasonable and provide a solid basis for further work.

What the authors describe in their work is how one can just add a MAC EXT layer to an existing MAC layer and get a mechanism that works well in many situations and is quite simple to build. From the viewpoint of reusing existing techniques this is just fine. But the underlying assumptions the authors use are too theoretical. Many real world difficulties are not considered at all or not enough. So I think it is not a fully realistic view how traffic looks like.

Safety messages need to be received with high reliability, otherwise they are no use. With the assumptions of the paper, I think that the results have to be taken with a grain of salt. They do show how far the development is and where there are still problems but the absolute numbers are moot. What works in their simulation is probably different to what works in reality.

They mention a truck between two cars that will just dampen the signal power very strongly. This is not covered by using free-space or two-ray model. Moreover, both models assume the broadcast area as perfect circles, which is unrealistic. Reflections and obstacles are not taken into account either. The shadowing model could have been used for better results [Ye00].

As cities install more and more 802.11a access points, the electromagnetic pollution will cause additional interference and thus complicate reliable reception of the safety messages, because the frequency bands are adjacent. For example: In 802.11b you have 13 channels but only 3 of them don't overlap. If there are more sources of interference, there are bigger problems to receive signals properly.

Another situation comes to my mind which has to be taken into consideration but isn't covered by the paper:
Vehicles on bridges that lead over another road cause additional interference which is not covered by the model. And they may get safety messages from below that don't have anything to do with their traffic situation. Here a digital map would be needed that helps filtering out safety messages that obviously come from another road.

The problem is that these situations are really important. In many cases there will be scenarios not matching with the simulation. The designers of the physical layer will do their

best but they won't be able to provide a completely faultless data stream to the MAC layer. I think that one should investigate into other, more sophisticated MAC protocols.

Another open issue is an adaptive method to control the message range. The vehicle has to notice that road characteristics have changed. But how good that works, heavily depends on the estimator that models the surrounding traffic situation.

What's more is that safety traffic differentiates itself into two priority classes which has not been modelled yet.

The paper gives a good insight of where the problems are, but fails to deliver something that is of instantaneous practical use.

References

[BaVa01] Soheila V. Bana and Pravin Varaiya. Space Division Multiple Access (SDMA) for Robust Ad hoc Vehicle Communication Networks. *The IEEE fourth international conference on intelligent transportation systems*, 2001. http://paleale.eecs.berkeley.edu/~varaiya/papers_ps.dir/Bana_Varaiya.pdf .

[Nara02] Mort Naraghi-Pour. Investigation of Media Access Control Protocols for Mobile Ad-hoc Networks. *NIST, Advanced Network Technologies Division*, 2002. http://w3.antd.nist.gov/pubs/mortrpt.pdf .

[Peng04] Yanlin Peng. Department of Electrical and Computer Engineering, Iowa State University, 2004. http://clue.eng.iastate.edu/~daji/seminar/new%20presentations/2004-09-22-Yanlin.ppt .

[Wu03] Dapeng Oliver Wu. Department of Electrical and Computer Engineering, University of Florida, 2003. http://www.wu.ece.ufl.edu/wireless/ad_hoc_networks.html .

[Wu04] Dr. Hongyi Wu. The Center for Advanced Computer Studies, University of Louisiana at Lafayette, 2004. http://www.cacs.louisiana.edu/~wu/576/MAC.pdf .

[XuMS04] Qing Xu, Tony Mak and Raja Sengupta. Vehicle-to-Vehicle Safety Messaging in DSRC. *VANET '04, Philadelphia, Pennsylvania, USA*, 2004.

[Ye00] Wei Ye. A chapter in ns Manual, 2000. http://www.isi.edu/~weiye/pub/propagation_ns.pdf .

List of Figures

List of Tables

Review of "Centimeter Vehicle Positioning and Lane Keeping"

Nils L. Roßmann

Abstract

Improving safety and driver assistance is today an important topic. More and more cars are equipped with navigation systems, but safety systems helping the driver to stay on the road/lane are still not available. This report describes how positioning systems can assist vision based systems in lane detection. Additionally positioning systems are very important for vehicular ad-hoc networks. In these context we are not interested in the relative position of the vehicle with regard to the lane boundaries, instead it is importand to calculate the absolute position and especially the information on which road and more precisely on which lane the car drives. The comparison with the visual system shows that positioning systems with access to very precise maps could fullfill the VANET requirements. As a consequence of higher precision in positioning systems there is also a higher map precision required. To refine the today available maps, a statistical approach is used which reaches a precision about 10cm.

1 Introduction

In vehicular ad-hoc networks it will be possible to inform the driver about the road-state in the near distance. To do this, every car collects information (e.g. speed, weather condition, lane surface, etc.) which could be requested by other cars or send out in dangerous situations. These messages could be received by many cars several kilometers away. After receiving a warning message from another car, the system has to decide if the information may concern the driver or not. To do this, especially the absolute position of the sender car and the own position is important. For example, when an accident happened in front of us, then it is important to know if it is on the same road, same lane or somewhere else. Crashes due to bad view are also a mayor problem. When there is a slower car on the same lane, not recognized by the driver, a car equipped with communication devices and a precise positioning system could warn the driver in time. So the positioning system precision is very important.

Beside this, lane keeping and departure warning would be an important safety improvement. Especially on bad wheather conditions and/or bad view. With such a system it is also possible to issue lane departure warnings if the driver falls asleep. For such application the relative position with regard to the lane borders is needed, which could be calculated by vision based systems.

Vision based Lane recognition is a good studied topic and there are many proven solutions for lane departure warning. There also exists systems which allow autonomous driving with a speed up to 160km/h. But there are still unsolved problems like a bad view due to rain/snow etc. or missing lane markings. So another system is needed to assist the vision based system. The idea is, using a very precise positioning system and a precise map to calculate the distance to the lane borders and compare the results with the vision based system. The authors implement a positioning system based on GPS, inertial measurement units and precise maps and a vision based system. This system ha s been tested on the A8 near Stuttgart.

The whole system will be described in the following sections. We start with the positioning system because for vehicular ad-hoc networks the absolute position is needed. If the positioning system is good enough to assist the vision based system, than it is also good enough to determine the current road and lane of a car. After describing the positioning system, a short introduction to the vision based system is given and then re results will be analyzed in concern of the vehicular ad-hoc network requirements.

2 Positioning System

Talking about nationwide or global positioning naturally GPS comes up. Due to it's global availability it's the most used system. Due to its low precision, there are several improvements needed to get the required precision, though the used approaches are described in the following sections. Another problem with GPS is, that there are outage gaps (e.g. in tunnels or in cities with high buildings around). To bridge these gaps, inertial measurement units could be used. They allow a high accuracy for a short time period, so they are sufficient to bridge short gaps. After calculating the absolute position of the car, it is necessary to determine the current location with regard to the map, which is done by a map matching algorithm. For navigation systems, it is sufficient to determine the current road, but to determine the current lane, a lane level precise map is needed. The whole system is described in the following sections.

2.1 GPS-Basics

The GPS system operates with 24-28 spacecrafts in an orbit about 20.000km around the earth. It's design guaranties that there are at least 4 and up to 12 spacecrafts in range.

Every spacecraft sends a signal which includes the send-time (t_i) and all information to calculate the spacecrafts position (x_i, y_i, z_i. With 4 spacecrafts it is possible to calculate the users position and the exact time without precise clock in the GPS-receiver.

So we have to solve four equations with the four unknown variables x,y,z,t:

$$\sqrt{(x_1 - x)^2 + (y_1 - y)^2 + (z_1 - z)^2} = c(t_1 - t) = d_1$$

$$\sqrt{(x_2 - x)^2 + (y_2 - y)^2 + (z_2 - z)^2} = c(t_2 - t) = d_2$$

$$\sqrt{(x_3 - x)^2 + (y_3 - y)^2 + (z_3 - z)^2} = c(t_3 - t) = d_3$$

$$\sqrt{(x_4 - x)^2 + (y_4 - y)^2 + (z_4 - z)^2} = c(t_4 - t) = d_4$$

d_i : distance to spacecraft i x_i, y_i, z_i : spacecraft i coordinates and time x, y, z, t : user coordinates and time c : speed of light

On the left side, the three-dimensional distance is expressed by the difference between the known spacecraft-position and the unknown user-position. This is equal to the distance calculated by multiplying the propagation delay with the light speed. By resolving these equations we get the users position and the exact time. If the GPS-Receiver have access to a precise clock, then three satellites where sufficient to calculate the position. This approach is called trilateration (see figure 1).

GPS offers two services: Precise Positioning Service (PPS) and Standard Positioning Service (SPS). PPS is encrypted and can only used by the US-/NATO-military. Therefore each

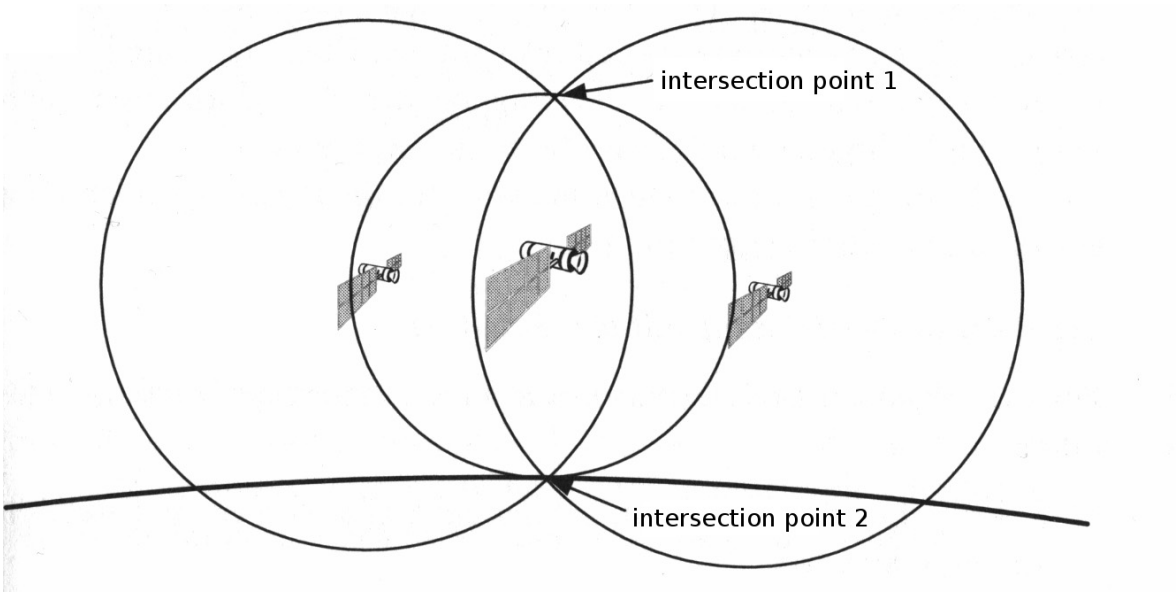


Figure 1: trilateration, source: [Roth02]

spacecrafts sends on two frequencies, the base signal on L1 (1575.42MHz) and a correction signal on L2 (1227.60MHz). By comparing the phase difference between the L1 and L2 signals, the error caused by the ionosphere can be calculated and minimized. At the moment, only the L1 frequency is usable for civil use. The military accuracy is unknown and the civil accuracy is about 10m.

2.2 EU Galileo

Galileo (see [DoHä04]) will be a satellite system similar to GPS developed by the European Union. The goal is to create a positioning system compatible with GPS but independent from the US-military interrests. It is planned to send unencrypted on 4 frequencies with encrypted (commercial) add-ons to reduce errors. With two-frequency receivers a horizontal accuracy about 4m and a vertical accuracy about 8m should be reached. The time accuracy should be about 10ns relative to UTC.

2.3 GPS as assistance for vision based Lane Recognition

Is it possible to reach the same accuracy like a vision based system? A standard mobile GPS receiver for civil usage only reaches an accuracy about 10m. This is good enough for navigation systems, but not acceptable for lane departure warnings, not even for safety related warnings in ad-hoc networks. In addition to normal GPS there exist solutions to improve the accuracy. These will be described in the following sections.

2.3.1 Kalman-Filter

When more then four spacecrafts are available it is possible to improve the accuracy with statistic methods. One of the most importand method's is the Kalman-Filter developed by Rudolf E. Kalman 1960. The Kalman-Filter is a statistical filter who estimates the states of a physical system (e.g. position, velocity) from noisy observations over time and minimizes the mean-square estimation error. The next State X_{t+1} is a linear function of the current state X_t

plus some gaussian noise. In addition to most other filters (e.g. low-pass-filters) the Kalman-Filter is a purely time-domain filter. There is no transformation in the frequency-domain and back needed.

The standard Kalman-Filter is designed for linear models. For non-linear models the extended-Kalman-Filter must be used. The extended-Kalman-Filter essentially linearizes the non-linear function around the current estimate.

2.3.2 DGPS

Most errors occur due to influences in the ionosphere and troposphere. To eliminate these errors we can use a second receiver on a fixed base-station with known coordinates. The base station should be located as near as possible to the mobile GPS user, so that the same deviation occur. The base-station also calculates it's position based on the GPS-Data. Then it calculates the difference between the known exact position and the GPS-position. This difference is sent to the mobile user so that he can include the difference in his calculation and improve the accuracy. With DGPS a accuracy about 5cm in real-time is still possible and with post-processing down to mm-range. In Germany, DGPS correction data is available nationwide from SAPOS ([SAP]).

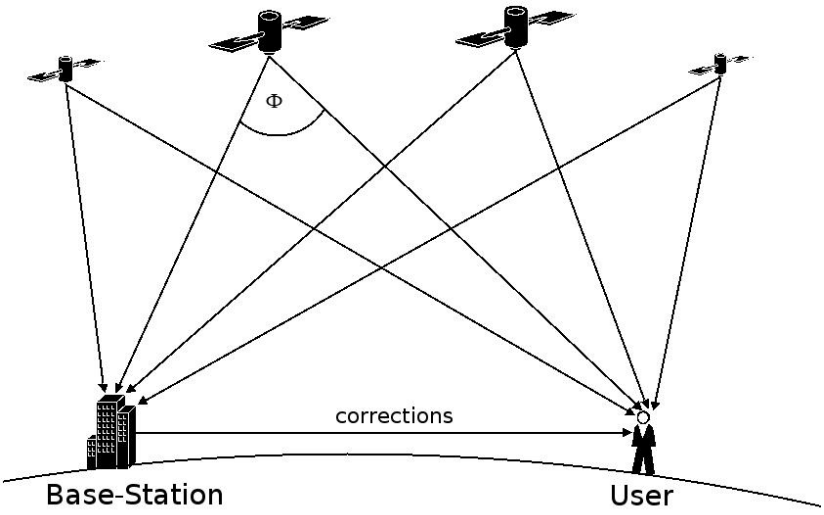


Figure 2: DGPS

2.4 Inertial Measurement Units (IMU)

To bridge short GPS-outage gaps and to improve the accuracy, inertial measurement units are used. A IMU consists of gyroscopes and accelerometers for all three axis (x,y,z) as shown in figure 3. A Gyroscope detects angular movement. Accelerometers detect changes in velocity. Going out from a known position, it is possible to calculate the current position even if we can't receive correct GPS data. Especially in cities with high buildings this is very important, because then it is often possible that less than 4 satellites could be received. To obtain a high precision an accuracy less or equal than 1deg/hr for the gyroscope and 1m or less for the accelerometers are needed. Currently used IMU's for navigation systems only offer an accuracy about 1 deg/sec for the gyroscope and 200mg for the accelerometers.

Further requirements are a small coverage type (0,5"x0,5") so that it could be placed near the GPS-antenna and it should be cheap. These requirements could be nearly met by MEMS IMU's (see [Cone00]) which are built using microscopic electromechanical structures. So, cost and volume availability can be reached in 2008.

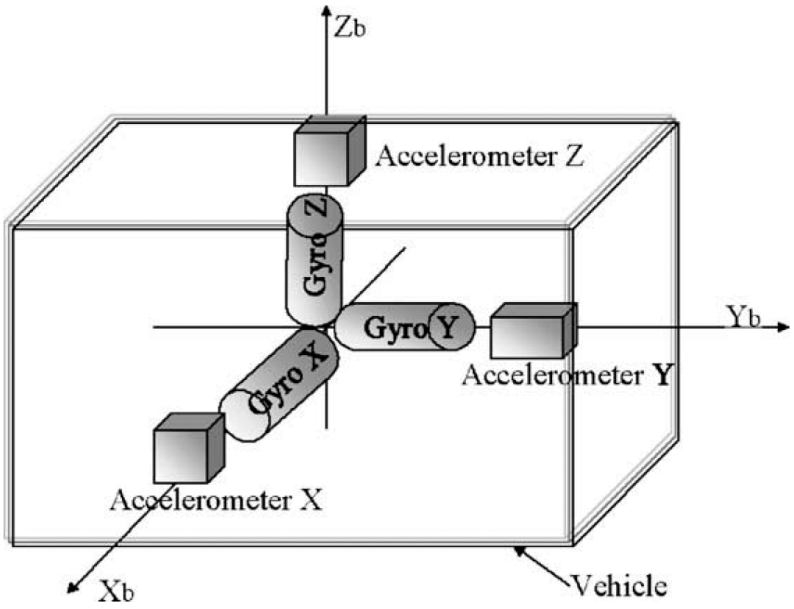


Figure 3: IMU for tree axis

2.5 Maps/Map Matching

To calculate the distance to the lane borders it is necessary to match the calculated position with a map. For safety applications like lane departure warning an accuracy about 10cm is also necessary. Current maps provide only 5m for metropolitan and 50m for less populated areas. With such a low accuracy map matching is very difficult and error-prone, especially when there are more possible roads nearby. With precise maps and precise positioning systems map matching will be easier. A technology to make precise maps exists today, but it is very expensive for nationwide-deployment. An alternative could be to use statistic methods and standard positioning systems with standard accuracy (about 1,5m) in private and commercial vehicles. These vehicles collect data during theire normal drives. Based on the data from multiple drivers the lane network refines dynamically. This method is cost effective, efficient and scalable. A precision about 10cm has been demonstrated (see [WSMO$^+$03]).

2.6 System Integrity

An important requirement is system integrity. We have to determine wether the system has met the expected performance. To do this, there are several crosschecks necessary. So first the DGPS-data is crosschecked with the INS (Inertial Navigation System) data. Then the integrated GPS/INS data is crosschecked with a dynamic model of the vehicle.

To simplify this model, they assume that there are no sideslips at the rear tires. In curves, the corresponding noise level will be increased. As parameters mass, speed and angular rate, are used to describe the vehicle. With such a model, it is possible to define a Kalman-Filter where the integrated INS/GPS output serves as observations. As output the Kalman-Filter

estimates the error of the INS/GPS data which is used for integrity checks. The integrity monitor is shown in figure 4 and a test-run with the corresponding integrity level is shown in figure 5.

Figure 4: Integrity Monitor

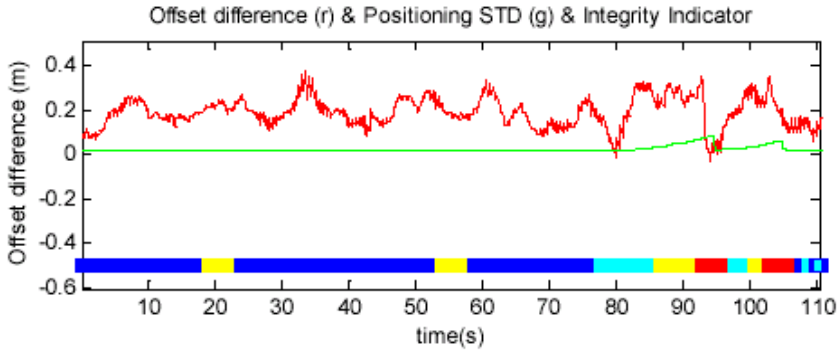


Figure 5: Positioning offset (TestA)

In figure 5 there is an jump in the There are about 30 cm position jumps at the time of 93 s due to system calibration. However the jump is captured by the integrity indicator. There are a couple of other integrity warnings which can be considered as false alarm because of the tight threshold setting."

3 Introduction to Visual based Lane Recognition

The goal of a VBLR-system is to determine the relative position of the vehicle with respect to the lane. This is done by detecting the lane features like lane markings and boundaries. To improve the image processing an extended Kalman-Filter estimates the lane borders and allows creating a search window. To define the Kalman-Filter a road model is required.

3.1 Clothoid road model

To define such a model, we take into account that most highways are built under the constraint of slowly changing curvatures with clothoids. Clothoids are used to 'connect' straights with circular arcs. Without clothoids, there would be a curvatures jump between the straight and the circular arc which cannot compensated by the driver (see figure 6).

The clothoid road model can be approximated by a third order polynomial. P(X,Y) stand for a lane border point (X and Y axis are shown in figure 6).

$$X = \pm\frac{W}{2} + X_{off} + \Theta Y + C_0\frac{Y^2}{2} + C_1\frac{Y^3}{6}$$

X_{off}: lateral offset of the vehicle origin with respect to the lane centerline, Θ: vehicle's yaw angle relative to the lane axis (small), C_0: curvature, C_1: curvature rate

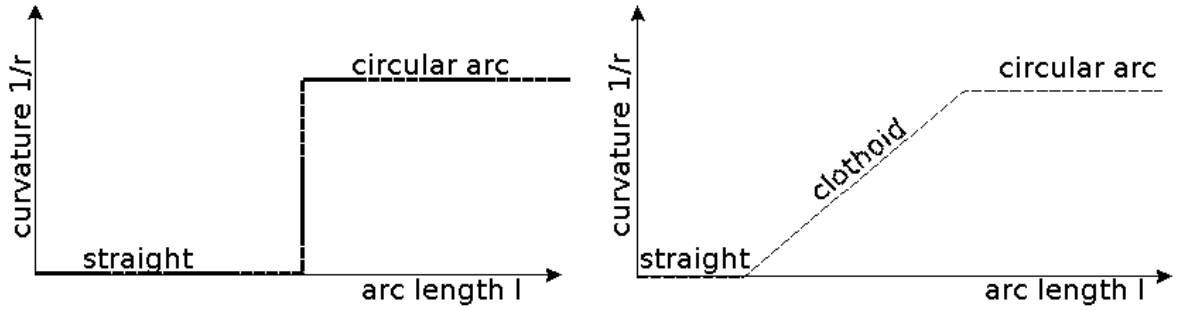


Figure 6: Clothoid road model

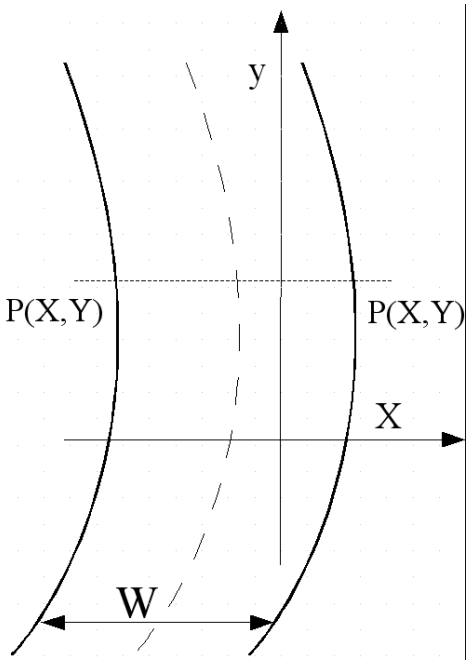


Figure 7: Road parameters

Based on the clothoid model and the Kalman-Filter search lines are placed in the image. The lines are centered on the prediction of the Kalman-Filter. The line length is based on the variance calculated by the Kalman-Filter. Then a dark-light-dark (DLD) transition is used to detect the white or yellow lane mark on the gray lane surface (example image: 8).

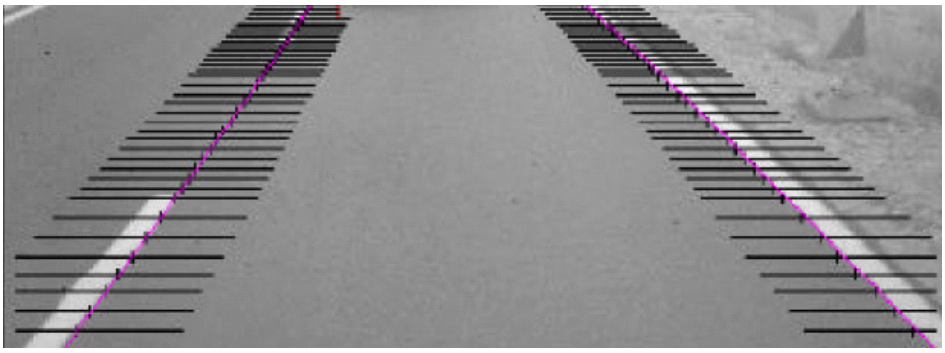


Figure 8: Search lines, source [rnJSMö04]

4 Wrap-Up

4.1 System Overview

The complete system consists of the VBLR and the positioning system (see figure 9). The VBLR uses the camera-data as input and detect the lane boundaries with help of the road model. With this information it calculates the relative position of the vehicle.

The positioning system uses DGPS- and INS-data as input. With this data and the help of the vehicle model it calculates the absolute position. With the absolute position and precise maps it is possible to do the map matching and find the current position with regard to the map and so the relative position with respect to the lane can be calculated. This system has been implemented and tested on the A8 near Stuttgart. These tests showed that the difference between the vision based system and the positioning system is about 15cm.

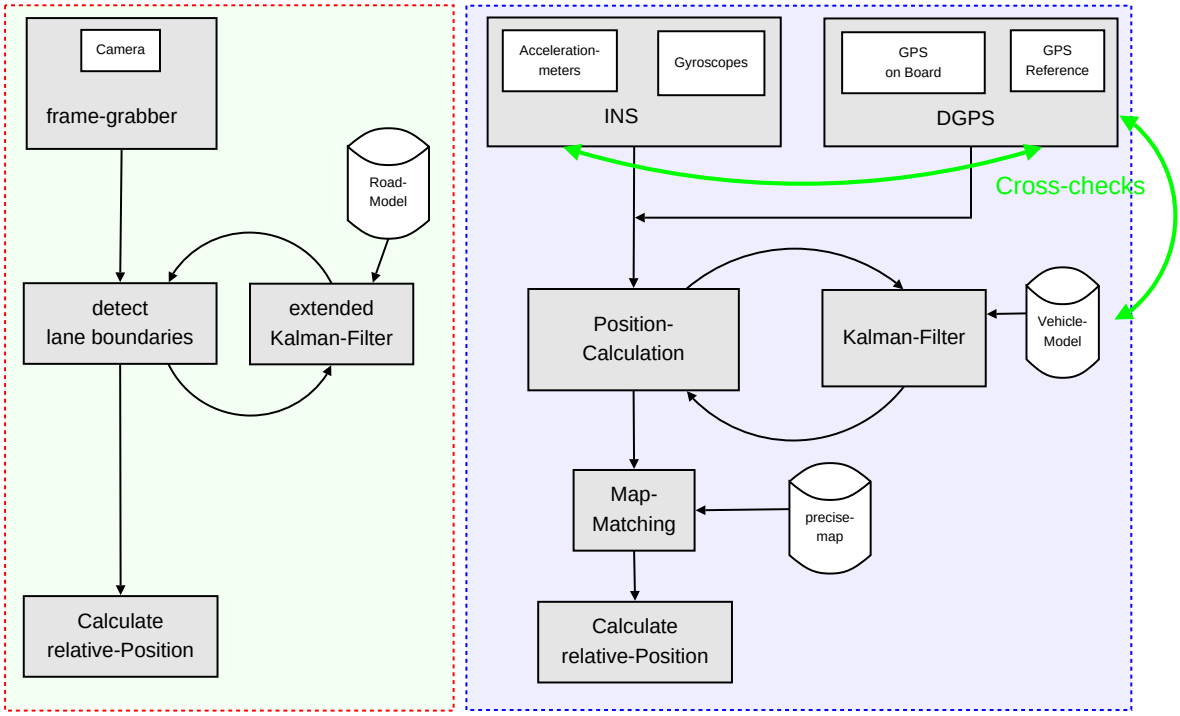


Figure 9: System Overview

4.2 Conclusion

The tests showed that centimeter vehicle positioning can used to assist vision systems for lane-keeping applications and could help to achieve robustness and smoothness. Before such a system is usable, precise maps must be created. An ultimate deployment can be expected within the next decade.

4.3 Discussion

The paper gives an good overview over the current available positioning technique. Nevertheless it's hard to understand if you are not familiar with the topic especially with statistic filters. Also the road model could be explained a bit more detailed. Beside this, it's an good starting point to this topic.

The positioning system is good enough to determine the current lane which allow applications like crash warnings in ad-hoc networks. A precision of about 15cm would be sufficient for the demands in ad-hoc networks. But there where no tests on smaller roads and not enough information to evaluate the overall accuracy. If we assume that vision based systems are very precise, then an offset of 15cm is still good enough. So the main unsolved problem is the bad map quality, which must be improved before such a system could be used. Although it is impressing how precise the statistical maps are, I think they cannot be expected in the next few years for a whole country like germany. Due to security concerns it is very important, that the data for the maps is trustful. Beside this, it also must ensured that the tracked data is stored anonymous due to data privacy. And also when such maps are available and created with trustful data it must be assured that the received GPS- and DPGS-data can be trusted. Perhaps it is possible to include a signature in the GPS data-stream.

References

[Cone00] J. Conelly. Micromechanical Sensors in Tactical GN&C Applications. AIAA-2000-4381, 2000.

[DoHä04] Hans Dodel and Dieter Häupler. *Satellitennavigation.* Hüthig. 2004.

[rnJSMö04] Prof. Dr. rer. nat. Jörg Schütte and Dipl.-Ing. Nikolaus Möhler. Lane dedection for Road Vehicles. http://www.ivi.fhg.de/frames/english/projects/produktbl/fahrspurerkennung_strasse_en.pdf, 2004.

[Roth02] Jörg Roth. *Mobile Computing.* dpunkt Verlag. 2002.

[RuNo03] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach (Second Edition).* Prentice Hall. 2003.

[SAP] SAPOS - Satellitenpositionierungsdienst der deutschen Landesvermessung. http://www.sapos.de.

[VlPH01] L Vlacic, M Parent and F Harashima. *Intelligent Vehicle Technologies.* Butterworth Heinemann. 2001.

[WeBi04] Greg Welch and Gary Bishop. An Introduction to the Kalman Filter. http://www.cs.unc.edu/ welch/media/pdf/kalman_intro.pdf, 2004.

[WSMO+03] IEEE (Hrsg.). Centimeter Vehicle Positioning and Lane Keeping. preceedings, Intelligent Transportation Systems, 2003, 10 2003.

List of Figures

Review of "Detecting and Correcting Malicious Data in VANETs"

Gerrit Göbel

Abstract

A way to exchange data between cars is building a network, especially a mobile ad hoc network. This may lead to problems, which are different to those of a cellular network. To meet required performance, a vehicular ad hoc network (VANET) has to rely heavily on node-to-node communication. This, however, makes it easy for attackers to broadcast malicious traffic in order to creep into it. Hence the approaches try to ensure the communication based on the fact that malicious nodes may be present. Thus, the nodes have to verify incoming information. For this, a model based on physical and statistical properties is given, with whose the nodes can search for possible explanations of the received data. They accept the most likely ones. The method is based on several assumptions, such as nodes are able to distinguish nodes from other nodes and a parsimony argument which will be described later. Furthermore, there are several ways to trouble the network. Attackers can spoof the other participants by pretending the existence of others nodes (spoof nodes), which will confirm the malicious information to be correct. On the other hand, attackers can intercept messages and forward information with new content. To conclude, several examples are given for a better comprehension of the system.

Keywords

Data validation, location verification, privacy, model of the Vanet, attackers

1 Introduction

1.1 Problems of ad hoc networks unlike in cellular networks

An ad hoc network is a number of wireless mobile nodes, which may build a temporary network without infrastructure and without central administration. If the nodes have to be mobile, e.g. in cars, this will be called a vehicular ad hoc network. In addition, there may be physical node collusions and insider attacks which would not exist in non mobile ad hoc networks.[Chen04]. In cellular networks, beside the assured transfer of the data a further approach is the security of applications like e-commerce, route planning etc. The validation of these data will be reached by key management solutions and authentication of network "insiders". In the VANET, however, this would be the wrong approach, because there are much easier attacks to the system than modifying data, such as producing wrong announcements about traffic jams, dangers etc. or intercepting messages and forwarding new, or selected information. Especially the VANET has to provide security against strong, colluding adversaries which may well be trusted members of the network.

1.2 Conditions for the model

The verification of data exchanges in a VANET is supported by sensor driven technologies. Sensors offer the possibility to detect the direction from where the message arrives, and so

the source may be found. Furthermore, every participating node has to validate the data it receives (of course, the node always trusts the observations which it made itself). The test is based on physical, safety conscious and statistical conditions (e.g. two nodes can never occupy the same place at the same time, information from nodes which travel faster than at a predefined velocity will be ignored, ...). These conditions lead to a model of the VANET, which can be constructed offline and so, to a large extent, when manufacturing the vehicles. Every participating node has to maintain this model and all incoming information will be compared with it.

1.3 The parsimony argument

If the incoming information is not consistent with the model, the node uses a principle which is called parsimony argument. This means that the probability that many honest nodes are present besides a few malicious ones is more likely than the other way around. So the node will always believe in a small number of attackers and thinks the explanation with the fewest malicious nodes to be the best.

1.4 The Sybil attack

An attacker may reproduce itself several times and spoof a node with the assumption of having got many messages from several different nodes. So the malicious node can bolster its view of VANET and outwit the model. To arrange this situation, a malicious node can send its messages multiple times. To counteract the Sybil attack, it is important for the participating nodes to differentiate nodes from each other. Here, the sensor capability comes into play which can differentiate the directions the data came from. The sensors combinated with the parsimony argument and the model of the VANET make the defense against the Sybil attack effective, because inconsistencies may occur and the malicious node can be detected.[MuMa04]

1.5 Network density

In every network, distributed algorithms work better if the density increases. This supports especially the connectivity of ad hoc networks because of some special properties like asymmetric connections. The communication pathways in ad hoc networks are not naturally bidirectional. The higher the density is the more likely it is that other pathways are used. Thus, it is reasonable to establish redundancy between the nodes. The mobility combined with a high density makes it difficult for the colluding malicious nodes to maintain a configuration where several attackers try to spoof a number of honest nodes as shown in the figure 1.

2 Solutions from other scientists

There exist other approaches to defend attacks like the Sybil attack, but in general they are not strong enough to manage all possible situations which may appear in a configuration like VANET. [Phil04]

2.1 Redundancy check

The redundancy check tests if information came several times from the same source. This works if every node can be definitely identified. But the Sybil attack for example can send messages with distinct identities, so the redundancy check will fail.

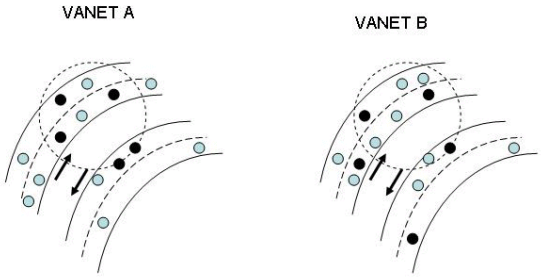


Figure 1: VANET A: The black circles represent the malicious nodes, which surround some honest nodes. In this configuration, an succesful attack would be likely because the majority of the attacker. They can interact and so convince the honest nodes of false information. VANET B: It is very expensive to maintain the configuration in A, the vehicles drove further and the number of colluding malicious nodes in rate to the honest nodes decreased. Hence, the probability of a succesful attack is reduced

2.2 Resource testing

The resource test reviews which resources such as storage, computational or communication ability a node has. If the collectivity of a certain resource of various senders has the capacity of one node, the test proves that an attack may have appeared. This check fails if the attacker can increase his own resources, which is not quite difficult to arrange.

2.2.1 Radio resource testing

The radio resource test is based on the assumption that every node broadcasts its information over a certain frequency. If a message comes several times with the same frequency, the source must be the same, the information will be accepted only once. If the attacker has the possibility to send over more than one frequency, this check is not sufficient.

2.3 Registration

A quite good approach would be to register every participant of the VANET. If every node may be identified definitely (by his license plate or other criteria), a malicious node will have no chance to attack a VANET in a simple way. On one hand this would solve the complexity of some problems with attacks, on the other hand the complexity to manage the registration all the time and to update it with every new participant and delete the leaving ones would exceed the benefit by far and make such an ad hoc network absolutely impracticable.

2.4 Position verification

Position verification allows the network to verify the position of each node. The data that come from the same location are assumed to belong to one and the same participant. But the nodes are static and do not develop this approach.

3 Kinds of adversaries in a Vanet

In this chapter the different types of attacks will be described, as well as the possibility for the attackers to exploit the description of the methods VANET uses, in order to prepare for their attacks. An attacker in VANET is successful, if he accomplishes to convince other nodes of his malicious data. In other approaches, already spoofed nodes as well as Sybil nodes may be used to control the malicious information. In this approach the validation is not based on the accumulation of agreed data, but the model is based on probabilities of attack scenarios, which may appear in a VANET. As a result of that it is more reasonable to distinguish attacks based on their type, their possible targets, the dimension the attack can get and the possible outcomes. They will be described in the following articles.

3.1 Different types of attacks

There are many ways to influence the consistency of the VANET data. A malicious node could tell other nodes about traffic jams or dangers which do not exist or does not say the truth about its real position. The detection of such nodes depends among other things on the capacity of the sensors, the evaluation of the exact direction of another node, however, by measuring the distance is difficult. Furthermore the exchange of new messages, which are used to validate new data depends to the actual density of the network and other environmental influences. The mobility inherently constitutes a problem of the accurate measurement. So it may appear that an attacker keeps undetected for some time. This case has to be caught by the model.

3.2 Possible targets of attacks

An attacker could be very strong and collude with other malicious nodes over high quality out-of-band channel for example. The attacks which are locally arranged are more likely, because the probability to interfere with other proximate honest nodes is lower. But it is more difficult to maintain the configuration of a situation because of the mobility of the nodes. So the malicious node may have several distant targets but the data of the neighboring nodes could make the attack ineffective.

3.3 Dimension of an attack

The scope of an attack depends on the number of nodes which are influenced by the malicious data around a malicious node. If there are not many nodes in this area, the dimension of the attack cannot be very large, but if false information circulate beyond that local area it is called an extended attack. The approach of VANET can avert a local attack which can lead to an extended.

3.4 Possible outcome of an attack

If an attack has happened, three results are possible. An attack may be undetected if a node was isolated or completely surrounded by malicious nodes. In this case, the attacked node will accept the malicious information. An attack may be detected by various nodes, but leaves an uncertainty about the received data. The third possibility is the detection of the attack, consequently no node can be influenced by malicious data. If there are honest nodes between the malicious ones surrounding an honest node, it may have a chance by validating the data with the parsimony argument. If enough honest nodes are present, the attack will fail.

3.5 Developing a Model of adversaries

On the basis of a model a node can make an efficient evaluation of arriving messages, being aware of the possibility of an attack. Due to this a model of adversaries is given with the challenge to calculate the most likely attack. Considering a statistical basis or combinatorial instruments, an ordering relation will be developed, where the explanation with the fewest malicious nodes is on the top and which will be actualized all the time. The fundamental question is, whether the attacker can use the knowledge about the model of adversaries, for his own advantage: can he change his way of attack to trick the model or can he send other messages which are more likely in the model? To avoid such a situation, the model has to be strong enough to make the adaptation for the attacker to the new situation more expensive than a normal attack. In addition, the model not only has to actualize the probabilities but also to remark and to react to changes in the pattern of attack of the adversaries. Hence, an adversary has to be observed for some time to include the usual "arms race". But in spite of all the complexity of treatment of adversaries and their evaluation the possibility of a mundane attack should not be forgotten.

4 Distinguishability

A very important condition in order to notify an attack is to be able to distinguish between nodes. A Sybil attack is not possible if the node can differentiate the malicious node and the spoof nodes. There are several assumptions which should be made. Every node observes its environment, and if a message comes in, the node has to combine the observations with the message. So it can check if the incoming message really came from the source. Furthermore, the node has to be able to differentiate his direct neighbors. As a result of the assumptions, there were two different cases to examine. On one hand the ability to localy distinguish what is around a node, and on the other hand the extended variant.

4.1 Local distinguishability

As mentioned above, the node must realize which other node sent the data. This problem can be solved by different solutions. For example transmission of light can be used (infrared or in the visible light spectrum) if the vehicles are equipped with cameras, which realize the exact direction where the signals came from. So the node can define the source of the message, which is a very important condition. It allows to distinguish the next neighbors, otherwise the node is vulnerable against attacks like the Sybil attack, or the consistency of his information may not be given any longer, if a confusion occurs for example. Other physical signals may be used to calculate the distance to the object, like ultrasonic rays or radar. But this could lead to problems, because the nodes may increase or decrease the strength of their signal.

4.2 Extended distinguishability

The node not only has to know what is in his direct neigborhood, it also has to know what happen beyond its neighbors. So the nodes exchange information between each other. Every node observes its environment and sends these data to the other. Normally, data exchange over larger distances may be limited by bandwidth and latency. But in this approach optimal conditions are assumed.

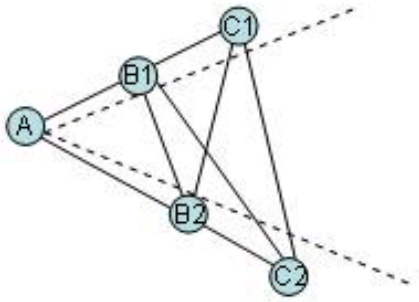


Figure 2: Node A recognizes his direct neighbors B1 and B2, but he is not able to reach C. A builds up the communication to B1 and B2 and B itself measures the position of C. Because of the data exchange between A, B and C all nodes know the position of each other.

4.2.1 Network density

As usual in ad hoc networks the pathways between nodes are redundant. If every possibility to connect between nodes is used, the graph of communication is complete. Also malicious nodes may be present, but their misconduct, perhaps not to forward data, should be corrected by routing protocols. The methods which are developed for VANET work naturally better if the connectivity of the graph is quite good, the parsimony argument will be used more effectively.

4.2.2 Mutual identification

Every node should have a pair of private and public keys, which can be refreshed as often as needed. Now, if two nodes had a close enough contact, they can exchange their pair of keys and for the time of validity of the keys they can exchange data in a trusted way. The privacy is pretty good in this approach, because the pairs of keys are quite short-lived. The node decides when to change the key and where to send it again. Also the problem has to be solved, how to refresh the keys several times (which is necessary for privacy) if two nodes exchanged data for a longer period of time. An advantage is the possibility of authentication not only over the time but also over a certain distance, because the source can be identified by the key, even if he was not in the direct environment of the node (for the time of the validity of the key) regardless from where the information came. Furthermore, if a node sends information, which will be identified later as malicious, all its data will be classified as questionable. In the model strong adversaries are allowed, which can exchange key pairs after an authentication to attack a node. But if a larger group works with the same key and tries to send the information as every node came from the same point, there is the possibility that one of the group is identified as malicious, and then nobody can send further information. That means, collaborating attacks may be counterproductive.

4.3 Privacy

An important problem in VANET is privacy. In the approach of Philippe Golle, Dan Greene and Jessica Staddon [Phil04], it is realized by decentralisation, i.e. data which need a good privacy do not have to flow to a centralized device which could be spied out. The information

can be sent directly, thus an attack on some nodes would not lead to a crash of the system and the information of all nodes are not saved at only one certain device.

To track a particular node the attacker should have many nodes around the target, thus they can detect the position precisely. This would be very expensive to maintain, because every node is mobile and it may move everytime. Because the mobility it is reasonable to care about the extended distinguishability. A key pair is used to authenticate the communication partners, it is refreshed continuously. This assures that an attacker cannot identify the node as long as he wants to. A longer identification of the nodes is not needed. The nodes have to exchange the pair of keys really carefully to maintain the privacy. There are several ways to do that: The pair of keys may be changed at synchronized times, the nodes can include gaps into the messages before changing the keys and they can exchange keys just in the moment, the distance is close enough.

5 Developing the model

After having pointed out the aspects in and around VANET the following model is proposed.

Distance: The distance between two nodes P1 and P2 is defined as the Euclidian space $\parallel P1 - P2 \parallel$

Event: $E(D, f)$ is defined as an event where D is the data and f is a continuous function $f : T \rightarrow P$ which is the locator function that specifies the location of the event over the lifetime $T \in R$. The lifetime of an event can be a real short time, maybe just a point of time or it is a longer time period ($T = t$ or $T = [t_0, t_1]$). The data in an event may be, for example, the identity or speed of the node at the location given by $f(T)$.

Node: (N, f, p) where N identifies the node uniquely whereby $N \in N$, f the function defined at event and p is the radius in which the node makes his observations, $p \in R$

Assertion: Every node can observe its environment inside the given radius. If something happens like the detection of another node, it broadcasts a message to the other nodes, the content of the message is called assertion. If Node O_i observes an event, it produces a message $< (D, f) >_{O_i}$. Every node, which get messages has to validate the assertion before accepting it. The lifetime of an event has to be a part of the lifetime of the node which produce the assertion and for every single point of time of the event, it has to happen inside the radius of the node. In a technical way, it means that (n_i, f_i, p_i) is a node and the event is $E = (D, f)$. T_i is the lifetime of the node N_i and T is the lifetime of the event E. If $T \subseteq T_i$ and $\forall t \in T$, $\parallel f(t) - f_i(t) \parallel \leq p_i$ then the node can accept the assertion $< (D, f) >_{N_i}$.

So nodes may exchange assertions with one another, in the practice in contrast to the theory the exchange of data is limited by bandwidth and latency considerations. In this approach these factores are ignored. So every assertion made by a node is available for every node - in the strict sense in the local area, because an assertion is most important for the neighborhood.

5.1 Model of the VANET

The model of the VANET is developed to define the events or sets of events which are possible. This could be based on rules or on statistical properties of events.

The model of the VANET is a function which maps the set of all existing events to the values valid or invalid. Formally let ϵ be the set of all existing events $M : \epsilon \rightarrow \{valid, invalid\}$. If there are several events $\{E_1, ..., E_n\}$ and there is a subset of ϵ, they will be consistent with

the model if $M(E_1, ..., E_n) = valid$ and inconsistent otherwise. If the return value is just binary, the assertion of a set of events is not really accurate, so the model will be extended with probabilities, so the return value will be a value between 0 and 1.

So every node can validate its own observations and share them with others. The problem in this approach arises from malicious nodes, which also can record events, that did not happen. They are a risk for the consistency of the global database of events.

5.2 Meaning of a set of events

Every set of events can be identified by a number $H \in N$, and every set is partitioned into two subsets, the hypotheses of validity H^+ and the invalid or malicious ones H^-. Every event in a set of assertions $K = \{< E_1 >_{O_1}, ..., < E_n >_{On}\}$ has an explanation $Exp_N(K) = \{< E_1 >_{O_1}^{h_1}, ..., < E_n >_{O_n}^{h_n}\}$ whereby $h_i \in H$ so that the subset of assertions tagged with hypotheses of validity is consistent with the model of the VANET.

$$Exp_N^{H^+}(K) = \{< E_i >_{O_i}^{h_i} \in Exp_N(K) | h_i \in H^+\}$$

And because H^+ is the valid subset of hypotheses, the map of $(Exp_N^{H^+}(K))$ of M is valid:

$$M(Exp_N^{H^+}(K)) = valid$$

Naturally, every explanation is defined with respect to a particular node N, because every node makes his own hypotheses of events (for example every node declares his own assertions as true).

The explanations which a node gets will be ordered by the node, depending on the statistical method used. It specifies a total ordering based on some scoring of the explanations. There are several methods, for example scoring explanations based on their simplicity.

If a collection of data K is invalid under a model of VANET M and an additional ordering collection of explanations of K is given, then there are two possibilities to proceed. On one hand the data is declared as invalid because there was an error. On the other hand the errors may be corrected when the best explanations use the assertions which are labelled with H^+. If there are several best explanations, the assertion labelled with H^+ can be split and a subset of K can be corrected.

6 Examples

The methods of the VANET should be constituted by some examples. For a better comprehension, two cases will be described, the one where sensors work quite well and the other case if the data transmission is complicated by weaker sensor capabilities. The second example illustrates the importance of distinguishability.

6.1 First example

In this example it is assumed, that nodes may specify the exact location of all neighbors. Furthermore the the location of a node which is sensed can be associated with its public key, because location sensing is bound with communication.

In the database K are tuples: $K = \{< N_1, \vec{x}_1 >_{O_1}, < N_2, \vec{x}_2 >_{O_2}, < N_3, \vec{x}_3 >_{O_{3...}}\}$ Every of these assertions $A_i =< N_i, \vec{x}_i >_{o_i}$ means, that the node O_i suggest having an observed node

at the position $\vec{x}_i$. If the observed node is beyond the observing range of the node, it has the value unobserved. then may be deleted from the database. For every node some entries are already made, because they can make assertions about themselves. Hence, there is a function L who maps N_i to the position $\vec{x}_i$ if the assertion is reflexive: $< N_i, \vec{x}_i >_{N_i} \rightarrow L(N_i) = \vec{x}_i$. Furthermore it is predefined that K has reflexive entries for every node and every non reflexive assertion $< N_i, \vec{x}_i >_{O_i}$ is consistent to the reflexive assertion of N_i. That means that the position of the observed node is specified if the Euclidian distance of the position of the node $\vec{x}_i$ and the position of another node $L(O_i)$ is smaller than the radius of N_i, and unobserved otherwise. Formally it means: $\vec{x}_i = L(N_i)$ if $\parallel \vec{x}_i - L(O_i) \parallel \leq p$

Potentially there are malicious nodes present too, so the explanation has to label the assertions with designations, truthful, malicious or spoof. Any criteria has to be satisfied by the labelled tuples $< N_i, \vec{x}_i >_{O_i}^{h_i}$ whereby $h_i \in t, m, s$ (true, malicious, spoof), so naturally the observations which the node made has to be labelled as true (if $O_i = N$ then $h_i = t$).Furthermore, the observers labelled as spoof do not appear in other tuples.

In some cases, it could be helpful to add some other tuples, who get the label added ($a \in H^+$). Every reflexive tuple that has been labelled m can be replaced by this new added, which supply a correct location that is consistent with any other truthful observations.

To complete the model of adversary, the method has to count the labels "malicious" which a node got for one or several tuples from different observers. The explanation with the fewest malicious labels seems to be the simplest one. If there are enough observations in the database, the malicious node will be detected by the data in $Exp_N^*(K)$, depending on the correctness of the positions of all nodes.

The advantage of this model of adversaries is, that the malicious node has not much possibilities to attack, if one attack fails, or rather one assertion is labelled as malicious, all the other assertions will be labelled as malicious too, so there is no second chance and the probability to detect an attacker early is higher. Furthermore, the spoof labels in the explanations are ignored so the higher ranked explanations have more spoof and fewer malicious assertions.

If there are only a few malicious nodes, there will be a simple algorithm to find the best explanation. Beginning with node N a breadth first search is starting, traversing the arcs from the observers O_i to the node N_i as long as N_i is not labelled as malicious. All reached nodes will be labelled as truthful, all other nodes as malicious. The algorithm can terminate if it found explanations of the same size which pass the consistency test.

In the following example some basic rules are important as usual: The probability to detect the attack is depending on the density of the network. If some nodes do not have good information about the nodes in the neighborhood, the detection of the malicious node will fail or will be complicated. In this case, the malicious node generates additional spoof nodes to enforce his attack. The spoof nodes record their observations into the database, but the real nodes on their part do not make records of the spoof nodes. That will lead to inconsistencies and will uncover the attack. The methods for detection could find the two explanations shown in figure 4. The malicious node tries to support an illusory node by creating spoof nodes. In this case the explanations with fewer malicious nodes seems to be more likely, because one malicious node beside several honest nodes may occur more often than in the other way around. The parsimony argument supports the right decision by assuming the smallest number of malicious nodes.

6.2 Second example

In the previous example, the sensors of the nodes provided good measurements, even beyond the direct neighbors. This time the possibility is considered that the sensors can just recognize

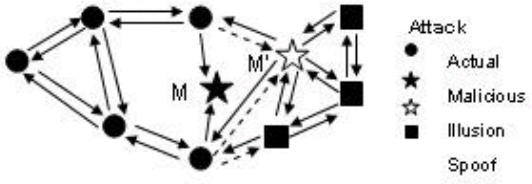


Figure 3: The malicious node M produces new spoof nodes to support a false location M'. The arrows constitute the observations which were made by the nodes, the dashed arrows show missing observations.

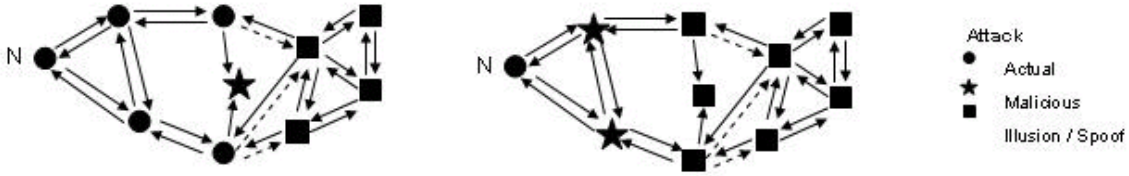


Figure 4: Two possible explanations for the conflicting observations, many others are possible too.

nodes in the environment without the certain direction. To compensate this matter of fact is to add the region R into the database $(N_i, R_i)_{O_i}$ whereby the observer O_i assumes, that N_i is positioned midst of the circle R_i. The exact size of the circle depends on the sensor technology. This approach leads to the detection of malicious nodes even if their data are not recorded into the database. Hence, it is possible that malicious nodes are totally unrecognized and not to find in the data, as well as new data can be added by the test methods, letter labelled with "a" (added). A record could be $< N_i, S_i >^a_{N_i}$. S_i is a region where N_i is possibly positioned, the dimension depends again on the sensors. Also in this example, the explanation with the fewest malicious nodes is prefered. A further attribute is established, "i" for illusion. In this method a malicious node can produce virtual nodes, which appear for the time of the presence of the malicious node as real.

$$< N_i, R >^i_{O_i} \rightarrow \exists k :< N_k, S_k >^a_{N_k} \wedge (Sk \subseteq R_i)$$

This record averts that an unreal node appears in the database as a real node.

Now, there is a systematical tryal to find good explanations by assuming malicious nodes by degrees and record them with the attribute added into the database. At the node N, the breadth first search is started and marks every assertion as illusion if the observer is midst the surrounding of a malicious node N_k and if its area is a subset of $R_i(L(N) = S_k$ whereby S_k subset of $R_i)$. If the assertion is reachable, it is labelled with truthful, otherwise as malicious. After this, the explanation with the fewest malicious nodes is used to correct the data.

The parsimony argument is not really reasonable to use in this case, a node in the surrounding of a malicious node could be influenced totally, its own world may be a complete illusion produced by the malicious node. Hence, the approach is modified. As previous, the explanation with the fewest malicious nodes is chosen with respect to the most honest nodes. An

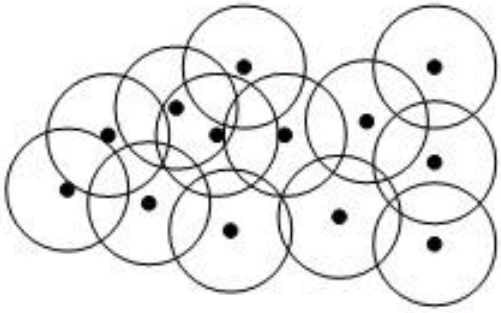


Figure 5: A possible configuration in a plane based on fixed range tests, every node is assumed to be midst a circle

explanation with only one malicious but also just one honest node may be on the top of the ordering. But by this modification there is not every problem solved. The attack would be really expensive, because every honest node had to have a malicious node in direct neighborhood, which simulates an illusory world. Especially in a mobile network, it would be very complex if not just impossible.

A further approach is based on probabilities where above-average much or least nodes in an area are more unlikely and this influences the ordering. Generally, to find the right explanation is not simple at all. It depends on the topology, the surrounding, the accurateness of the sensor data and the actual position of the nodes, because a unusual spreading of the nodes in the area is not impossible. There is always the potential for a malicious node to change the position a little and thereby to stay undetected.

7 Conclusion

The approach of Philippe Golle, Dan Greene and Jessica Staddon [Phil04] covers all sources of error, which may appear to trouble the VANET. Every node itself collects data from its environment and sends it to other nodes. By the use of the sensor technology the source of the data can be detected. Every node can evaluate incoming messages. For this a model is given which is based on physical and stochastic properties. If inconsistent data appear, the most likely explanation will be found on the basis of a parsimony algorithm. The most important assumption, however, for this is the satisfactory density of the network because the methods just work effective if enough nodes take part and exchange information.

7.1 Conclusion in own words

The approach of the three scientists is based on some theoretical assumptions, which are not practicable in the reality. The bandwidth as well as the latency are only in the model assumed as optimal. In the reality even the efficient, fast and exact wireless data transmission is not categorical possible. Those depend on the properties of the environment as well as the technology used which also may be restricted under a financial aspect (the technology has to be affordable for the rank and file). This leads to a main problem, the scale, because the system will not provide reasonable values if the density is not high enough and the attractiveness of the system is restricted strongly. Furthermore, if an attacker has the possibility to apply transmitter blocking signals, an area may be paralyzed, depending on the transmission

technology used. These signals can interfere the broadcast and the message do not arrive its receiver, it can be fragmentary or faulty.

The theory is sophisticated and the model covers possible scenarios in a VANET, but the application in practice is just visible in real trials. VANET is based on frequent exchange of information, the higher the density of information is, the higher is the probability to detect attackers. If only few nodes are participating the succes may stay away. Hence, there have to be many nodes directly in the beginning of the initiation. The frictionless start of a wide system can be very difficult, problems may occur by installing it and especially in the first time. This was visible for example at TollCollect. The entering was postponed lasting for month and new problems occur often in the media.

7.2 Ideas of one's own

To support the model of the VANET, certain vehicles could be declared as generally trustworthy, for example cars which are on the road all the time. These nodes, prepared with a special key, always kept under surveillance and controlled may forward the validation of certain messages, reduce processor load as uncertainty and in a fair quantity release the system as well as make it more safe. A vehicle, surrounded by several attackers, would be certainly spoofed, even if there are few honest nodes around. The existence of base nodes, always en route, collecting data by observations and by the way loaded with information could restore the attacked node to consistency, because the information which come from it were generally trustworthy. This one could broadcast information about the attacker detected with the data of a base node and so increase the exclusion of malicious nodes. Instead of using vehicles for the base, already installed systems like TollCollect may be upgraded or additional used. A base node could send all its information to such a base station (distributed devices installed once in a while on the road, e.g. at the points where the sensors of TollCollect are placed) and every node who comes the way could get new trustworthy data.

7.3 Own opinion of the paper

The technical report of the three scientists is really exhausting to read and to apprehend. The structure of the paper is from my point of view a little bit confusing, they explain for the comprehension important things secondary (e.g. how does the Sybil attack really work, which possibilities exist to arrange such an attack) and other sections like the examples are so complicated that the real convenience is not directly clear. Some points are just mentioned marginally (e.g. position verification) so that the real understanding just arise if additional papers are read. Without having previous knowlede the paper is very hard to read and it may be better structured to support the understanding. What is missing in the paper are substantial concrete examples, there is mainly the theory described.

References

[Chen04] Qi Chen (Hrsg.). Informationssysteme in Ad-hoc Netzwerken. Termpaper, University of Karlsruhe, 2004.

[MuMa04] C. Siva Ram Murthy and B. S. Manoj. *Ad Hoc Wireless Networks - Architectures and Protocols*. Prentice Hall. 2004.

[Phil04] Jessica Staddon Philipp Golle, Dan Greene (Hrsg.). Detecting and Correcting Malicious Data in VANETs. Technischer Bericht, VANET 04, Oktober 2004.

List of Figures

Review of "The Security and Privacy of Smart Vehicles"

Torben Brumm

Abstract

Vehicular ad-hoc networks seem to be evolving quite fast and become a part of driving aid in the future. But privacy and security of these networks are only rarely examined by a few papers. "The Security and Privacy of Smart Vehicles" by Jean-Pierre Hubaux, Srjdan Čapkun, and Jun Luo [HuvL04], which deals with *electronic license plates* (a kind of identification in these networks based on changing pseudonyms) and location verification (where a car is situated at a specific moment) is one of them and will be summarized and commented in this report. After that, I will present my proposal for a protocol for these *electronic license plates* which handles authentication and changing pseudonyms and uses cars and base stations established by a central authority (which also needs to be involved in the protocol).

1 Introduction

Car technology has developed significantly during the last years and will develop even further in the next several years. Some developments increase our safety (e.g. airbags), some our comfort (e.g. navigation systems), and some of them are for controlling purposes (e.g. license plates (which are not a real "technology" nowadays but they will be one as soon as they become electronic)).

Especially communication devices (e.g. GPS receivers) become more and more important for our behavior during driving. They are (or will be) also used by the officials for collecting tolls or enforcing the law. It is just a matter of time when all our cars are equipped with one or more communication devices which will identify them towards each other (and towards stations near the road) or exchange even more complicated messages like traffic reports. The possibilities are nearly unlimited. Communication technology can improve the safety in our streets, ease the work of the police, and make driving less stressful than it is nowadays. We will look into a brighter future...or won't we? More communication means enforcing the risk of communicating wrong data or obtaining information out of the given communicated messages. While the technology is improving constantly, the matters of security and privacy have been examined by very few people. But we need do ask ourselves: "Do these new technologies also raise new problems concerning security and privacy or are they even more secure than the old methods?"

One paper which discusses the aspects of security and privacy is "The Security and Privacy of Smart Vehicles" by Jean-Pierre Hubaux, Srjdan Čapkun, and Jun Luo [HuvL04]. I will summarize the results of their work in section 3 after explaining some basic expressions used in this report in section 2. In section 4 I will take a closer look at their results and give my opinion on them. After that, I will give you a more concrete solution of their *electronic license plates* in section 5. Section 6 concludes this report.

2 Security Basics

In this section, I would like to describe some expressions I use in this report.

Tamper-resistant hardware basically is hardware put into a safe so that no one can manipulate the hardware itself. It doesn't have to be a real safe but a physical protection against misuse. Any attempt to destroy the protection and getting access to the hardware (e.g. to get a key out of it) should result in destroying the data and/or alarming someone (the owner or the authorities for example).

A *one-way hash function* is a function that uses one value to calculate another one out of it (very fast as it is a hash function). But it is impossible to determine the original value out of the resulting value.

A *challenge* is a value (e.g. a set of bits) generated by a participant in a network. It sends this *challenge* (maybe after manipulating it somehow, e.g. by using a hash function) to someone else in its network and waits for a very specific answer. Different *challenges* require different answers, so the communication partner has to calculate the answer after receiving the *challenge*.

A *symmetric key* is a key two communication partners share in a network. They both use this key for both encrypting and decrypting the messages they send to each other.

3 Paper summarization

The paper [HuvL04] gives a short overview about problems concerning security and privacy of *smart vehicles*. The authors do not invent anything new, they just summarize ideas and take a closer look at them concentrating on the aspects of security and privacy.

They often use the term *smart vehicle*. A *smart vehicle* is equipped with an event data recorder (EDR), a GPS receiver, some sensors, a communication device, and a computing device.

The EDR is similar to to a plane's *black box* which records everything concerning the vehicle to reconstruct a possible crash. The GPS receiver is installed to know the vehicle's position all the time while the sensors detect obstacles nearby the car during driving or parking. For communication with other cars or stations along the road we need a communication device which is capable of building safe mobile ad hoc networks. The computing device is responsible for supervising all necessary operations like protocol execution during communication.

With this *smart vehicles* driving would become safer (e.g. by the sensors warning the driver) and more comfortable (e.g. *smart vehicles* would be able to get an own personal route which avoids traffic jams using the communication device and its positioning system).

The two main topics discussed in the paper are *electronic license plates* and location verification (and the corresponding privacy and security problems). The following two sections will give you a overview about their results concerning these two topics.

3.1 Electronic license plates

Smart vehicles would need an identification while they are using their communication device. In the paper, this identification is called *electronic license plate* because it identifies itself towards other cars or roads just like normal license plates do towards passing cars or people. They could ease collecting tolls as it would be possible to check who entered when which toll

road. Another objective would be to identify drivers who have fled an accident by "asking" nearby parked cars what they have recorded (and who has passed them).

But there are some security and privacy issues these *electronic license plates* raise. In order to use this identification for authentication, the authorities would have to give each car several keys and certificates to ensure a secure communication between them. The structure of this system has to be very carefully designed as it requires a similar privacy level as mobile phones for example, but it also needs to be capable of exchanging messages in real-time because the car's location needs to be exact (for avoiding accidents for example) and the communication has to be fast before cars get out of range (e.g. when there are very few cars on the road). A concrete solution of this will be presented in section 5. The authors don't think that this *electronic license plates* will be a privacy problem as long as they are designed careful enough (other systems like mobile phones or the internet are very popular despite the fact that they are as private as the system of *electronic license plates*).

To become more specific: One problem is anonymity which is solved by assigning pseudonyms to the cars (which have to change over time to protect the driver's privacy). We can calculate anonymity which is done in the paper using an anonymity metric based on entropy. Their computed degree of anonymity d is

$$d = \frac{-\sum_{i=1}^{N} p_i \log_2 p_i}{\log_2 N}$$

where N is the anonymity set's size and p_i is the probability of the pseudonym being the ith value of the anonymity set. The degree of anonymity represents the amount of information the system presents (unwantedly) for a given pseudonym.

Another problem is a physical attack on the *electronic license plate* (disabling for example). But this is even a minor problem compared to the license plates we use nowadays and which we can "disable" (steal) very easily, because we are able to protect it better on the inside of the vehicle or even trigger an alarm or alter law enforcement when there is a try to disable it.

It's quite similar with stealing and using the identity of someone else. Today it is quite easy, *electronic license plates* can be made much more resistant to such attacks by storing them in *tamper-resistant hardware* and certifying them. With modern authentication protocols it will be much harder to identify yourself with someone else's identity than it is nowadays with stolen license plates. Stolen license plates look normal, but stolen identities can't be used as long as they are not able to fool the authentication protocols as well.

The last mentioned problem in the paper (which can't be solved up to now) is a denial of service attack which would jam the communicated signals.

3.2 Location verification

The second part of their paper concerns location verification and the security issues accompanying it. Two general solutions for location verification are explained:

The first one uses GPS (as mentioned above) which is quite spread even today and then sends this data to base stations or other cars. But there are some drawbacks like the availability in urban areas which block the signal (tunnels for example) or the *tamper-resistant hardware* which is not very safe until today. But the main problem with GPS is that they are vulnerable to several attacks (jamming, spoofing, etc.) which can also be enforced by relatively unsophisticated attackers. The most dangerous attack of those is using a GPS satellite simulator

which can produce stronger fake signals than the real ones. Simple attacks could be stopped by software changes, the clever ones can't.

The second solution uses verifiable multilateration. For this solution, base stations spread all over the country are needed, but we could get rid of the *tamper-resistant hardware* in the car and install cheaper normal hardware. The solution works as follows: Four base stations perform a communication with the car to obtain the distance between itself and the car. With these four values it is possible to calculate the actual position of the car which is located inside the triangular pyramid formed by the base stations. For a correct position in two dimensions only three base stations are needed (Figure 1 shows an example).

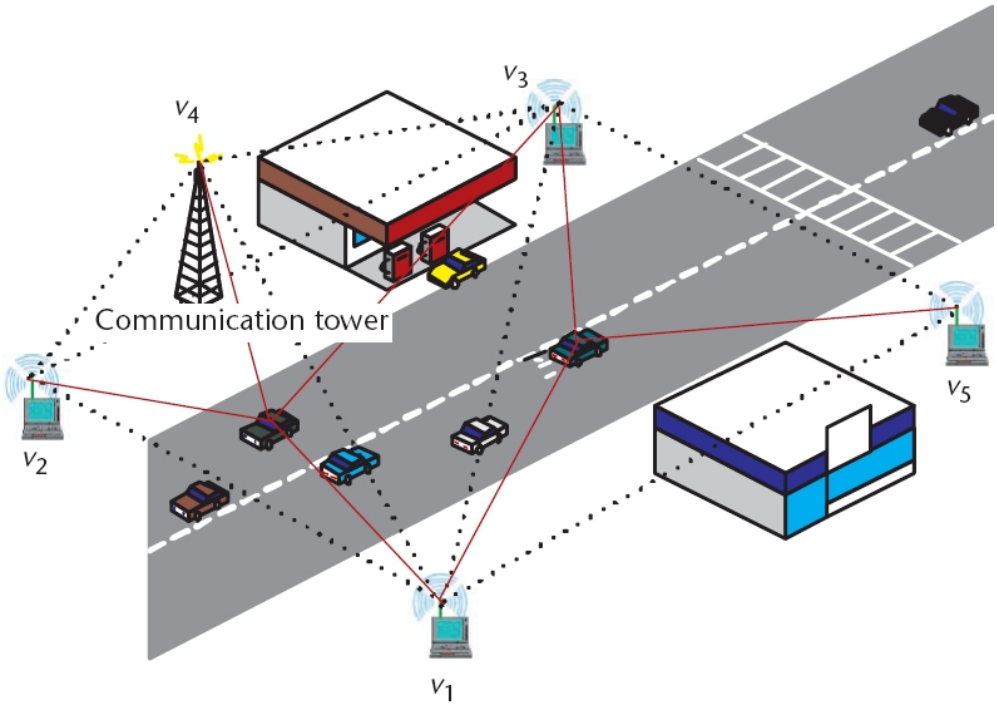


Figure 1: Base stations v_1, v_2, v_3, and v_4 determine the left car's location in three dimensions. The right car's location is determined in two dimension by v_1, v_3, and v_5. (Taken from [HuvL04])

The communication protocol performed between the base station and the car proposed in the paper (based on distance-bounding protocols by Stefan Brands and David Chaum [BrCh93]) works as follows (after establishing a connection and mutual authentication): The car generates two random values N_c and N_c' which it hashes with a *one-way hash function* h. It sends the result to the base station which answers by sending a randomly generated value N_v, the *challenge*. The car responds to this *challenge* with $N_c \oplus N_v$ and after that sends the signed N_c' to the base station. The base station measures the time between sending the *challenge* and receiving the answer and calculates the distance to the car out of this value (also considering the calculation time of $N_c \oplus N_v$ which is relatively short). To control the correctness of the messages, the base station verifies the signature of the last received message and hashes N_c and N_c' (which it knows by now) with the same hash-function h as the car. If the result matches the first received message (the result of the hash function calculated by the car), everything was correct. Table 1 presents a short version.

Following this protocol it is impossible for a car to convince a base station that it is closer to the base station than it really is (as light travels at finite speed). It is only possible to hold the answer for the *challenge* to make the base station believe that the car is further away than it really is. But as there are several base stations involved in a location verification and

Source	Destination	Message
C generates N_c and N'_c and calculates $h(N_c, N'_c)$		
C	V	$h(N_c, N'_c)$
C generates N_v		
V	C	N_v
C	V	$N_c \oplus N_v$
V measures time between sending N_v and receiving $N_c \oplus N_v$		
C	V	N_c (signed)
V verifies signature, calculates $h(N_c, N'_c)$, and compares it to received $h(N_c, N'_c)$		

Table 1: Distance-bounding protocol between a car C and a verifying base station V

getting further away from one means getting closer to another one, it is impossible to tell the system a different location than the real one, it is only possible to offer an unprecise location with wrong values which of course will be noticed.

4 Paper evaluation

After summarizing the paper, let me add some comments to their results and the ideas they mention.

The idea of *smart vehicles* is really interesting and would be a real benefit for the driver, but who could afford such a car in the near future? Today, these technologies are only available in some cars (some technologies are further spread than others, but even a GPS receiver is still something special for a car). So, as mentioned in the paper, very probably there will be very few of this kind of cars with all its technologies at the beginning which will result in quite a small benefit for the owners as most of the features of the car rely on communication. And if there is nobody to communicate with, we can't use it. So in my opinion it would be much cleverer to reduce the required components to a minimum to make the ideas work. But as I can't really tell which component to take out without limiting the possibilities we have when using *smart vehicles*, I have to agree to their proposal which is to establish the services first which are able to be used when only few cars are equipped. When these services are used by enough cars, we can create services which need more cars and so on until every car is equipped with the technology of *smart vehicles*.

4.1 Electronic license plates

In my opinion *electronic license plates* are (if designed carefully enough) not a bigger security problem than today's license plates. As mentioned in the paper, physical attacks are much easier to control and using someone else's identity would also become a more complicated problem. But there is one critical problem which they only mention in one sentence: a denial of service attack could kill the whole use of the system. As it is possible to jam every signal, we are able to jam all the signals at our desires. For example: A driver causes an accident and flees the scene (as mentioned in section 3.1). But as he flees, he uses another communication device (not built into the car) to jam all his identification signals. No other car will be able to store his identity and the whole concept has no use which means it is important to establish communication protocols which are quite difficult to jam (impossible is impossible). This way only few people can take advantage of this security problem. But for sure it will continue to exist and should be further examined (before establishing an *electronic license plates* network).

What about privacy concerns? In my opinion, these *electronic license plates* are preserving privacy as much as the license plates we use today. If we take a look at today's license plates, we will normally find information about the country the car is from, some additional information where the car is from (for example the city in Germany) and a combination of some letters and numbers. So it is possible to determine the origin of the car with one look. To figure out the owner of the car we would need a database which translates the alphanumerical combination into a person's data (which is only accessible by the authorities). *Electronic license plates* will also offer a code which you have to translate into personal data (and only the authorities have access to it). Depending on the design and complexity of this code, it can be easier or harder to retrieve information out of it. It also could be possible to eliminate direct information about the car's origin (like the city) to raise the privacy level. Monitoring locations would also be quite the same. Whether we install a video camera to record all passing license plates or whether we install a communication device which records all *electronic license plates* of passing cars, the privacy is neither higher nor lower, only the time we need to evaluate the data differs. But there is one privacy drawback we will experience with *electronic license plates*: Depending on the protocol it could be needed to communicate someone's *electronic license plate* over a huge distance within a short period of time. All the cars on the way from the car which wants to identify itself to the point where it needs to be identified could listen to it and track the car's location. Periodically changing pseudonyms (in a carefully designed protocol) can help preventing this problem but *if* the tracker figures out how the mechanism works which changes the pseudonym, he could track a car with minor effort (The protocol in section 5 uses more or less random new pseudonyms). Today it is much more difficult to track a car's location.

4.2 Location verification

I have to agree that GPS receivers are quite insecure and easy to fool. As GPS signals are very weak (10^{-16} Watts at the Earth's surface [WaJo03]), they are very easy to override with stronger GPS satellite simulators. As these simulators are quite cheap ($10,000-$50,000 [WaJo03]), we need to install more intelligent GPS receivers in our cars to be able to decide whether it is a real signal or a (much too strong) fake signal. Without these changes I don't see a possibility to use GPS for more purposes than it is used today without taking a high security risk. Even today it is dangerous to rely on GPS in certain ways.

Verifiable multilateration seems to be the better solution as it does not accept any false data (only inaccurate positions are possible). But is this solution really easy to realize? We would need base stations all over the country to be able to determine a vehicle's location at every time. The network would have to be without any holes or a car could disappear in a region where there are no base stations. Even the established mobile phone networks can't guarantee a 100% availability, how could a new network suffice this requirement? To use this solution efficiently, we would have to build all base stations at once (and connect them somehow) which would cost much more than any company or country would like to spend on a project that can only determine the location of our cars. Only if we reused a network which is already there or if we used the new network for more than just location determination, there would be a (financial) possibility to establish this kind of network. The GPS solution would also need some base stations to communicate their location to, but it only needs one reachable (and reachable can also mean "routed through other cars") base station in opposition to three or four directly reachable base stations during one location verification.

One thing we should not forget in both solutions is that only authorities should be able to build (and run) these base stations. As soon as it is possible to run your own private base station, you could act just like a real one and intercept the position data (GPS solution) or find out about the distance between you and the car (verifiable multilateration). As it will

be impossible to keep the hardware only available to the authorities, one should think about this problem when designing the protocol which establishes the connection between car and base station.

4.3 General comments

I think the authors had some good ideas concerning privacy and security for *smart vehicles*. I would like to emphasize their point that privacy and security aspects did not stop the internet or mobile phones. In both networks it is technically possible to determine at least your identification (and sometimes your location), but the networks are very widespread all over the world. And we can't deny that it will be possible to determine identification and location for any car if we establish the solutions mentioned in the paper. However, as long as it remains difficult enough, there is a possibility that this solutions are really going to be used someday.

The structure of the paper is very good, and it is very easy to understand. Their points are convincing, I never thought that any of the content is wrong. But I think they could have done better. Of course this is a good overview about problems of security and privacy, but for its length, it only has few information that really concern privacy and security (maybe that's why I never thought of anything to be wrong – few information means few possibilities to give wrong information). Their introduction to this matter (with all its information about *smart vehicles*) is in my opinion much too long. People who are concerned about security and privacy of *smart vehicles* should know what possibilities there are to use these cars. They also look at only two points of privacy and security. This wouldn't be bad if they had looked deeper into them (especially on the matter of *electronic license plates* they just touch the surface in my opinion).

On the other hand, I have to emphasize that this paper is the only one that really looked into this topic of privacy and security. The only other one I found concerning this topic was "Security Issues in a Future Vehicular Network" by El Zarki et al. [ZMTV02], but it just touches some security issues while totally ignoring the privacy part. To be more precise, they think that there is no confidential data transferred in a vehicular network.

5 Electronic license plates – A protocol

As the idea of *electronic license plates* is only roughly described, I would like to discuss in this report a more concrete solution. Of course we need a central authority that administrates all information about all cars. It is responsible for all changes concerning *electronic license plates* like changing pseudonyms or registering new cars and is the only authority that is able to connect a car's pseudonym to its owner or to certify a pseudonym. In my solution, it has to store (in addition to the owner's personal data) the current *symmetric key* K_c, the current pseudonym P_c and a *one-way hash function* F that does not change over time for each car. The car itself stores K_c, P_c, F and a certificate C_c for its current pseudonym. All the internal data should never be visible to the car's owner or any other person.

5.1 Establishing connections – authentication phase

The vehicle periodically sends out its pseudonym along with the corresponding certificate to identify itself against other vehicles (for example to be registered after an accident) or base stations that belong to the central authority. (Cars also send out other data like traffic

information etc., but I will focus on the *electronic license plates*.) If it wants to communicate with a base station, it adds this communication request to the outgoing message. As soon as a base station receives this connection request or wants to establish this connection for other reasons (e.g. to inform the vehicle of an upcoming traffic jam), it contacts its central database and reports that it wants to start communicating with the vehicle with the pseudonym P_c. The central authority generates a *challenge* Ch and calculates $F(Ch)$ (of course by using the F corresponding to P_c). Then it provides to the base station these two values and the current *symmetric key* K_c. B now sends the encrypted Ch to the car. The car calculates $F(Ch)$ and encrypts it with K_c. After having received this value from the vehicle, the base station compares it to the one received from the central authority. If they match, both communication partners can communicate using K_c for encryption. A short version of this protocol can be found in table 2.

Source	Destination	Message	Used key
V	B	P_c, C_c(, "New connection please!")	-
B	C	"Give me the connection data for P_c please!"	-
C generates Ch, and calculates $F(Ch)$			
C	B	Ch, $F(Ch)$, K_c	-
B	V	Ch	K_c
V	B	$F(Ch)$	K_c
B compares values, on match: connection established			

Table 2: Establishing an encrypted connection between base station B and vehicle V using central authority C

5.2 Changing pseudonyms

As mentioned in section 3.1, pseudonyms have to change over time or else anybody who once found out about the connection between pseudonym and real car would always know where to find the car.

By following the upcoming protocol, a vehicle is able to change its pseudonym and the current *symmetric key* as soon as it wants to. First we need an encrypted connection between the car and a base station (use an existing one or create one as explained in section 5.1). Now the vehicle sends its request to change both the pseudonym and the key to the base station which passes the message to its central authority. The central authority randomly chooses a new pseudonym P_n(that is still available) and a new random key K_n, generates a valid certificate C_n for the new pseudonym, and locks this car's data so that no other base station can apply for another new pseudonym for this car. After receiving these three values from the central authority, the base station passes the new pseudonym and the new key to the vehicle. As soon as it received the confirmation (already encrypted using the new key K_n) of getting this data from the vehicle, it sends the new certificate C_n to the car. Now the new pseudonym, key and certificate can be used instead of the old ones. To complete this change, the car needs to send out its new identification including P_n and C_n and the request to establish a new encrypted connection. The central authority will unlock the car's data and only use the new pseudonym after receiving the corresponding message from the base station. Table 3 gives you a short version of the protocol.

Source	Destination	Message	Used key
V	B	"New pseudonym please!"	K_c
B	C	"New pseudonym for P_c please!"	-
C chooses P_n and K_n, generates C_n, and locks data			
C	B	P_n, K_n, C_n	-
B	V	P_n, K_n	K_c
V	B	"okay"	K_n
B	V	C_n	K_n
V	B	P_n, C_n, "New connection please!"	-
B	C	"Give me the connection data for P_c!"	-
C unlocks data			

Table 3: Changing pseudonyms (vehicle V, base station B, and central authority C)

5.3 Additional explanations

5.3.1 Establishing connections – authentication phase

I based this protocol on the GSM authentication protocol [MoST94] which also has three involved parties (similar to my three parties car, base station, and central authority). The authentication works nearly the same with two differences: First we don't use keys per session that change every time we contact a base station like GSM does. I don't think this is necessary as we are able to change the key as soon as the vehicle wants a new one. But we get an advantage out of this which leads us to the second difference: the *challenge* and the corresponding answer are communicated encrypted. The advantage of this change is that our secret function F is better protected. Without encryption it would be possible to fake a base station and challenge a car as often as needed to find out about F because we receive the corresponding answer from the car. With encryption the car finds out about the fake base station as soon as it decrypts the *challenge* and the text doesn't match the usual *challenge* pattern. So it can ignore this *challenge* instead of giving any information about its secret function F.

5.3.2 Changing pseudonyms

The pseudonym changing protocol is designed to be able to be interrupted at any time without the possibility of having a car with no valid pseudonym and certificate. As we have established an encrypted connection before demanding a new pseudonym, we don't need to worry about security. So I will concentrate on explaining how the system works in case of an interrupted connection.

As long as the communication is interrupted before C locks the data, nothing important has happened and the car has to repeat its request for a new pseudonym. If the interrupt happens after unlocking the data, the change of pseudonym and key is finished and we can use the new values just like the old ones before the change.

But if we interrupt the change while the data is locked, there could be a problem with the locked data. However, I will prove that it is not the case. When we lose the connection after we received the new certificate for the new pseudonym, the change is done. As soon as we try connect to another base station with our new pseudonym, the central authority will unlock the data and also use the new pseudonym instead of the old one (as if the car had contacted the base station used for changing its pseudonym).

If the connection is interrupted before the vehicle has its new certificate, the change is incomplete. The car will still send out its old certified pseudonym to identify itself to its surroundings (and will have to request a new pseudonym if wanted). When a base station requests the corresponding connection data from the central authority, the central authority realizes that the old pseudonym is still in use, deletes the newly generated values, and unlocks the data (the old data). A timeout mechanism (which would delete the new data after a while of not having a connection attempt with the new data) is not needed. It could even result in errors if the car just left the range of the base station from which it got its new pseudonym and can't find a new base station to communicate with.

The last case is that a car sends out its old pseudonym and after that receives its new certificate. As this could result in a very unpleasant situation (the car uses its new pseudonym which the central authority has deleted), we simply disallow sending out the old pseudonym during a change of pseudonym. If the change does not finish early enough before we need to identify ourselves, we have to use the old pseudonym. But in this case, the change is handled as if it never happened, meaning the car needs to demand a brand new pseudonym over a new connection to a base station (by establishing a new connection we delete the "old new" values).

5.3.3 Other design decisions

There are some other things that need to be looked at when designing these *electronic license plates* which are not necessary for the protocol. In order to be able to use the recordings of other cars that noticed a car fleeing from an accident, the central authority would have to store all pseudonyms ever used by a car along with the corresponding date and time. Depending on the frequency of the car's pseudonyms change, this could be a huge amount of data per car (much more than just the personal data). But the change needs to be quite frequent to avoid leaving a trace, so we have to be very careful when we decide how high this frequency shall be. A change of pseudonyms should not take place at a certain time (e.g. exactly once per hour) but should be at least partially random. Else anybody tracking a car could just wait until this certain time and then use the new pseudonym this car sends out at this specific time. It should be the same with the frequency of sending out your own identification. If you send out your pseudonym e.g. exactly once per minute, an attacker could just wait until the correct moment to get the new pseudonym you just received.

Of course there are other design criteria like length of key and pseudonym (longer means more security but also more communication) or the infrastructure of base stations and the central authority (How many connections to cars are possible at once?), but these are quite specific and not needed for this protocol outline.

6 Conclusion

As only few people have ever worked on the issue of privacy and security in communicating cars, I think the paper gives a good approach to the issue. The protocol which gives a more concrete solution of authenticating and changing pseudonyms than the idea of *electronic license plates* mentioned in the paper is surely not designed to be the only solution in this matter, but it is a first attempt. Until we really have all our cars equipped with the needed technology, many years will pass and mobile communication will advance. We also need to find someone who can finance this whole infrastructure of base stations before we can develop *electronic license plates* efficiently. Until then we will stick to our old fashioned license plates...

References

[BrCh93] Stefan Brands and David Chaum. Distance-Bounding Protocols. In *Theory and Application of Chryptographic Techniques*. Springer-Verlag, 1993.

[HuvL04] Jean-Pierre Hubaux, Srdjan Čapkun and Jun Luo. The Security and Privacy of Smart Vehicles. *IEEE Security & Privacy Magazine* 2(3), 2004, S. 49–55.

[MoST94] Refik Molva, Didier Samfat and Gene Tsudik. Authentication of Mobile Users. *IEEE Network Magazine, Special Issue on Mobile Communications Technologies* 8(2), 1994.

[WaJo03] Jon S. Warner and Roger G. Johnston (Hrsg.). Think GPS Cargo Tracking = High Security? Think Again. tech. report, Los Alamos National Laboratory, September 2003.

[ZMTV02] Magda El Zarki, Sharad Mehrotra, Gene Tsudik and Nalini Venkatasubramanian. Security Issues in a Future Vehicular Network. *European Wireless*, 2002.

List of Figures

List of Tables

Review of "MDDV: A Mobility-Centric Data Dissemination Algorithm for Vehicular Networks"

Jean-Thomas Célette

Abstract

Due to the technical progress of electronic in-vehicle equipment as well as communication and localization devices like GPS, driving has become far more comfortable and secure. These advances more and more also focus on vehicle to vehicle communication, meaning the ability of two or more cars to exchange information about presence, location and potential dangers or to exchange such information with roadside access points. A major problem in the conception of such peer to peer vehicle networks is caused by the constant mobility of the cars and the unpredictable density of the network users. The MDDV algorithm has been designed to handle and exploit this particular property. The main idea lies in using the mobility of the network to help transporting the information by using geographical information to enhance message forwarding. MDDV makes every network user a router by combining both network-specific knowledge, geopositional information and the message purpose.

1 Introduction

During the past twenty years, computers and related applications have found their place in nearly every task. Not at last have they been used to enhance security and comfort in vehicles. The latest interests focus on in car-communication. Mobile internet access and the ability to exchange data between multiple vehicles is more and more on the way to become a challenge for modern car design. On the VANET conference and on the ACM MobiCom workshop different ideas and research results about these Vehicle related Networks are presented and discussed. Topics concern network architectures for Mobile Ad Hoc Networks, Feasibility studies, performance test, general questions about the physical network layers and the communication protocols themselves.

A usual approach to mobile communication is similar to mobile computing in wireless networks. Roadside access points, that are part of a big wired network, provide a wireless network for moving vehicles. Thus, vehicles use the access points as gateways, as a mobile computer uses a wireless access point. A different approach to mobile networks is the design of mobile ad hoc networks, meaning that vehicles communicate directly with other vehicles. This approach compensates the lack in the density of access points and is cheaper to realize. But on the other hand, complex protocols have to be designed to provide message forwarding and other services. Due to the unpredictable and unreliable structure of networks composed of mobile nodes, the challenges of vehicle to vehicle networks are quite different to those of classical static network topologies. Different proposals have been made for message dissemination in such mobile ad hoc networks. Message dissemination basically means the transport of a data block from one place/node to another place/node. MDDV is an algorithm that actually exploits these compromising properties of the vehicle ad hoc networks to enhance

message dissemination in mobile networks. In this text we will describe and explain this algorithm.

This text is a technical report that was made in the framework of the VANET seminar at the University of Karlsruhe. It is mainly based on a paper by Hao Wu, Richard Fujimoto, Randall Guensler and Michael Hunter called MDDV: A Mobility-Centric Data Dissemination Algorithm for Vehicular Networks that was presented at the VANET Conference in October 2004. The paper offers an easy approach to V2V networks and to MDDV. It presents the algorithm and some testing results in a well structured text. The terminology of this technical report will mostly be the same as in the referred paper. In the second section of this text we present the paper, resume it and explain or illustrate some parts of it. In the third section we will discuss it, mainly by evaluating the simulation results presented in the paper and in other related documents. In section four we will make proposals for future development based on MDDV, by mainly trying to augment its reliability. Advantages of hybrid architectures will also be discussed in section 4. We will finally conclude the paper in section 5.

2 Presentation of the Paper

2.1 Vehicle to Vehicle Networks (V2V)

In its second section, the referred paper presents properties and global goals of V2V ad-hoc networks. MDDV requires some global and environmental information including the current position of the car, a geographical road map, as well as engine statistics, crash reports and so on. These informations are provided by technical equipment such as a GPS device and different sensors that are assumed to be available on the vehicle. With the GPS device a vehicle knows its position at any time, and thus the position can be used to compute the behavior. Anyway the vehicles only know their own position, but not the position of their neighbors, in order to maintain a minimum of privacy. Of course each vehicle running MDDV needs a wireless network device like a IEEE 802.11 compliant network interface. However, MDDV does not assume itself to be available on every vehicle, as at the beginning only a few will be equipped.

MDDVs goal is to achieve message forwarding on a very special type of network. Therefore it is based and uses the very special characteristics of this network. In the paper these characteristics are assumed to be as follows:

- the nodes are constantly moving and their movement can be used for message transportation. The mobility is regular and predictable as vehicles move along roads and are constrained by traffic regulations
- the network architecture is constantly changing due to the high mobility of the nodes
- the nodes are following predefined roads therefore their movement can be assimilated to a one-dimensional trajectory
- the total distance between nodes can be very large; the networks size can in principle be as big as the road network
- the nodes are unequally repartited: in some areas the node density is very high and between such areas it can be very low so that the topology can be assimilated to interconnected islands [DoTH02].
- the equipped vehicles can be failing (whole vehicle, part of the equipment, network connection)

- the power consumption of the equipment does not require special power sources and the power sources are reliable (no loss of power)

The authors of the paper say that these properties influence the design of the V2V networks. The partitioned highly mobile network topology make big logical structures (as the network graph) obsolete. It makes no sense to precompute the route of a message through a network if the network structure may change every moment causing the route not to exist anymore. The longer (in terms of nodes) a route is, the more probable it is that one segment is cut. So it is rather desirable to use a localized routing algorithm that only compounds nodes that are directly accessible.

As a directly accessible node is not completely reliable and may not be able to replicate a message any further, it is useful to replicate the message to augment its chance of reaching the destination node.

The main design parameters and objectives concern the effectiveness of the V2V network. The authors distinguish four primary design objectives: a *low delay*, a *high reliability*, a *low need of memory* and a *reduced message overhead*, meaning the ratio of unnecessarily sent messages. The delay means the time that a node needs to forward an information and it is usually low when the reliability is high. A big high message overhead also increases the memory needs, as it has to be stored on multiple nodes. The reliability mainly depends on the networks structure and increases with the message redundancy, e.g. with the overhead. It seems that the first parameters (delay and reliability) oppose to the overhead and memory needs. But of course the overhead leads to a higher network load and a higher network load would decrease the reliability and increase the delay times. Thus the overhead (redundancy) needs to be controlled to make sure the network does not get jammed by messages replicating in any way or nodes sending the same message more than once. To enable such a controlling system, information has to be placed in the message header, which will results in higher memory needs. Anyway, such information is no bigger than a few bytes, and the messages themselves being potentially short, they can be stored without needing much memory, compared to current memory sizes.

Other design parameters concern the dissemination of the messages. A main difference between usual peer to peer (P2P) networks and the V2V networks is that in P2P networks communication happens between two specific nodes in both directions, whereas in V2V networks it often involves one vehicle sending a message to one or more vehicles in a certain area. Those can be specific vehicles or might not. Using a usual IP-similar protocol would highly increase the network traffic as only one specific node can be reached. This would not satisfy the applications of V2V networks.In the paper the authors distinguish between four data dissemination semantics:

- unicast with *precise location* means sending a message to a specific node at a specific location so that the message reaches its destination before a specific time
- unicast with *approximate location* means sending a message to a specific node which actual location is unknown but a previous location is known and its actual location can be guessed in a certain area
- multicast means the dissemination of a message to all vehicle in a certain region: for example all vehicles within a circle around a specified position
- anycast means sending a message to at least one node of a certain type in a specific region: for example "'send to any police car"'

Each of these message forwarding systems has its own application in V2V networks. Multicast can be useful for warning vehicles in a certain area of to request information (e.g. about traffic density) in a certain area. Later in the text a difference is made between local multicast and multicast for a remote location. An application for local multicast could be to send a warning to all surrounding cars. Another application could be to use information of the cars stability program (ESP, ASC and how these systems are called) to automatically warn surrounding cars when the road seems to be icy. Unicast can be used to send such information back or to warn a nearby car. Anycasts application can be the sending of distress messages. Of course many more application of these dissemination semantics can be found.

Those properties of V2V networks and the design objectives for the forwarding algorithm bring up three main properties of MDDV: as the nodes are highly mobile and the communication does not happen between specific nodes but between nodes that are chosen because of their location, MDDV can not be node centric as we would expect it from IP-routing protocols but needs to belocation centric e.g. the main transmission criteria is that a message gets closer to its destination region. The second main characteristic of MDDV is that for transmission over numerous segments, a node can not know if an immediate connection is possible. Therefore *opportunistic forwarding* has been introduced meaning that messages are stored until the opportunity to send the message to some other nodes presents itself. The third property concerns the number of nodes the message is forwarded to; if there is no certainty that there is a way to the destination over the node with the closest position to the destination, the message has to be transferred to another node. Therefore messages have to be redundantly forwarded.

2.2 Presentation of MDDV

The third section of the paper is the biggest one and it presents the MDDV algorithm itself. It explains how multicast to a distant region is done. A pseudo code implementation is shown in the section 4.

As described in the previous section, one of the main objectives is to deliver messages as soon as possible. Therefore it is necessary to determinate the path that can carry the message the fastest. The easiest way to handle this seems to be taking the shortest path. But the traffic density has a main influence on the message propagation, as a high node density would allow a faster message propagation. To handle factors as traffic density our authors have introduced the dissemination length. Like the path length is used as the edge-weight when solving shortest path problems the dissemination length decreases when a road offers good possibilities to transport a message, e.g. when it has a higher traffic density. As traffic density is difficult to determinate our authors assume that the more lanes a road has, the higher the traffic density is e.g. a four lane highway will not only be faster for vehicles to drive but also for messages to be forwarded than a one lane country road. The formula that determines the dissemination length d(A, B) between a point A and a point B is the heuristic formula

$d(A,B) = r(A,B) \cdot (m - (m-1)(i^p + cj^p))\ 0 < c < 1$

where $r(A,B)$ is the road length of the chosen path. The global idea of the formula is that the road distance is multiplied with a formula based on other parameters:

- i is the number of lanes from A to B and j the number of lanes from B to A
- p and m are some formula specific parameters that are used to influence the the weight of the parameters. p associates the car density to the number of lanes. It is unlikely that the number of lanes will be linear on the number of cars per road-length-unit. In the tests p has been set to 0.1 meaning that on a 2 lane road there are 1.07 times as much

cars than on a 1 lane road. m influences the weight that the lane-number parameters have compare to the road length and has been set to 5 for testing.

- as vehicles moving in the opposite propagation direction are less useful than vehicles moving in that direction the parameter c has been introduced to lower the advantage caused by opposite lanes. In the test c has been set to 0.05.

The dissemination length of a path joining two distant regions will be the sum of the dissemination length of the road segments. Of course the dissemination length requires that a precise road map with lane information is available to the system.

2.3 Message heads in localized networks

As presented in the the section 2.1 MDDV was going to be a localized algorithm as only local knowledge is present at computation time. Anyway, the collective behavior of all the nodes involved in a message forwarding has to reach the objectives. In the semantic of multicast the algorithm has two main objectives: bringing the message to the destination region (here called forwarding phase) and then transmitting it to every node in there (here called propagation phase).

While propagation is rather easy to implement, message forwarding is rather complex. In theory it would be best to only let the node that is the closest to the destination region (here called message head) send the message to avoid information overflow. In practical this would not work because there is no perfect knowledge, meaning that not every vehicle can know the real message head at any time. If the message head leaves its trajectory or stops or becomes inoperative, there is no message head left and the message forwarding would stop and the message would get lost. Also if the message head travels in the wrong direction, sending the message before it crosses another vehicle, the other vehicle would not know it is the new message head unless it knows that the other vehicle is moving away from the destination region.

To handle these problems MDDV allows multiple vehicles to forward the message as long as they are near the message head.

Information about the message head, that is constantly updated, has to be available to every message holder. The best way to handle this is to have some metadata being carried within the message. This metadata would be, agreeing to our authors the location of the last message head when it sent the message and the corresponding time. The message head information would only be updated when a message head moves closer towards the destination region or when a new vehicle becomes the message head. To design this, the notions of message head candidate and non-message head candidate have been introduced. A message head candidate is a node which can with some probability assume that it is the real message head, e.g. when it is closer to the destination region than the current message head. For this purpose the *message head pair (l, t)* has been introduced where l gives the location of the actual message head and t the time at which the message head had the position l. Thus, the authors consider two state-swhiches:

A non-message head becomes a message head:

- if it receives a message with a message head location (l) that is farther from the destination region
- if it passes the message head location (l) within a time that is defined as a system parameter

A message head becomes a non-message head:

- if it receives the same message with another message head pair (l, t) where l is closer to the destination region
- if it leaves the trajectory or moves away from the destination region trajectory

It is also required that a message holder updates its messages head pair if a newer one is received.

2.4 Data transmission

The next concern is about the transmission protocol used for data exchange. As an exchange protocol based on logical structures like trees, graphs or clusters can not be implemented in an unpredictable network like V2V, MDDV will have to focus on peer-to-peer connections and only be able to directly communicate with other nodes that are in reach. This makes the transmission protocol different to known protocols like IP where packets are addressed to a gateway if the destination host is not reachable within the network. In addition to that, the transmission must be able to address a message to any reachable host (as presented in 2.1) e.g. must be *location centric*. Thus, multiple receivers must be able to receive the same message within one transmission. That broadcasting property influences the possible data exchange semantics: a three way interaction would be difficult to implement, as the message sender would have to interact with multiple receivers at one time, and the idea of broadcasting could not be realized. Therefore, the best way to allow broadcasting is a one-way data exchange. The sender simply sends the messages when receivers are reachable (remember that all nodes are supposed to broadcast their id, independently from the transmission). This also supports the general idea of data dissemination and opportunistic forwarding and reduces the transmitted data to its minimum, which can be compared to a reduction of the overhead.

Furthermore, the data exchange algorithm is distributed in two phases: the forwarding phase, objecting the transmission from the source to the destination region, and the propagation phase, making sure that every vehicle in the destination area receives the message.

2.4.1 The forwarding phase

The authors distinguish 2 states that both enable message propagation during the forwarding phase: the active and the passive state. In the active state, a node replicates a message as soon as it receives a newer version of it and also sends the newest version of the message that it has stored in its memory when an older version of the message is received (older here means that the l parameter is farther from the destination area. It also resends the message each time a new vehicle shows up. They call this the *full protocol*. A node in the passive state will only transmit the message if it receives an older version of the message e.g. for error correction.

A node is in the active state as long as its location and time are within a certain gap from the message head pair. When this distance (location and time) is transgressed the node turns to the passive state as long as it stays in another certain distance from the message head pair. Out of this region the node does not transmit the message anymore.

2.4.2 Propagation phases

The propagation phase is much easier. The node does either transmit the message or does not: as long as both time and position are within a certain distance from the message head pair, the message is transmitted. Out of this distance, it isn't.

2.5 Implementation of MDDV

In the text a pseudo code implementation, that solves a part of the multicast problem, is provided. It implements three lists or queues as data structures to keep track of events:

- a neighbor list that contains the list of the neighboring nodes in their order of appearance
- an opportunistic message list containing the messages that have already been transmitted and are waiting to be transmitted to new neighbors
- a scheduled message list, that contains the message that have to be transmitted in the order they were inserted

Periodically the node sends the scheduled messages or sends messages that are stored in the opportunistic message list and that were last sent before a new neighbor showed up.

The second part of the implementation regards what is done when a new message arrives.When a new message arrives, the algorithm behaves differently depending on if the message was already known or not.

- When the message has already been received, the new message head pair is compared to the message head pair of the stored message. If the message head pair of the new message is significantly newer than the message in memory, the message is added to the scheduled list. If it is insignificantly newer or insignificantly older it is added to the opportunistic message list (and only the newest head pair is kept). If it is significantly older the message is added to the scheduled list (with the newer head pair).
- When the message is unknown, it is whether added to the opportunistic list or to the scheduled list, depending on its state, active or passive as defined in the previous section. The case that the message is not eligible to be transmitted at all is not implemented.

2.6 Discussion

Section 5 of our referred text discusses the design of the algorithm and resumes the potential failures of the algorithm. It also gives an idea about what future research work has to be done, which is also discussed in sections 7 and 8 of our text. That question will also be addressed in one of the following sections of this report.

In its second part section 5 discusses the other data dissemination semantics that have been addressed in the section 2. It globally says that the other data transmission semantics can be handled by modifying multicast. Unicast corresponds to the forwarding phase of multicast, scan is a special case of unicast and anycast is similar to scan. Of course, the localized approach of multicast (when the source is within the destination region) can be regarded as a special case of the discussed multicast. The similarity of these semantics offers new design possibilities which will be mentioned in part 3 of this report.

2.7 Realized Testing

In the text the authors have dedicated a section to the realized testing of MDDV. Unfortunately only a few results of the test are shown, but let's first explain the testing system.

No real testing has been done on MDDV yet, but a simulation system was built up. It basically consist of two interacting software systems, one called CORSIM [t08a], used to simulate the

behavior of the vehicles, and the other, QualNet [t08b] used to simulate the behavior of a Wireless Network. According to the frederal highway association (FHWA) CORSIM is a comprehensive microscopic traffic simulation, applicable to surface streets, freeways, and integrated networks with a complete selection of control devices (i.e., stop/yield sign, traffic signals, and ramp metering). CORSIM also simulates traffic and traffic control systems using commonly accepted vehicle and driver behavior models. For the testing, the MDDV team has simulated the morning rush hour traffic in the northwest quadrant of Atlanta, corresponding to 12 KM of the I-75 Corridor and 160 KM of streets.

MDDV has been implemented as an application level protocol to be run on the described testing system. Beside MDDV two other data-dissemination schemes have been implemented for comparison purposes. Both are unrealistic. The first one, called central intelligence scheme, is based on the assumption that any car, knows the identity of the message head at any time. So there is no loss of message possible and only the message head propagating the message. This is of course not realistic because is there is no way to have such a perfect knowledge in a V2V network.

The second approach is the P2P scheme that does not implement any message head at all so that every vehicle propagates the message every time another node comes across.

MDDV is a mixture of the two other schemes, that's why comparison makes sense. The concrete simulation involved a strain of 40 geographical-temporal multicasts with a message size of 512 bytes over an average road distance of 6.5 km. The radio range was set at 250m with an IEEE 802.11DCF device. The message expiration time was set to 480 seconds (8 minutes), meaning that after that time the message would not be forwarded anymore.

3 Evaluation

3.1 Test results

To discuss the provided test results let us first introduce some notions:

- the *delivery ratio* is the fraction of the message that are delivered within the expiration time compared to the total amount of messages that were sent
- the *penetration ratio* is the fraction of MDDV capable cars compared to the total amount of cars on the road
- the *delay* is the time that a message takes until it reaches the destination.

The test results shown in figures — show the relation between quality criteria as delay or delivery ratio with the penetration ratio. The figures are taken from the referred text.

In Figure 1 the delivery ratio is at 60% for a penetration ratio of 10% which is quite insufficient. At a penetration ratio equal of greater to 20% the delivery ratio is more or less constant at 90% which is an acceptable quote. But a penetration rate of 20% is extremely high because it would take a long time until this quote of equipped cars is achieved. Unfortunately there is no such data available for the other idealized dissemination semantics.

Figure 2 shows the maximum, minimum and average dissemination time of the messages, where messages that have reached the timeout are not taken in account. The average is between 100 and 200 seconds, while the delay sinks when the penetration ratio augments. On the first sight 200 seconds (at a penetration ratio of 20%), is very long if one remembers that the average distance is 6.5 km. 200 seconds being only one way dissemination, a car

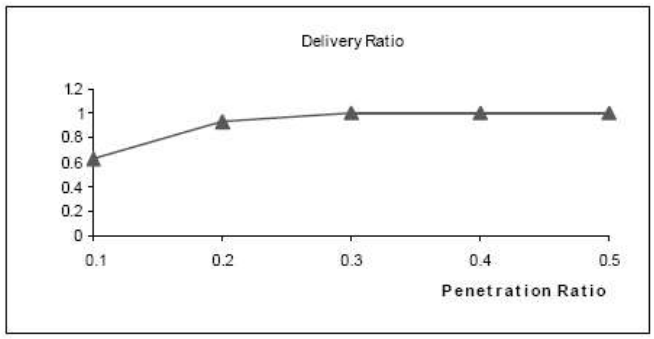


Figure 1: Delivery Ratio

requesting traffic information for a location that is 6.5 km away will wait approximately 400 seconds for the answer. In [WLHF+04] current average speeds on a highway like I-75 are 60 mph (96.5608 km/h) and 15 to 25 mph (24.1402 to 40.2336 km/h) on a street. In 400 seconds a vehicle can respectively cover a distance of 10.7 km at 60, 4,470 km at 25 and 2.6 km at 15 mph. So obviously such a request will not make sense on the I-75 because the destination region will be passed before the sender gets the reply. At 25 mph the results might arrive too late but at the low speed of 15 mph they might still be useful. A second aspect of the test results shown in figure 2 is the difference between the average delay time and the maximum and minimum delay. The minimum delay is near to 0 seconds, whereas it must be considered that this could be due to the destination region being chosen next to the source node, as 6.5 km is only the average distance. Also the maximum delay of delivered messages being close to the expiration time can simply be due to the same fact as no upper limit is given for the destination region. The empirical variance for a fixed road distance would be more significant for the testing. Indeed, other studies on data dissemination as [WLHF+04] show similar results: "'the message propagation delay is highly variable except when vehicle density becomes saturated. A particular delay may be well below or above the average, depending on prevailing traffic conditions."'

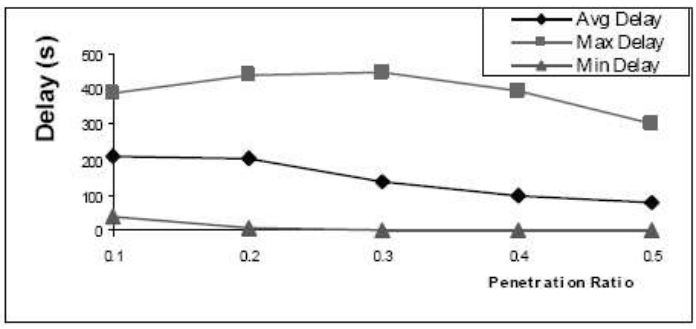


Figure 2: Dissemination Delay

This variance in the message delay highly affects the reliability of the protocol. Furthermore it can empirically be admitted that the delivery ratio falls and the dissemination delay linearly grows when the dissemination path gets longer. The average delivery ratio of 0.6 at 10% penetration ratio would show awful results on longer paths. A request over a 13 km route, which is a reasonable distance to obtain traffic information Would have a delivery ratio of 0.6*0.6 = 0.36 . If we assume that that request is made by an anycast, so that only one

vehicle answers it, the delivery ration of the answer would be even lower. According to the penetration ratio of GPS devices in cars in the early 90s, it cannot be assumed that more than 10% of vehicles are equipped with MDDV and the required devices within the first year on the market. And these bad results would discourage the consumers so that the delivery ratios at 20% penetration ratio (almost 100%) will never be reached, and that is a major problem opposing to the integration of such a system.

The other test results showing MDDV in comparison with the two idealized dissemination algorithm are less significant. Figure 3 shows the delivery ratio of MDDV normalized against this of P2P and it appears to be lower than the delivery ratio of MDDV itself. This is due to the fact that the test use an MDDV implementation with reduced message overhead, what actually makes MDDV nearly comparable to central intelligence -but without the perfect knowledge. Thus this delivery ratio is not that significant.

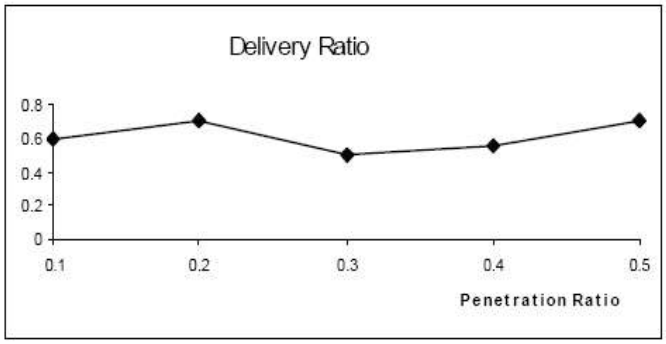


Figure 3: Normalized Delivery Ratio

Figure 4 showing the message overhead for this particularly overhead-low MDDV implementation shows that the overhead is nearly as low as the one of central intelligence which cannot be surprising. On the other hand, it appears that the message overhead is very low, mostly $< 5\%$. Such a small message overhead can not be affecting the network reliability. It is also astonishing that Central Intelligence has any message overhead at all: as every vehicle has perfect knowledge of the message head at any time, no unnecessary message should be sent out. The minimum overhead at 20% penetration rate fits with the maximum in the delivery ratio, but cannot be explained with these test results.

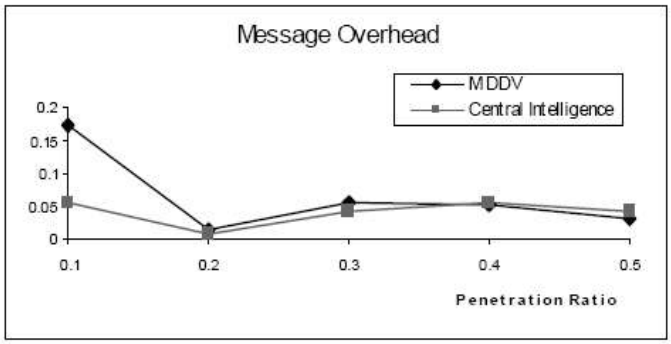


Figure 4: Normalized Message Overhead

The simulation results show that with the right parameters, MDDV can be a reliable dissemination algorithm. Prognostics on dynamic MDDV parameters are made in section 4.

The results also show that the effectiveness of MDDV mainly depends on the penetration ratio. Similar results are shown in [WLHF+04]. They confirm that a minimum penetration ratio of 20% is needed for acceptable results. Figure 6 shows a very low connectivity for a penetration ratio less than 15% and a 100% connectivity while the penetration ratio is at least 20%. Similar results can be observed for the delay time.

[WLHF+04] also reveals another weakness of V2V connectivity that also affects MDDV. All the simulations we have considered yet, assume whether morning or evening rush-hour in Atlanta. Thus the reliability of V2V networks in general mainly depends on traffic density, and at low circulation time like nighttime, the reliability is much worse. Simulation results shown in [WuFR03] also confirm this.

The simulation results published in the text also do not show how the dissemination-path-length affects the reliability as only an average distance of 6.5 km is given. It can be admitted that reliability gets worse when the path gets longer, but it is important to know if it come to a physical limit when using MDDV, and if system parameters need to be adapted for long distance messages (e.g. Having a timeout that is proportional to the dissemination length). On the other hand, when roadside access points can be used in hybrid architectures the dissemination path that consists over the V2V network is limited as well.

Another problem that's not addressed by the simulation is the given road infrastructure. All simulations that we've seen are made in Atlanta, a big city, and in big cities you've always got vehicles on the road. What about country roads? Is a V2V network possible on a country road? We can assume it isn't when the traffic density is too low. A possibility to handle this problem will also be presented in section 4.4.

3.2 Dissemination Path

In the text it is unclear whether multiple dissemination paths can be used. Section 3.2 says that the path that will be taken is the path with the smallest sum of weights. Actually it is not specified whether this path is computed once by the message source and packed into the message, which will result in a larger message size, or if it is recomputed at every hop, which means more computing time, as this problem is NP-complete and needs a high computing time. In the case the message is recomputed, the path may not be unique. The reach of the wireless device is supposed to be 250m. 250m is a lot, and thus it is possible that two nodes on two different roads receive the message and each node propagates the message which may result in multiple trajectories.

3.3 Message head

Let us consider a vehicle that drives slowly on a road, transporting a message. Assume that a fast vehicle comes on the road, broadcasting its ID. The slow vehicle will send the message, the fast one receive it, but before the faster vehicle sends it with a new message head pair the two vehicles come out of reach. This way the same message might get forwarded redundantly. This phenomena might get amplified when the vehicles drive in a city with buildings and other vehicles on the road that might suddenly jam a transmission. Changing the lane is sufficient to suddenly have a truck in the way that might break the transmission. In general, it can be said that wireless ad-hoc networks have a limited reliability by themselves and that property is not watched in the algorithm design. All these things may affect the message overhead, because redundant forwarding actually doubles the number of transmitted messages.

These aspects of wireless ad hoc networks have also not been taken into account for the simulation.

4 Further development

The MDDV algorithm provides a service for message dissemination in V2V networks. Thus applications for MDDV have to be designed and implemented. An important aspect of V2V communication and MDDV in particular is that it provides only a one-way connection. MDDV provides a service similar to UDP in IP connections. An other aspect of V2V communication is a easily changing reliability that hardly depends on the localized network architecture and node repartition. In general, such a system prefers short messages, but when the connectivity is good e.g. when the nodes are close to each other bigger messages can be exchanged and even connection orientated point to point transmissions could be implemented. The following parts ask general questions on protocols that could be based on MDDV and propose some design parameters.

4.1 Dynamic message length and predefined messages

An approach to make MDDV scalable to the connectivity degree consists in building a protocol on MDDV that provides basic services, and also the ability to transfer longer messages. Such a system could allow it to distant vehicles with an unreliable connectivity to only exchange basic information with a very short message template. Vehicles being close to each other, that therefore benefit on a high connectivity, resulting in a fast connection could run higher level protocols like TCP/IP or similar protocols to exchange higher amounts of data like movies or propagate internet access or permit inter-vehicle network gaming as suggested in [WLHF+04]. This way we would implement short messages for unreliable connectivity and longer messages when the connectivity is good. This option could well be implemented in the message header, and some basic messages like a be careful or police car request or what is the traffic like at position p? could be implemented as standard messages to make them very short. Of course the message format must provide different possibilities depending on the message's semantic. Unicast does not make sense when combined with a traffic information request or a police car request.

To start making a proposal about the message format let us first think what all types of messages need in commen on their header. Basically there would be the sender ID, where Ids must be unique as MAC-Addresses and would therefore need 12 HEX numbers e.g. 6 bytes. As well a message should always contain the senders location and the sending time. Time and date can be stored in 4 bytes and the two position variables in 2 bytes each. The expiration time needs another 2 bytes. To implement predefined messages another byte can be used (256 possibilities should be enough). We will call this byte the message describer, and beside the standard messages it could also hold the dissemination semantic. Longer messages can be handled as a special case, and another one or two bytes would specify the length. If the message describer needs arguments, we admit these must immediately follow the message describer. Such arguments can consist in addresses (only for unicast), in locations (unicast, anycast, multicast), and other arguments like direction, speed and so on. At last the message format should provide an error check system or even some redundancy, because the service is connection-less and an erroneous message can hardly be corrected. This should not make problems as the total message size for short messages as we have proposed them is very small as it should not exceed 30 bytes.

The smaller a message is, the less it jams the network, and the more it can be replicated without producing an important overhead. Thus, it can be thought to implement MDDV to let it prefer short messages and to let them take multiples paths, whereas long messages forwarding is optimized to reduce the overhead. Of course it must be thought that some long messages need a high reliability too, and, by the way, what would it bring to have a traffic

information request being delivered within seconds when the answer never arrives? The other question is about the achievement when using multiple paths.

4.2 Multiple paths

In the text a suggestion was made about using multiple path, and in the Evaluation part we have seen that multiple path use is even possible when it is not explicitly wanted. If you want to have multiple road-paths to be used to disseminate a message, there are different possibilities to do this. We present the following two:

- A car hearing that the message head is closer to the destination but seems to be on an other road (remember that every car has a GPS and a roadmap), could continue broadcast the message on its way, obeying to the algorithm defined in the text. This way the message could take multiple roads without having to mainly change the MDDV algorithm.
- A car could keep the message and stay in the active forwarding state even if it moves away from the destination region. Conditions like being closer to the destination region than when the message was first heard and having not heard the message since at least x meters ($x > 250$) would prevent from a multicast in *any* direction. This semantic would be favorable to using multiple paths.

The main advantage of such a multi-path dissemination algorithm is the augmented reliability. Especially in few and far between traffic conditions where the dissemination can easily be stopped due to the low density of cars (imagine only one car transports the message and then leaves the path), multi-path offers redundancy and thus a higher reliability. Of course the message overhead grows, and multi-path dissemination is only favorable in certain conditions and it has to be adaptive to the traffic situation.

4.3 Priority based MDDV

The two previous sections both seem to point at one thing: to rise the reliability (especially in reducing the message delay) of certain messages without creating to much message overhead and to implement multiple path forwarding for some messages and not for some others, the concept of message priority has to be introduced.

Some applications of MDDV naturally require a higher priority level. Calling a police car, or broadcasting emergency messages is highly more important than having an internet connection over a few vehicles or chatting with a neighbor. Of course a priority system could easily be abused if the sender decides what priority level his message has. Just imagine exchanging data with a nearby car over a few cars with MDDV. To get a more reliable connection who wouldn't think about giving his messages a higher priority? A way to solve this would be to implement standard priorities, e.g. emergency messages naturally have a higher priority than traffic situation request while those have a higher priority than user transfers like data exchange. This can easily be implemented if a message format as the one described in 3.1 is used. As the message type would be defined in the message header a vehicle that just forwards the message could immediately find out the message type, and thus find out its priority and deduce the forwarding semantic to use. In addition to that, no big amount of hidden data could be in such a predefined message, because by themselves such predefined messages are short and only provide room for very few arguments.

For user type messages a fairness system, that prefers short messages or messages from senders that haven't sent a message yet, could be implemented the the network traffic is high an

network jams occur. This would also prevent users from monopolizing the network for unimportant purposes.

Depending on the priority level, multi-path forwarding could be favored or not. The parameters proposed in the previous section on multiple-paths could be adapted depending on the priority level. A message with a higher priority would more likely take many paths while a lower prioritized message would be disseminated on less different road-ways. The times T_1, T_2, T_3 and the distances l_1, l_2, l_3 defined in the text could as well be depending on the priority level. This way a message holder would forward a message with a higher priority level more often as a message with a low priority level, and it would also be more likely that the message takes multiple paths.

These suggestions are proposals for further development of V2V networks using MDDV. Of course further analyses and simulations have to be done to see whether these suggestions are accurate or not, and if they are, studies and research would have to be done to determinate values of the parameters.

4.4 MDDV and traffic density

As we've seen, the reliability of MDDV highly depends on traffic density. In the previous sections we've discussed possibilities to adapt the reliability of MDDV. The next question we ask is: does it make sense to adapt MDDV parameters depending on traffic density?

High traffic density augments reliability, but also increases the message overhead. Thus it could be considered to use different dissemination schemes depending on the traffic situation. But how can we evaluate the traffic situation? Of course, before forwarding a message a node could send a traffic information request to choose the appropriate dissemination semantic, but this would result in more messages that can be considered as overhead and would would not solve the problem. Furthermore what would it bring to send a traffic information request to send another traffic information request? The second point is that traffic density is a localized problem, as the traffic situation may change within short ranges. Anyway it might not be enough to adapt the dissemination scheme to the local traffic situation, because this could end in leading the message to a place where the message never comes out because of a lack of equipped vehicles. This is a problem that has to be faced. Other equipment providing traffic information like radio channels or maps containing prognostic traffic information for each road segment could help solving the problem. Anyway, one thing remains with certainty: on a path with a lack of vehicles the reliability will surely be compromised. But when a node has traffic information on the whole path that the message should follow, it can decide about the dissemination semantic to use. When a path does not have sufficient traffic density or when a road (or worse: a bridge) is closed ,the semantic an be adapted, multi pathing enabled. And the more traffic information a vehicle has, the more loss of message can be prevented.

In an area with *very* low traffic density, a solution could be to couple the navigation system and MDDV. When a car has its navigation turned on, long term prognostics can be made about the vehicles path. Thus a message that has to be transferred through a region with a poor traffic density could be hold by a car which will likely cross the region, e.g. If its navigation system has computed a path through this region. Such an implementation would of course be subject to privacy concerns, that would have to be solved.

4.5 Hybrid architectures

The last facet of MDDV we will discuss in this paper, is the deployment of MDDV in hybrid architectures. Hybrid V2V networks consist of V2V ad hoc networks coupled with roadside

access points that are all connected to a wired network. The main advantage of this network architecture is that messages can be forwarded over long distances with a low delay and a high delivery rate. The delay in the wired network is usually so low (usually a few milliseconds to a few seconds) that it can be neglected compared to the V2V network that handles message-forwarding with MDDV. In result transferring messages over a long distance, is only compromised by the distance it covers using the V2V network. Thus, in hybrid architectures, transferring a message over 100 KM could be faster and with les overhead than another message transfer over 2 KM, depending on the distance to the next roadside access point.

Two approaches can be made to integrate MDDV in hybrid networks:

- each node is aware of the location of the access points (for example they could be on the map)
- or not each node knows the location of the next access point

In the first case MDDV could be configured to compute if it is preferable to send the message to the next access point or to forward it directly to the destination region, e.g. which V2V distance is longer. The access point would only have to compute the best situated access point and forward the message to it. The second access point would simply run the MDDV algorithm (with the difference that it has a fixed location).

In the second case the access point have to behave mostly in the same way, except that they have to catch the messages an forward them if it makes sense to do so. In the case an access point sends the message over the wired network it should notify the surrounding vehicles that they mustn't forward the message. The vehicles simply run MDDV.

The first approach would certainly be the better, because the delay should be lower when access points are taken into account for path computing. In general hybrid architectures provide a heavy advantage to pure ad hoc topologies. Similar results have been observed in [DoTH02]. Depending on their density, roadside access points could help to solve the problem we've addressed in 3. They could balance a low penetration ratio, which would facilitate the deployment of such a system.

5 Conclusion

MDDV is an efficient approach for data dissemination in mobile ad-hoc networks. MDDV uses the vehicles mobility and achieves its goal nearly as good as unrealistic idealized algorithms. Without having a too high message overhead it approaches the reliability and delay times of an idealized algorithm. But we've seen that it seems to be easily compromised by low traffic density and low penetration ratio so that it's deployment seems to be very difficult. Anyway it mustn't be forgotten that the idea of MDDV is still very young, and not much research has been done about it. And the simulation results seem to be promising while the penetration ratio is reaching 20% with high traffic density. In the last section of this report we have presented and accosted some methods that could augment the reliability of V2V message dissemination. Some of these proposals could be subject to more research and simulation in order to exploit the capabilities MDDV provides. Especially hybrid network architectures seem to compensate the weaknesses of V2V networks. Another benefiting approach is to augment the knowledge (on traffic situation, on access points or on vehicle movement), because it can easily be used to optimize the dissemination of a message. Multi path forwarding could also increase reliability and even seems easy to implement. In the report, we have also made suggestions about a special message format optimized for ad hoc networks and traffic purposes. This approach could be benefiting to any kind of partitioned, unpredictable and partially unreliable network.

References

[DoTH02] Olivier Dousse, Patrick Thiran and Martin Hasler. Connectivity in ad-hoc and hybrid networks. Technischer Bericht, 2002.

[t08a] CORSIM SIMULATOR. www.fhwa-tsis.com/corsim_page.htm.

[t08b] Qualnet SIMULATOR. http://www.scalable-networks.com/.

[WLHF+04] H. Wu, J. Lee, M. Hunter, R. Fujimoto and R. Guensler. Simulated Vehicle-to-Vehicle Message Propagation Efficiency on Atlanta's I-75 Corridor. Technischer Bericht, November 2004.

[WuFR03] Hao Wu, Richard Fujimoto and George Riley. Analytical Models for Information Propagation Vehicle-to-Vehicle Networks. Technischer Bericht, 2003.

List of Figures

Review of "Urban Multi-Hop Broadcast Protocol for Inter-Vehicle Communication Systems"

Ben Mayer

Abstract

The communication between vehicles which are outside their sending range makes it necessary to send messages over many repeater stations. These repeaters could be either stationary antennas or other vehicles. Besides you have to see that in urban areas it is very possible that high buildings shadow nodes which are in the transmission range and so you have to find a secure and stable way around them. This leeds to a new efficient protocol which is based on IEEE 802.11 and is called UMB (Urban Multi-Hop Broadcast Protocol). It wants to solve the problem of broadcast storm, hidden nodes and reliability problems in urban areas. This protocol divides the road portion inside the transmission range into sections, locates the furthest non-empty and takes then a vehicle inside. This vehicle gets now the duty to acknowledge and forward the messages of the sender. On intersections in the urban area there are repeaters which initiate new directional broadcasts if they are in the transmission range of the sender. How this protocol works in detail and why it seems to be very effective in comparison with other flooding based protocols I try to explain in the report.

1 Introduction

Not long ago the automotive industry became interested in Inter-Vehicle Communication Systems because they saw a chance to make their products safer, more comfortable and more entertaining, too. Broadcast is often used in such systems. Useful applications relying on broadcast could be the sharing of emergency, weather, traffic and road data, advertisement and announcements. When messages are send to locations which are beyond the transmission range then the multi-hopping method is used. (This can be compared with an airplane that wants to travel a long distance but it has not enough fuel to do it in one straight way. So it has to stop at several airports to get refilled and to restart the journey.) It is the fact that through packet loss, collisions and hidden nodes the receiver cannot get its message and beside this it is important that unnecessary retransmissions use a significant amount of the wireless resources. Hidden nodes in urban areas are for example vehicles at road segments which are shadowed by tall buildings and so cannot be reached by the sender. These obstacles make it important to carefully design a protocol which nearly solves the problems and make the traffic reliable and efficient.

Starting with a pure blind flooding without using any topology information you get redundancy, contention and collision problems. There exist techniques to improve blind flooding, but they are not effective for all ranges or node densities. And these two points change very quickly in inter-vehicle communication systems. The next step to fix the redundancy could be to include topology information to choose the nodes. But for high numbers of vehicles it becomes impractical because the speed of the protocol would decrease by more data which had to be processed. To get a solution for the hidden node and reliability problem you can

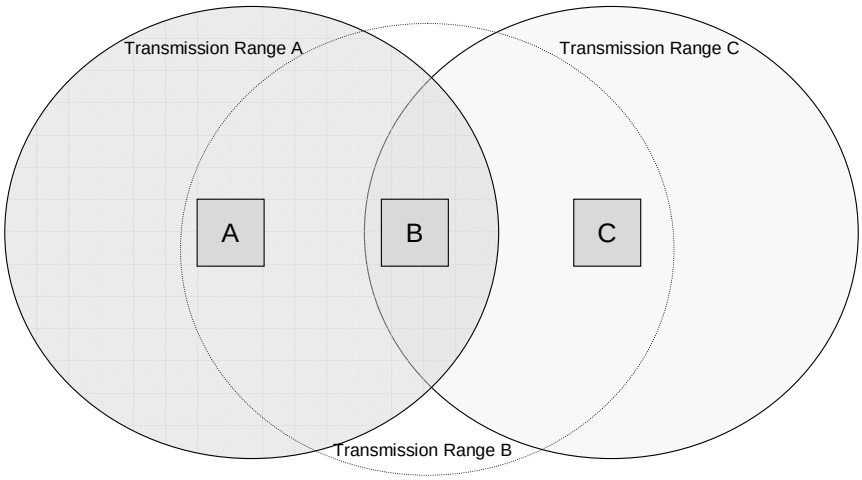


Figure 1: Shows the hidden node problem

introduce a handshake and acknowledgement mechanism like RTS / CTS (Ready To Send / Clear To Send). To make this clear look at Figure 1.

If a sender (A) wants to send a data packet to (B) it sends a short frame packet (RTS) with the information about the duration of the transmission. All participants which are in the sending direction get this packet and do not initiate sendings for a time based on this information. If B gets this frame packet it answers also with a frame packet (CTS) which also includes the information about the duration. Now the participants of the network which are not in the transmission range (for example C) of A but of B will be informed that a transmission will start soon. So they also do not send. If A gets the RTS packet it starts sending the data.

But as long there is more than one destination this could cause packet storms around the source. Now UMB tries to combine and expand the positive aspects of other protocol attempts. The UMB protocol consists of two parts

1. Directional Broadcast
2. Intersectional Broadcast

In the next part I try to explain the directional broadcast where sender nodes try to select the furthest node in the broadcast direction to assign the duty of forwarding and acknowledging without using any topology information. The intersectional broadcast is realised through repeaters at the intersections to forward packets to all road segments.

2 Protocol Description

Assumptions:

- Vehicles build an adhoc-network
- Simple repeaters at intersections

- Vehicles equipped with GPS receiver and road map

The most important goals of UMB are:

1. Avoiding collisions due to hidden nodes (with handshake mechanism)
2. Using the channel efficiently (with clever selection of packet forwarder)
3. Making the broadcast communication reliable (with ACK packet mechanism)
4. Sending messages in all directions at intersections (with simple repeaters)

3 Directional Broadcast

3.1 RTB/CTB Handshake

To avoid the hidden node problem there is a RTB (Request To Broadcast) / CTB (Clear To Broadcast) handshake implemented. The receiver of the sources RTB packet is the furthest away node among the nodes in the sending range. So all other nodes which are between source and receiver also receive the message and do not access the channel for a specified time in the packet. To select the furthest away node the road portion in the transmission range is divided into segments. These segments are only created in the direction of the dissemination. If there is more than one vehicle in the last segment then this segment is iteratively divided in smaller parts. If this method does not succeed then the remaining nodes in the last sub-segment enter a random phase. The protocol can adapt the iteration steps dependent to the traffic conditions. A RTB packet includes following

1. transmission duration
2. position of the source
3. broadcast direction

Source vehicle obey the IEEE 802.11 transmission rules (CSMA/CA) while attempting to send an RTB packet. That means that they monitor the channel before sending to avoid collisions.

When the vehicles on the road in the direction of the dissemination receive this packet they compute their distance to the source. They can do this because of the header information of the packet and knowing their position. Dependent on this distance,transmission range and number of segments they compute the length of a jamming signal which is called black-burst. Since two cars have not the same black-burst length it is now possible to select the furthest node. The vehicles send their black-burst after SIFS (Shortest Inter-Frame Space) which is the time to wait after hearing that the channel is empty. Now they change to listen-mode and listen to the channel. If when listening the channel is empty they recognize that they are the furthest node, because the furthest node sends the longest black-burst, which is computed as follows in the first iteration:

$$L_1 = \lfloor \frac{d}{Range} \times N_{max} \rfloor \times SlotTime \quad (1)$$

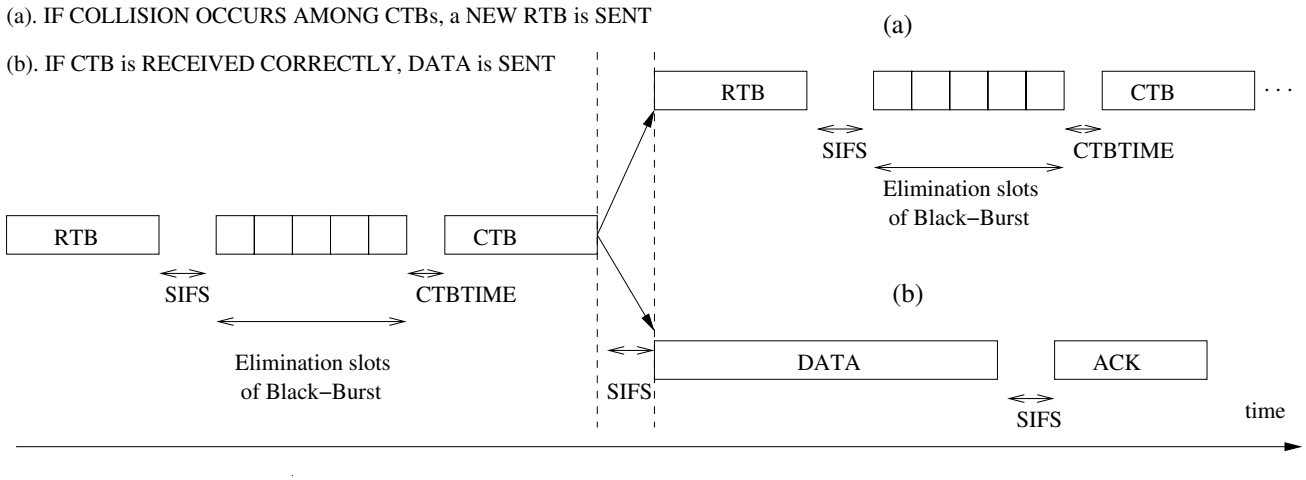


Figure 2: Sequence of packets. (a) Second RTB/CTB handshake (b) DATA/ACK.

d is distance from source to vehicle
Range is transmission range
N^{max} is the number of segments created
SlotTime is the length of one slot

After this selection this node can answer with a CTB packet. The other nodes which could hear the burst do not try to send a CTB packet. If there are many vehicles in one segment they will maybe all find the channel empty and all try to send a CTB. In that case the packets will collide though the source detects the transmission but cannot decode it and re-sends the RTB packet. But now only the nodes which sent the CTB packets join the next iteration, where the furthest non-empty segment is splitted into N^{max} sub-segments. This procedure repeats until the successful transmission of a CTB packet to the source or until D^{max} attempts. With this formula you can compute the black-burst length for the i-th iteration (L^i):

$$L_i = \lfloor \frac{d - Llongest_{i-1} \times W_{i-1}}{W_{i-1}} \times N_{max} \rfloor \times SlotTime \quad (2)$$

$$i = 2, 3, ..., D_max \quad (3)$$

$$W_i = \frac{Range}{N^i_{max}} \quad (4)$$

$Llongest^i$ is the longest blackburst before
W^i is the segment width in the i-th iteration
The source only indicates in its RTB packet that there was a collision before and the nodes which sent the longest black-bursts in the previous iteration are responsible to select and split the segment. Now if this segment based method does not come to a successful ending then the vehicles which sent a CTB in last iteration enter the random phase where they randomly choose a black-burst length from [0, Nmax - 1] slots. If there are collisions then the colliding nodes join the next phase and so on. This will only happen Ran^{max} times else the segment based method starts from beginning. In the most cases this should not happen. If the source gets no answer after sending the RTB packet it goes back to the first segment based iteration and tries again after a random waiting time. The maximum attempts are given through the RET^{max} parameter.

3.2 Transmission of DATA and ACK

After ending this handshake between source and corresponding node with receiving the CTB package the source node sends the broadcast packet. In this packet there are data and the ID of the node which has successfully sent the CTB. As said before this is the node who has to forward the packet and to acknowledge it by sending an ACK to the source. This ensures the reliability of the dissemination. Note that all nodes between also get the packets but do not forward or acknowledge. If the ACK packet gets lost before the source receives it then the source goes back to the first segment based iteration and starts again. But this can only happen $\mathrm{RET}^{\max}$ times.

4 Intersectional Broadcast

While disseminating the packet can be received by a repeater at an intersection and should be forwarded in all road directions except the direction where the packet came from. At intersections a repeater should always be the forwarder because of the best sight to the road. When a node is the forwarder and there is no repeater inside the transmission range then a normal directional broadcast is started as described before. But if the opportunity exists to send to a repeater the node sends it via the normal point-to-point IEEE 802.11 protocol. The positions of itself, repeaters and road segments are known by all nodes through their GPS systems and digital road map. This protocol uses RTS,CTS,DATA,ACK according to the UMB protocol. After receiving the broadcast the repeater starts the directional broadcasts in all other directions.

There could also appear situations where packets pass road segments many times. These loops should be avoided because of wasting band-width. There for the repeaters memorize the packet IDs they have forwarded and do not initiate broadcast for packets they already know. Another aspect is that bandwidth is lost because of nearby nodes which receive packets that repeaters forward more than once. Especially for long data this could waste significant bandwidth. So the UMB protocol has implemented a mechanism to decrease the waste: The repeater does not send the information in the DATA packet if the corresponding node has already overheared this packet. The corresponding node has only to set a bit in its CTB packet so the source can recognize that it has not to send the information. However the source has to send a short DATA packet to give the duty of forwarding to the corresponding node.

5 Performance testing

The developers of the UMB protocol have tested it in a special simulation environment. They have simulated the MAC Layer and Physical Layer with an event driven simulation library (CSIM) and the vehicle movement and road layout with a separate Matlab simulation. To get a good comparison to the UMB protocol they have modified the standard IEEE 802.11 Protocol to implement two other flooding based protocols. First there is 802.11-distance where collisions try to be avoided by computing the waiting time - before forwarding a packet - inversely proportional to the distance from the source. The waiting times are discrete because of being multiples of the Slottime. Second there is 802.11-random where the waiting time is randomly chosen between zero and the maximum Slottime. This two protocols have also in common that *every* node has the duty to forward the received packets. To test these three protocols on the road there are two road structures implemented, which you can see in the following graphic.

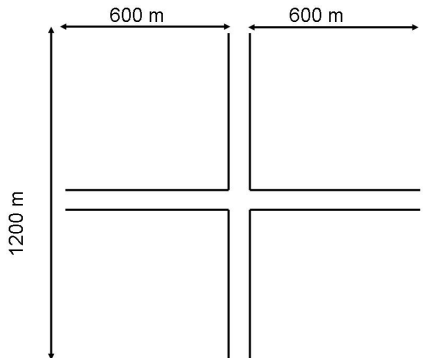


Figure 3: Road Structure I: one intersection, 1200 m x 1200 m

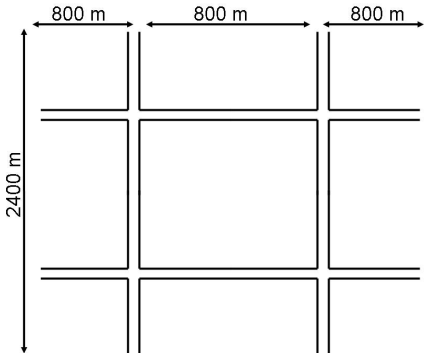


Figure 4: Road Structure II: four intersections, 2400 m x 2400m

A specified number of vehicles move on the road lanes with a constant speed. They do not overtake, turn around or stand still. The cars are randomly placed with exponentially distribution. Finally here are the simulation parameters the developers used to test their protocol:
transmission range = 400m
data rate = 1 Mbps
frame body = 2312 bytes
base protocol 802.11b
maxSlot = 32
simulation time = 60s
simulation repititions = 30
$RET^{max} = 15$
$N^{max} = 10$
$D^{max} = 2$
$Ran^{max} = 3$

Three metrics have been defined to compare the performance:

1. Success Percentage: Success Percentage of a packet is the number of cars which received this packet divided through all cars in the simulation
2. Packet Dissemination Speed (m/s): Speed of a packet. Distance travelled by the packet divided by the delay. The delay is the time elapsed since entering the source queue until the packet is received by another node.

3. Load Generated per Broadcast: The number of bits transmitted to send one packet to the whole network. Total number of bits divided through total number of generated packets. To consider the lost packets this value is also divided through the success percentage.

In the following graphics you can see the testing results for different arrangements.

5.1 Arrangement 1

- Intersections: 1
- Average Vehicle Density: 10 / lane
- Total vehicle number: 61

As you can see the UMB success percentage is constantly near by 100 percent. Only if the payload is high and the number of created packets increases some of them get lost. The two other protocols decrease very fast with increasing the packet number because they have no solution for the hidden node problem and their packets are not acknowledged by the receivers. The UMB normalized average load is very low that shows that the channel is efficiently used because of avoiding unnecessary multiple rebroadcasts. But the UMB protocol is slower in dissemination speed than the two others as long the payload (DATA) is comparable to RTB, CTB, and ACK overhead. If the DATA packet becomes longer the difference decreases to a minimum since the time used for the handshaking is little compared to the time used for sending the data.

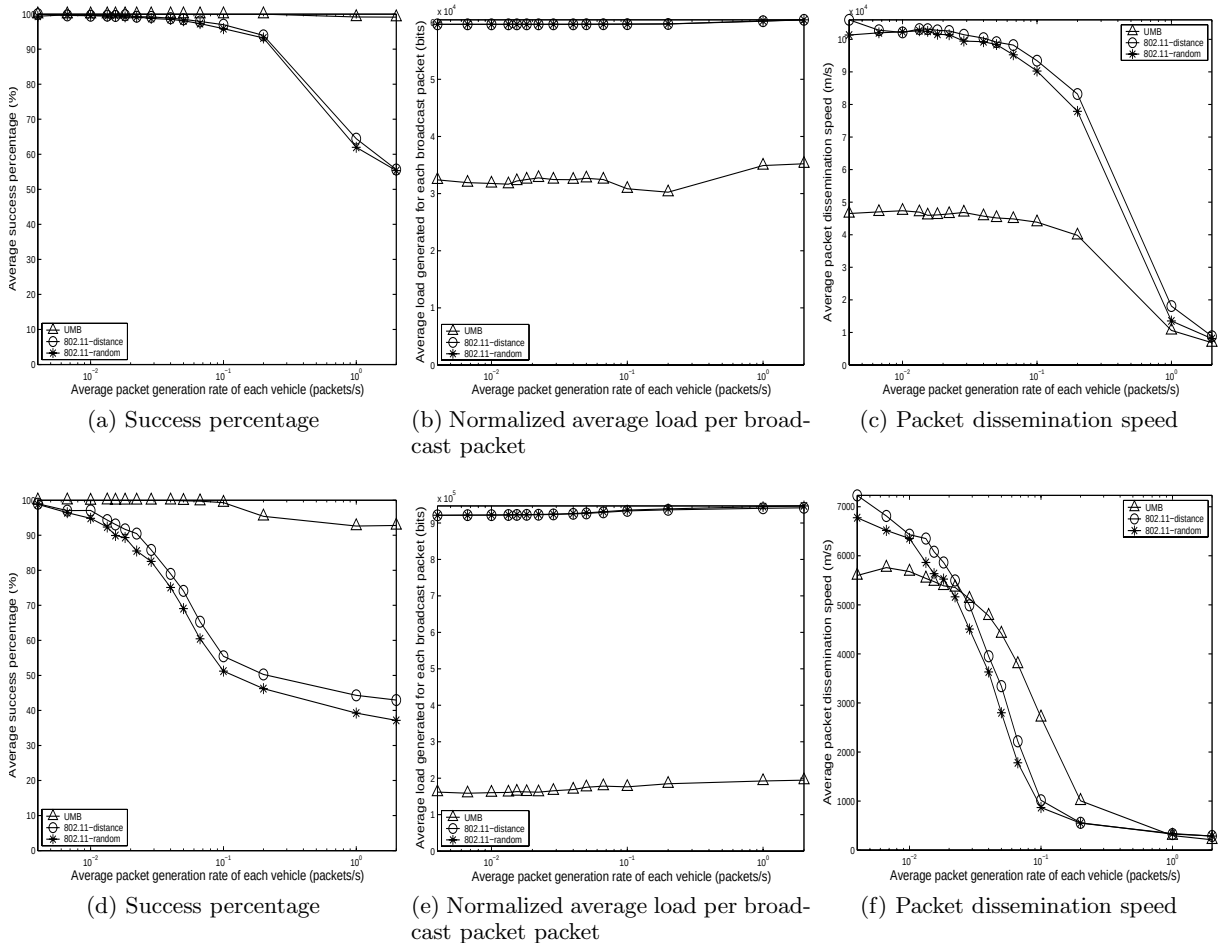
5.2 Arrangement 2

- Intersections: 1
- Average Vehicle Density: 33.3 / lane
- Total vehicle number: 160

When more vehicles are simulated you can observe that 802.11-randoms' and 802.11-distances' performance earlier decreases. The hidden node problem and the lack of acknowledgement become more and more important. UMB makes no significant changes. The UMB normalized average load keeps constant on a low good level because the forwarding duty is only at one vehicle and not at every car as implemented by the two other protocols, which obviously get a weaker performance because more cars have to re-broadcast more packets. Dissemination speed decreases when packet generation increases. That seems to be because the packets remain longer in the source queue and the sending rate is the bottle neck. If DATA is long the worse performance of UMB gets much better and is comparable to the others. This increase of the speed is a again a result of the comparison of the small overhead in comparison to the transferred information.

5.3 Arrangement 3

- Intersections: 4
- Average Vehicle Density: 10 / lane



(a) Success percentage (b) Normalized average load per broadcast packet (c) Packet dissemination speed

(d) Success percentage (e) Normalized average load per broadcast packet packet (f) Packet dissemination speed

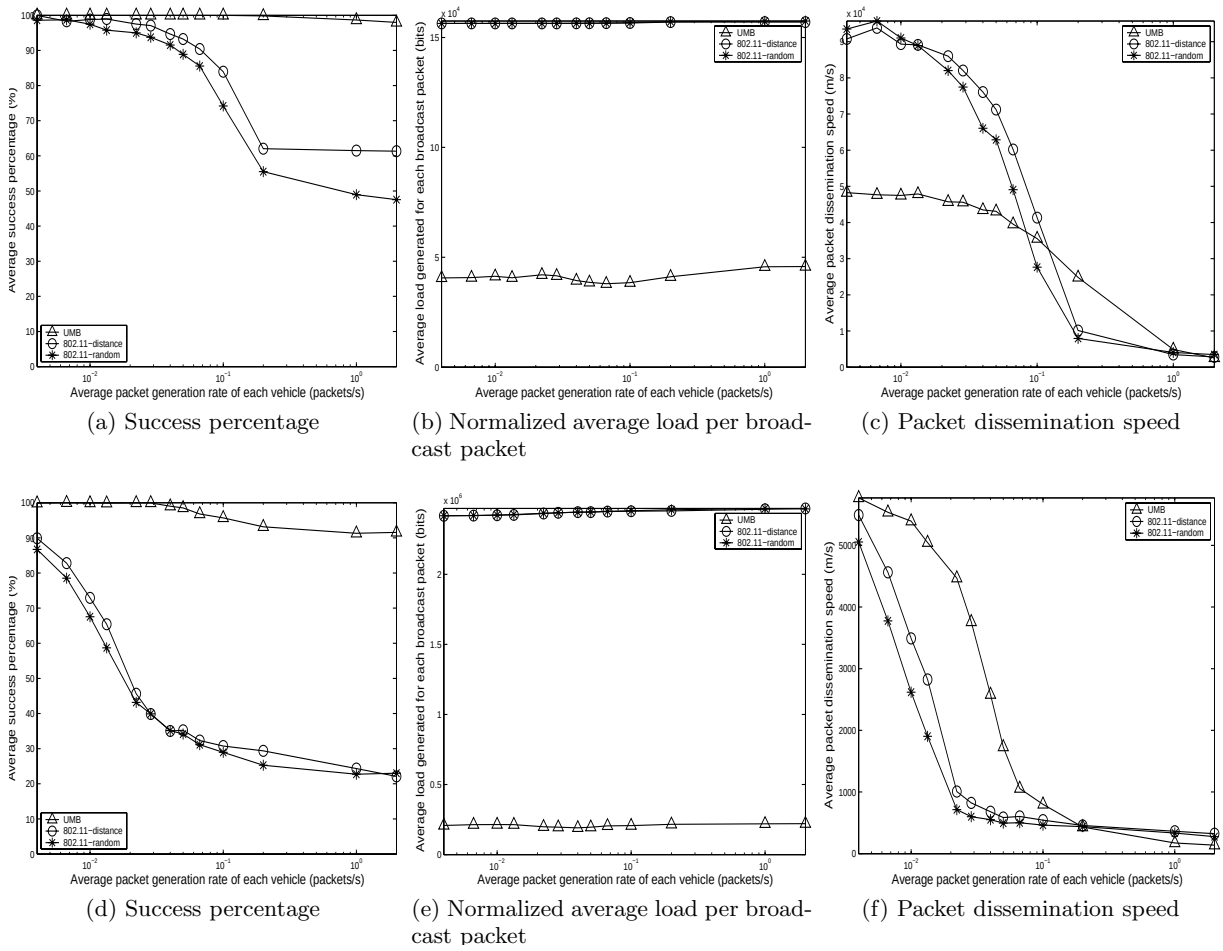
Figure 5: One intersection. Average vehicle density= 10 veh/km per lane. In (a), (b), (c) payload is 100 bytes and in (d), (e), (f) the payload is 2312 bytes

- Total vehicle number: 190

Now in the third arrangement packet loops could appear so the repeaters are equipped with a caching unit for the packet IDs. But because the effect of loops is avoided in the simulation there is obviously no unexpected result of the new arrangement. Because the road structure is bigger than in the one-intersection arrangements there are more cars on the street and so more packets are produced and the normalized average load increases. That always decreases the flooding based protocols' performance and the success percentage decreases because of more collisions and packet loss. The dissemination speed behaves as in the one-intersection arrangements.

5.4 Arrangement 4

- Intersections: 4
- Average Vehicle Density: 33.3 / lane



(a) Success percentage (b) Normalized average load per broadcast packet (c) Packet dissemination speed

(d) Success percentage (e) Normalized average load per broadcast packet (f) Packet dissemination speed

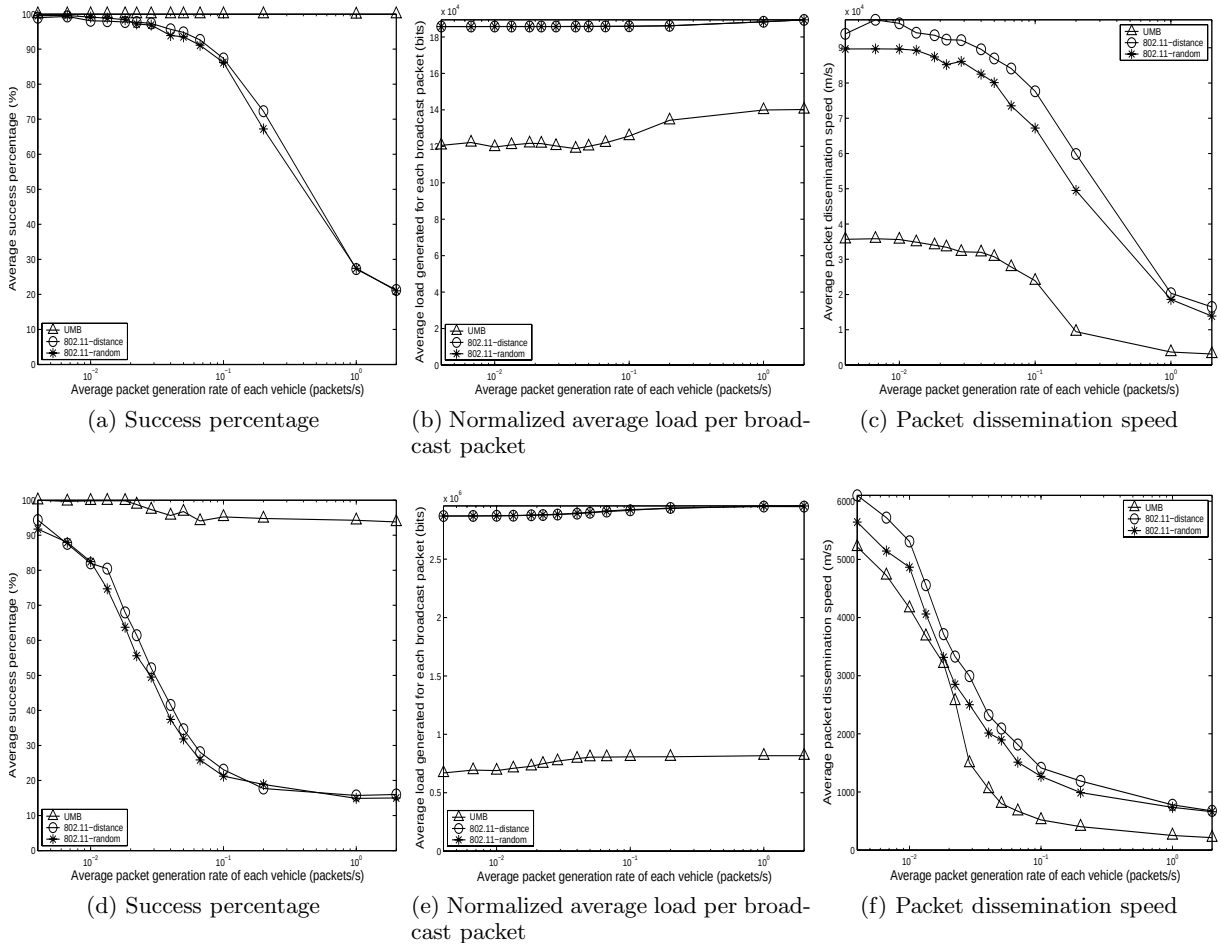
Figure 6: One intersection. Average vehicle density= 33.3 veh/km per lane and payload=100 bytes. In (a), (b), (c) payload is 100 bytes and in (d), (e), (f) the payload is 2312 bytes

- Total vehicle number: 619

Because more vehicles produce more packets the success percentage decreases by all three protocols. But the flooding based protocols are more affected because of the lack of control mechanism as handshaking and acknowledging. And this leads to packet loss and collisions. The normalized average load shows the same affection caused by the increasing vehicle density. At the packet dissemination speed you can see that for large DATA as seen before the RTB/CTB/ACK-overhead nearly disappears in comparison to the DATA transmission.

6 Conclusion and opinion of the author about the source text

As seen in the last diagrams the UMB protocol has obviously not to fear a comparison with flooding based protocols. In nearly all cases it is superior and shows a constant good performance. Increasing the vehicle density and number of intersections only has moderate effects. And because of giving the duty of forwarding to only one vehicle the protocol uses the bandwidth very efficient. But there is an aspect which could speak against the use of this protocol. First you have to install repeaters at every intersection which could be an economic problem because it is expensive. The protocol should be extended through a version which does not need the repeaters and as far as I read this work has already started. On the other



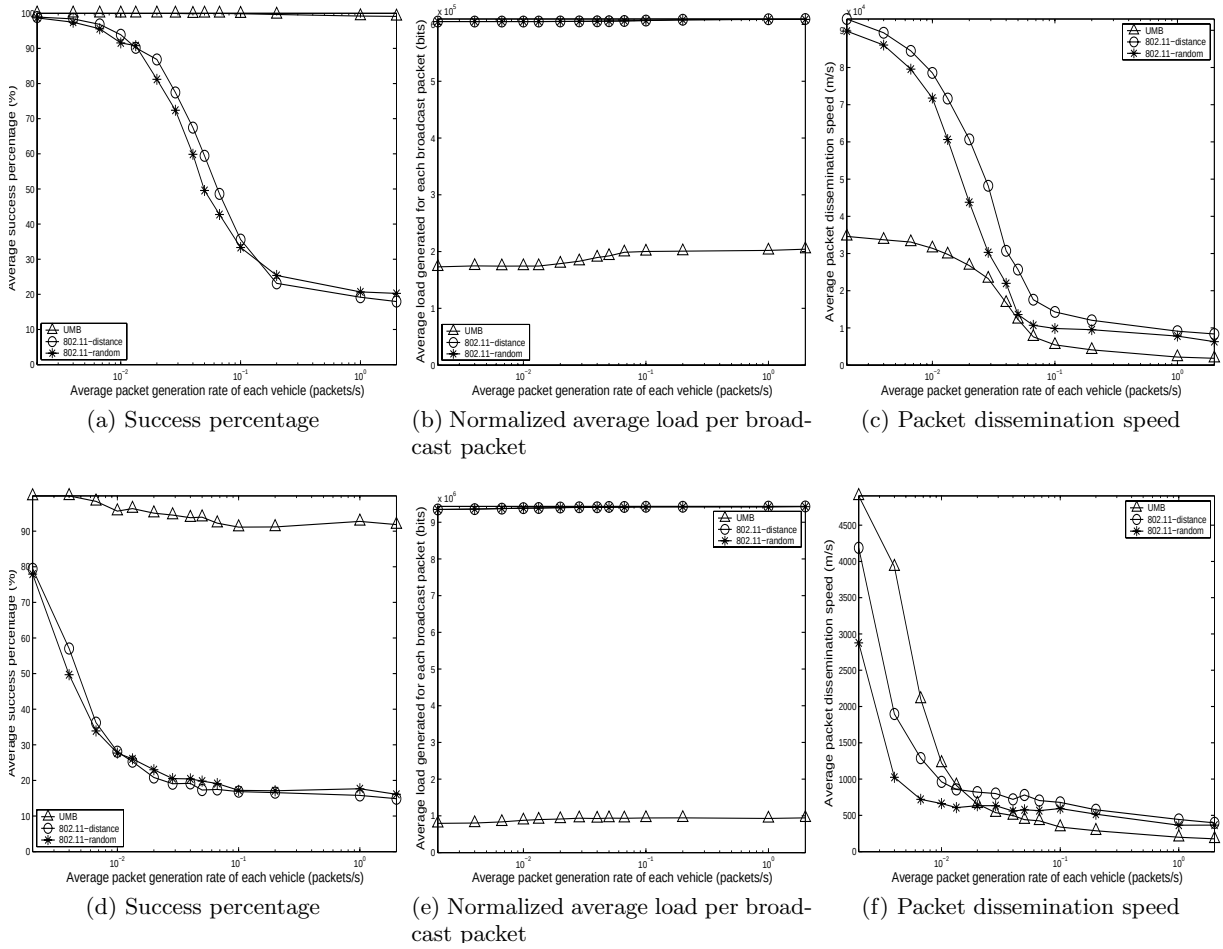
(a) Success percentage (b) Normalized average load per broadcast packet (c) Packet dissemination speed

(d) Success percentage (e) Normalized average load per broadcast packet (f) Packet dissemination speed

Figure 7: Four intersections. Average vehicle density= 10 veh/km per lane and payload=100 bytes. In (a), (b), (c) payload is 100 bytes and in (d), (e), (f) the payload is 2312 bytes

hand that would make it heavy to get such good performance and I am not convinced whether the protocol would be reliable for complex road architecture. This should be tested in more complex arrangements. Another aspect is that through the random parts of the protocol and the many possible restarts it is nearly impossible to guarantee quality of service. What I liked about my source text was the clear structure of the information and the scaling of details. But in my opinion the whole thing could be explained on half of the pages. The charts are too small and the differences cannot easily figured out. Anyway I had to use it because I had no data to produce the charts. Beside of that the charts could hardly be compared because of the arrangement on the paper. It is a pitty that they do not show the effect of loops in their arrangements but only the absence. That makes the third and fourth arrangement pritty senceless. They show very often and in detail the better performance of UMB by explaining the lines in the diagrams. But for example why the dissemination speed decreases by increasing the packet generation rate they do not explain. All in all they give a good introduction in their protocol but keep some secrets.

List of Figures



(a) Success percentage (b) Normalized average load per broadcast packet (c) Packet dissemination speed

(d) Success percentage (e) Normalized average load per broadcast packet (f) Packet dissemination speed

Figure 8: Four intersections. Average vehicle density= 33.3 veh/km per lane, In (a), (b), (c) payload is 100 bytes and in (d), (e), (f) the payload is 2312 bytes

Review of "Opportunistic Routing in Multi-Hop Wireless Networks"

Michael Bucko

Abstract

A new unicast routing technique for multi-hop wireless networks, called Extremely Opportunistic Routing (ExOR), is being illustrated in this paper. How can this technique advantageously be used in vehicular ad-hoc networks and what are possible weaknesses on the other hand?

1 Introduction

After a short introduction into unicast routing protocols, the differences between wired and wireless networks and an overview of the IEEE standard, a new unicast routing technique called Extremely Opportunistic Routing (ExOR) will be presented.

When a sender broadcasts a message, every receiver can listen and understand the sent message. Talking about unicast routing technique in wireless means that the sender broadcasts a message, and so every participant of the network can receive the packet, but the only one who is interested in, who wants to know what or who is able to decode the message is the addressed recipient.

Nearly every unicast routing protocol first chooses a path of nodes in the network in order to send the data through this pre-determined path. These techniques make a lot of sense if the underlying network is based on wired connections or each pair of nodes is only able to exchange data when nodes are linked. In a multi-hop network there always is at least one optimal route as long as every link has deterministic costs. A routing protocol is expected to perform well whilst it uses the optimal route for sending its data.

Using this technology seems to be interesting in wireless networks too, but there are a few major differences to care about between a wireless and a wired network. First of all, in a wireless network every node can directly communicate over radio with another node, maybe with a higher error rate. Another deviation is that all packets sent through radio are broadcasted and therefore there is no need to target a specific node. Last, radio communication between nodes is not deterministic; so packets sometimes do not arrive. In wired networks properties making this pre-determined route performing well, may not hold in wireless.

With the possibility of sending and receiving data in Local Area Networks (LANs) without any wires, speaking about wireless LAN, a new standard was published by the Institute for Electrical and Electronic Engineers (IEEE) in 1997. The IEEE 802.11 (WLAN) standard was added next to the existing wired network standards, i.e. IEEE 802.3 (Ethernet), IEEE 802.4 (Token-Bus) and the IEEE 802.5 (Token-Ring). The WLAN standard has of course been added and improved since its introduction. An 802.11 network is a system with a cellular architecture that is subdivided into cells. The simplest scenario is the (mobile) ad-hoc network functioning without any additional fixed infrastructure and where at least one

computer builds the self-organizing wireless network. Adding (fixed) access points (APs), which are usually connected with a distribution system (DS, in most of the cases an Ethernet), an interconnected WLAN, in the Standard named as Extended Service Set (ESS), is set up. A few ideas of how VANETs can benefit from ExOR and/or combinations are explained and touched in 4.

2 The IEEE 802.11 Standard

In order to understand why slight changes in the Media Access Control layer in Extremely Opportunistic Routing are necessary, later in 3.3, the 802.11 Standard is quickly refreshed. The IEEE 802.11 protocol covers the Media Access Control (MAC) and the physical layer (PHY), as any other 802.x protocol does. The Standard defines a single MAC which interacts with, currently, four PHYs where all of them are running at one and two Mega bits per second, two of them also at 5.5 up to 54 Mbps.
As the physical layer is exactly equivalent to the layer one (PHY layer) in the OSI-Model the MAC only partly covers the OSI-layer two - the data link layer.

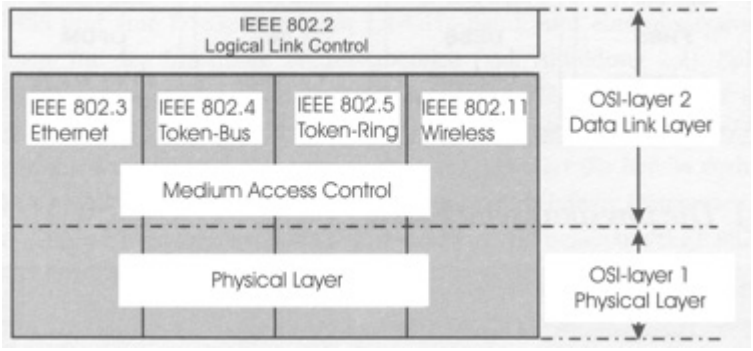


Figure 1: 802.11 in the OSI-model.

2.1 The physical layer

The wireless standard defines four PHYs however all have an interface to a uniform MAC layer. The PHYs Frequency Hopping Spread Spectrum (FHSS) and Direct Sequence Spread Spectrum (DSSS) both interact in the 2.4 GHz band and the third in infrared. In 1999 the Standard was extended by a PHY in the 5 GHz band called Orthogonal Frequency Division Multiplexing (OFDM).
The physical layer of wireless LANs has to ensure that users can operate neatly and smoothly as possible within the medium. Transmitting a bit correct and safe over the wireless from a sender to a recipient is the main task of the PHY. So in this layer the specific transmission medium and the modulation method are chosen to provide this task.

In order to supply an undisturbed transmission in an interference-prone medium the Spread Spectrum Technology uses a much wider frequency range than really needed. Due to spreading the signal with a code interference is reduced to a minimum.
Then, a bit is modulated on the electromagnetic wave where a harmonic wave is changed as long as the transmission of a bit lasts in its amplitude, frequency or phase.
The third task is to regulate the access on the medium. Three mechanisms, Time Division Multiple Access, Frequency Division Multiple Access and the Code Division Multiple Access, help to solve the multiple-access problem.

In the architecture directly under the MAC layer, the Physical Layer Convergence Procedure (PLCP) sub-layer communicates with the MAC, administrates the Physical Medium Dependent (PMD) sub-layer and extends frames received with important information for the physical sending. The PMQ sub-layer realises transmission and reception over the wireless medium with the respective technology and modulates or demodulates the frames.

2.2 The MAC layer

The MAC layer regulates the transmission of a whole message. Additionally to the key task, the regulation of collision-free access to the medium, the 802.11 MAC performs other actions that are usually done in upper layer protocols beyond the standard MAC protocols; such as fragmentation, packet retransmission and acknowledgments.

Within MAC, two basis access methods exist, following a centralized and a decentralized idea. The method based on the decentralized idea uses the Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA). In contrary to the other 802.x standards wireless cannot provide collision detection, because of wireless characteristics. Implementing a collision detection mechanism would need the implementation of full duplex radio. Since nodes cannot detect collisions during transmission and distinguish collisions from other disturbances, collisions have to be avoided in advance. A timer, called the Network Allocation Vector (NAV) regulates Collision Avoidance.

A station wanting to transmit data senses the medium. If the medium is busy, the station will delay its transmission, on the other hand if the medium is sensed free, the station has the allowance to send its data (Carrier Sense). Since several nodes simultaneously try to access the medium when they sense it free, collisions may appear where all sent data get lost (Multiple Access). The function of collision avoidance bases on the NAV-value. Every station stores the NAV in a special register and builds it on the Duration/ID field in the packet header, see Figure 3. All stations receive this information due to broadcasting, so they know how long transmissions will last and only try to send data when the NAV-value runs out. Collisions and disturbances can still turn up and therefore successful receipts of data will be communicated to the sender through acknowledgements (ACKs). If ACKs do not come, the sender tries to send the data again after a specific timeout. Stations do only acknowledge unicast packets since it would not make any sense doing the same with broadcast packets. The 802.11 standard distinguishes between data frames, for data transmission between users, control frames, for regulating the medium access, and management frames, such as acknowledgments. Management frames have a higher priority than the others do, because a station that sent some data needs an ACK within a certain time. After a transmission, stations wait an appropriate interval between two following frames, before they try to reserve the medium. Ensuring that management frames really have a higher priority than e.g. data frames, different intervals, Interframe Space (IFS), exist. The most important is the Short IFS (SIFS) and defines a minimum interval for ACKs and other management frames. Later in the 3.2.2 we see that this is a very important interval that has to be changed in order to implement this new routing protocol. If the SIFS passes by and the Distributed IFS (DIFS) expires as well, a node can now gain access to the medium in order to sending its data after a short random back off.

Another situation for collision that may occur is called the hidden state problem (Figure 2. As we already know, when one node occupies the medium, all other nodes in the range of the occupying node have to wait. The hidden state problem is a mistaken recognition of free medium although someone else already occupies the medium. This situation happens

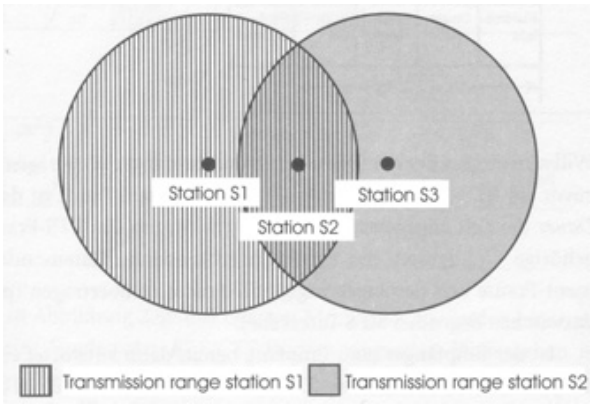


Figure 2: The hidden state problem.

when a node C does not lie in the same range as node A and vice versa. Node B connects both of them. Let us assume that A wants to send a message to B, because the medium is sensed free and all other collision avoidance criteria are fulfilled. However, in truth C tries or already transmits data to B, as a consequence medium is not free and a collision occurs at B. Therefore, A sends a request to B, where A is granting access for the medium for a specific duration. B hears this Request-To-Send (RTS) frame and sends back an allowance if medium is free and B is ready. This Clear-To-Send (CTS) frame again contains the duration so every other node will hear and update its NAV.

The most general MAC layer frame format in an ad-hoc network and a frame format for an acknowledgment are shown in Figure 3.

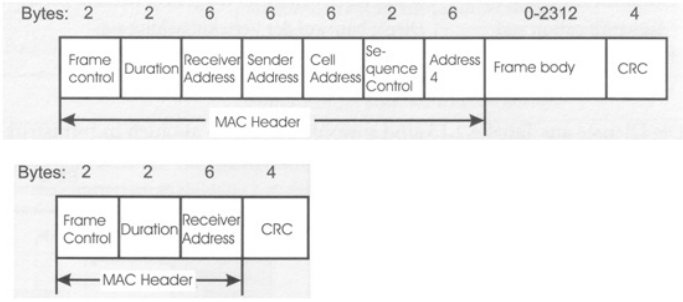


Figure 3: Data and Acknowledgment frame format in the MAC layer.

3 Contribution of the Paper: The ExOR Routing Technique

Extremely Opportunistic Routing is a unicast routing technique for multi-hop wireless networks which reduces the quantity of transmissions and is a protocol that wants to benefit from the wireless' characteristics. Each packet is forwarded through a sequence of nodes, where the ExOR determines the next node closest to destination forwarding the packet from all the nodes that have successfully received the packet. This behaviour skips numerous hops in multi-hop path which finally reduces the number of transmissions up to 65 % compared to the best possible pre-determined path routing. Important to realize is that ExOR does not predetermine the path ahead of sending the data, but during the sending process.

The most important ambition of ExOR is to ensure that only one receiver forwards the packet and that it is the receiver closest to destination.

3.1 The Protocol

The ExOR protocol is split into three stages. Each node in the network has a loss rate matrix for direct radio transmission between every pair of nodes. This matrix is updated periodically and generally built with a link-state flooding scheme.
The three stages are repeated so many times until the packet reaches its final destination.

3.2 The Protocol

The ExOR protocol is split into three stages. Each node in the network has a loss rate matrix for direct radio transmission between every pair of nodes. This matrix is updated periodically and generally built with a link-state flooding scheme.
The three stages are repeated so many times until the packet reaches its final destination.

3.2.1 Stage 1: Selection of the candidate forwarder set

The achieved, good performance of ExOR resides on the ability choosing a prioritized candidate set of nodes which can transport the packet closer to its destination. Selecting the forwarding candidates is based on the shortest number of hops and longest hops. The sender first looks for the shortest way to the destination node where equal paths have to compete with their delivery ratio. The highest priority in the candidate set receives the node that is the closest to destination. The sender writes this list in the package header - once again, the candidates are prioritized by distance. As soon as the first node is set, the ExOR deletes this

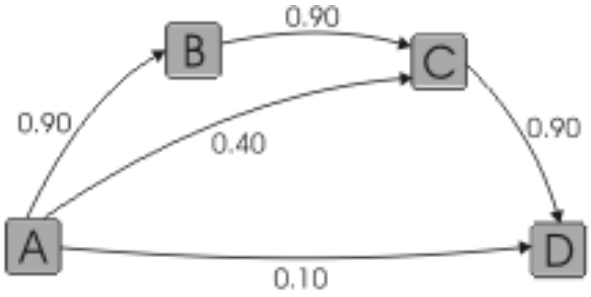


Figure 4: A simple network example, with delivery ratios.

highest priority candidate from the loss rate matrix and searches for the shortest path again. Then the first hop from this new shortest path is the candidate with the second priority. This procedure is repeated until all remaining positions in the candidate set are fixed. The candidate set is cached until the delivery ratio matrix is updated with its new values. Especially at changes of the topology, when nodes join or disappear in the network, the measures of loss rates and the flood statistics are updated.
A simple example is shown in Figure 4 where the sender A wants to transmit a packet to C. In this example, the candidate set size is three. D is the highest priority node due to its longest hop. Node A deletes node D from its delivery loss rate matrix and repeats the procedure. Then, A chooses C as the most important node in this new shortest path, so C receives second highest priority looking at all nodes. At the end, node B gets weakest priority. Finally, the candidate set is [D, C, B].

3.2.2 Stage 2: Acknowledging transmissions

Remember that one of the most important goals in ExOR is that only one receiver should forward the packet. But how can it be guaranteed and how do all nodes agree who is the next sender of the packet? Due to the fact that not every node from the candidate set receives the packet from the original sender, there must be a way that every node knows if he is the lucky one to send it. The solution is that a node broadcasts an acknowledgment (ACK) for successfully receiving the packet. As seen before the MAC in the 802.11 standard only provides a simple acknowledgment that replies: "'Yes I got it."' This simple ACK is not sufficient for choosing the forwarder; therefore the MAC from 802.11 must be modified. An acknowledgment must contain more information and the modified MAC has to reserve multiple slots of time to return an ACK. Why is it? Time between a SIFS and the next longer IFS is only long enough for returning one ACK and afterwards other data can be sent. However, we have a candidate set size of n members and all of them have to acknowledge the receipt of the packet. Consequently, to provide the possibility that n members are able to reply and no collisions are provoked, the answer interval has to be stretched. Therefore, we need something more than n times the ACK-transmitting time plus n times the time of a SIFS for guaranteeing the whole ACK sequence.
According to its position in the candidate set, the candidate delays the transmission of the ACK. So a node X sends an ACK with the ID of the highest priority successful recipient Y known to X. Before a candidate forwards the packet, he is listening to all ACK slots to know if there is a candidate that has a higher priority and therefore will be the forwarder. The fact that every member of the candidate set sends an ACK containing the ID of the highest-priority ACK sender heard until he had to send the ACK suppresses multiple forwarding. Now one might think: "'What happens if a candidate X with a low priority does not hear each or especially higher prioritized ACKs?"'
A further example illustrates that it does not matter. Assume that node A reads in a received packet that it is the highest-priority nominee. Recipient A immediately sends an ACK with its ID in it. B, the one with the second highest priority, does not get A's ACK, but a further candidate C receives the acknowledgment from A. After C's time slot B receives the ACK from C. If there was no modified MAC, B would now decide to be the forwarder and thus A and B would forward the packet. But luckily our ACKs do contain IDs; B indirectly knows through C's ACK that A also got the transmission.

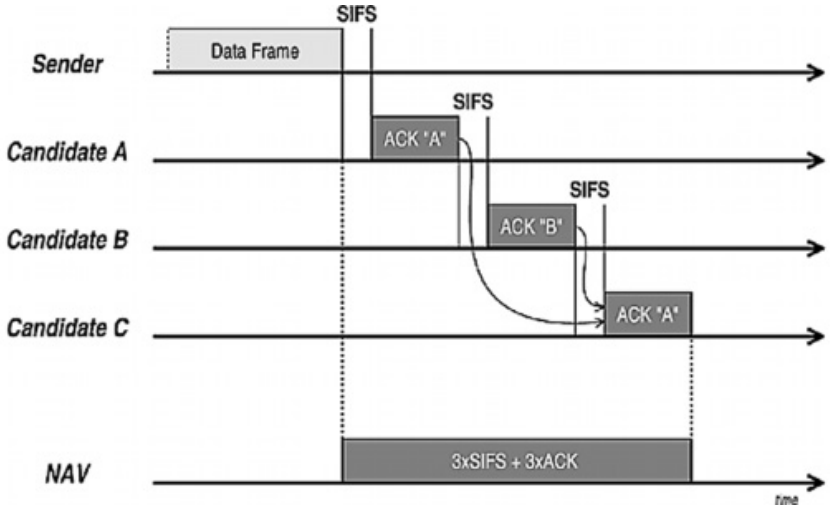


Figure 5: ExOR acknowledging sequence.

3.2.3 Stage 3: Decision of forwarding a received packet

Finally when the time slots for acknowledging the reception end up, each candidate locally decides with its information whether to forward the packet or sort it out. When the node decides to forward the packet because it has only received ACKs with lower priorities, it replaces the candidate set in the ExOR frame header with its new one. And finally transmits the packet.
Even though we have such a good protocol, sometimes a packet is transmitted more than once owing to acknowledgment reception failure. Each packet therefore contains a random nonce which the forwarding nodes store in their cache. Transmission is clear if the nonce is not stored in the nodes cache. As a result, a node will never send a packet multiple times as long it is not an explicit retransmission due to ACK failure.

3.3 Changes needed in the MAC Protocol

As already touched on, the Media Access Control from the 802.11 standard must be changed in a few details.
The data and the acknowledgment frame format have to be modified in the context of the Physical Layer Convergence Protocol (PLCP) frame. The frame control field has to indicate

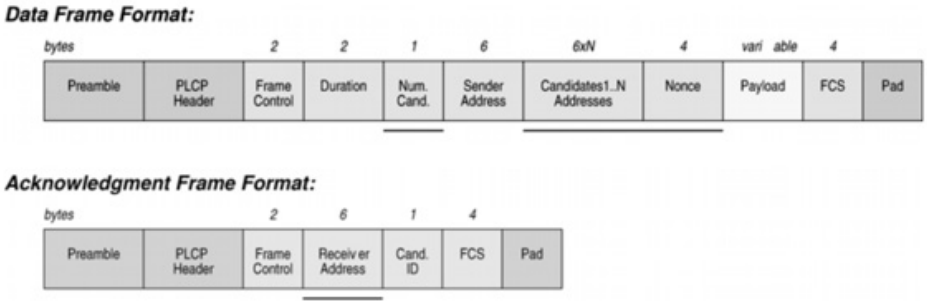


Figure 6: Modified MAC frame formats in ExOR.

that in spite of a standard an ExOR frame format follows; concretely a data or acknowledgment frame format. The two duration bytes in the data frame format will also allocate the time needed for multiple acknowledgments in the Network Allocation Vector. The time indicated in the Duration field, which every node in the network stores in the NAV is as much longer than the plain data transmission time as to provide to all forwarding candidates time enough to send their ACK. With adding the ACK sequence to data transmission time, explained in 3.2.2, prevents that other nodes lying in the same radio radius will not try to reserve the wireless medium for themselves too early.
Further the data frame format provides information about the candidate set, which is listed by priority. The packet's nonce with 4 bytes differs as well from the standard.
Due to short reply time slots, the acknowledgment frame format cannot be too big and thus only contains minimal changes that have the desired effect as described in section 3.2.2. One is the sender address of the data frame format. The second change is the position of the node in the candidate set belonging to the highest priority ACK heard so far.

The ExOR produces an overhead to the 802.11 Standard, but not a significant one and is cut out by far by its improvement.
Time of an ExOR data frame and acknowledgments is 8.3 % longer than unicast 802.11. The

physical layer operates at i.e. 6 Mbps with a SIFS time of 16 μs, preamble and PLCP header duration of 20 μs and 4 μs. The payload is a 1500 byte frame and the data frame has four candidates. The table in Figure 7 compares the transmission times.

	Data Frame	ACK(s)	Total
802.11 unicast	2085 µs	58 µs	2143 µs
ExOR w/4 candidates	2092 µs	228 µs	2320 µs

Figure 7: Transmission times for each type of frame and in total.

3.4 Evaluation and Simulation Results

Biswas and Morris, the authors of the paper [MoBi03], developed a simulation environment that models a multi-hop wireless network and its operating. The goal was to model a large, in respect of the diameter, network for getting a better impression of ExOR compared to the best possible pre-determined routing protocol. In this simulation environment they had the exact information about the delivery ratio and randomly placed 100 nodes in a plane seized 50x50 sq ft, which is about 15x15 square meters. Delivery ratio between nodes base on measures of the distance-to-delivery relationship in the work of *Ganesan et al.*, [WCEG+02] using medium transmission power. The distribution of packet reception rate over distance is non-uniform. The distribution of packet delivery rate over distance shows that probability of packet reception decreases almost linearly over distance rather than inverse-square, see Figure 8. In other situations or scenarios there can of course be other distributions, which might show a much bigger decrease than inverse-square. These details shall be enough for the moment as it goes too far in other topics - not used in this context here and import to understand now. Looking at the resulting distribution shows that ExOR performs well and

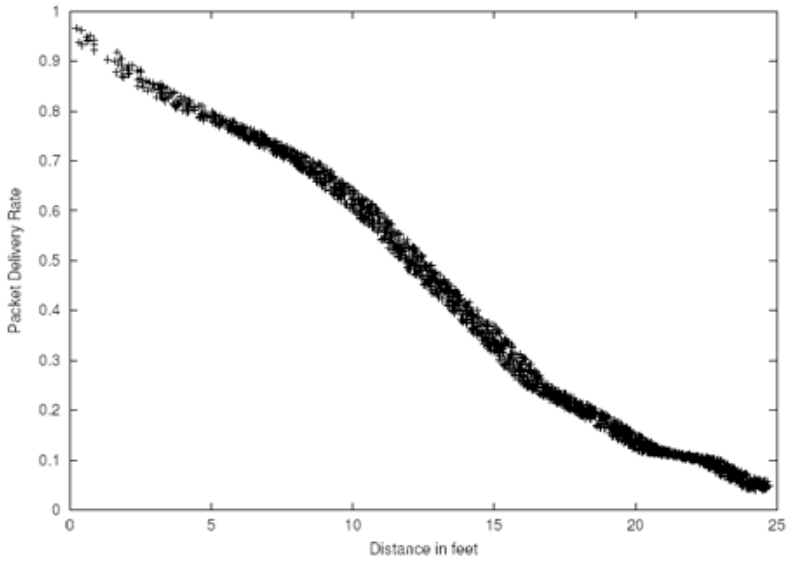


Figure 8: Delivery ratio vs distance using medium power (and in simulation).

even does when an inverse-square approximation is chosen. In cases where the falloff will be too big, the benefit of ExOR diminishes a lot due to the lack of long-distance links.

They assume all nodes are stationary during the simulation and therefore the delivery-ratio matrix (DRM) remains unchanged from the beginning to the end.
Leaving any radio propagation effects and packet-size-correlated losses aside - random losses of packets are only based on the delivery ratio. A packet is transmitted maximally 8 times. The rival of the ExOR path is the best possible pre-determined path, both do perfectly know

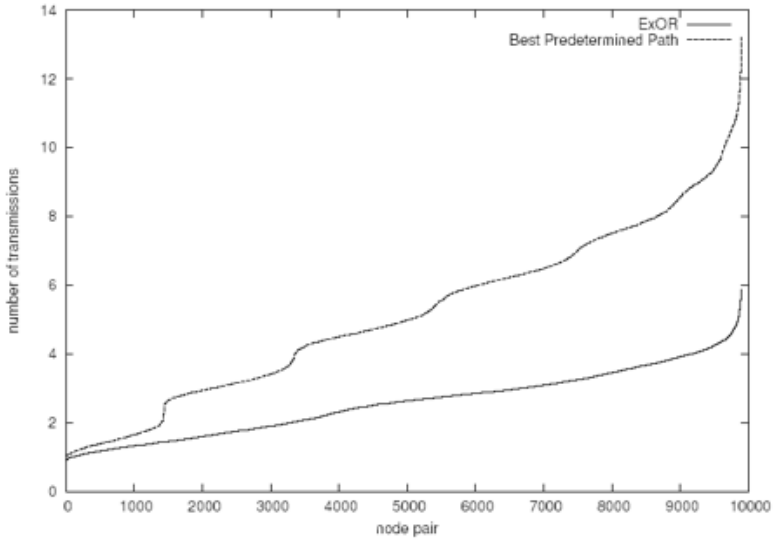


Figure 9: Number of transmission averaged over all pairs of nodes required to route a packet from source to destination.

the DRM and are both supplied with the identical DRM. For comparing ExOR with the best possible pre-determined path, they ran Dijkstra's shortest path algorithm in order to determine the total number of transmissions required on average to sending and acknowledging packets along a given route.
Biswas and Morris were focusing on the total number of transmissions between all pairs of nodes and the distribution of links used by both approaches.
In Figure 9 the average number of transmissions required to route 100 packets between every pair of nodes in the system, where ExOR uses a candidate set of eight members.
ExOR habitually performs much better and this up to 55 %. In even longer networks improvement up to 65 % are possible. This confirms that opportunistic routing really skips most short and intermediate hops. As in Figure 10 presented, the next survey is the distribution of successful transmission distances. This experiment contained 50 node pairs where we are looking at the five nodes farthest on either side of the network. Excluding short distances enables the plot to show us that ExOR forwards packets much farther than the best possible pre-determined route protocol.

4 Assessment and Conclusion

Event though ExOR is only simulated, Biswas and Morris plan to implement ExOR on hardware and are still negotiating for getting access to 802.11 MAC layer firmware, ExOR seems to work really efficiently. The very short simulation and evaluation section lets shining through that it is likely to perform well in real wireless networks and not only in simulation. Nevertheless, the fact ExOR simulation and experimentation were not caring about packet size rise a uncertainty about performance later on in real action as well as they were just

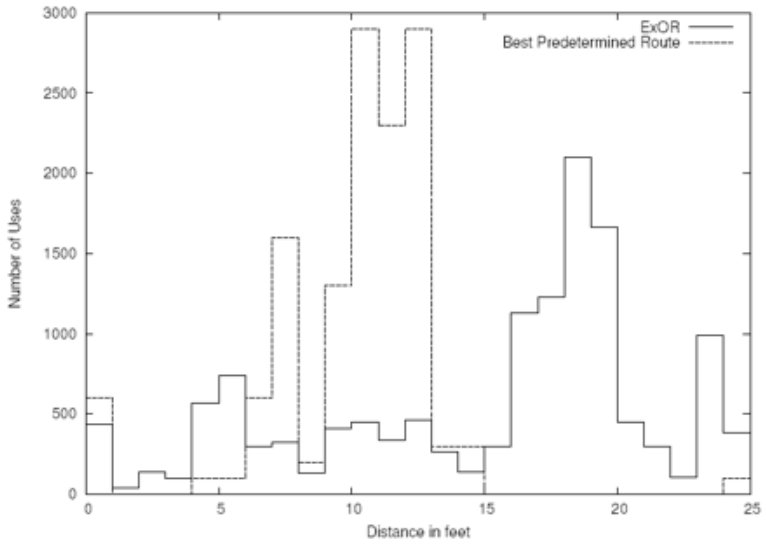


Figure 10: Histogramm of transmission distances using ExOR and the best possible predetermined routes.

simulating a plane of 50x50 sq ft. Criticising or even rating the paper is very difficult because the very important - evaluation, simulation and related work, are kept very short and information is put too concisely.

ExOR is definitely not a custom made protocol for VANETs, but its strengths are really worth considering implementing this unicast routing protocol in cars. As explained in foregoing sections the reduction of the number of transmissions for communication between the source and destination is an important improvement. Especially because in VANETs longer distances has to be served and bandwidth decreases as distance-supply increases.
Power supply is a criterion too, but in cars high transmission power is realizable. VANETs would really profit from the fact that packets are not transmitted multiple times any more. Those multiple sent packets can be used for sending other data instead. Forwarding acknowledgments could also ensure that the aimed car received the packet and show, which other cars did as well. If the acknowledgement reaches the original sender with all candidates stored in it, it would never trigger a new transmission of this information to one of those candidates again. This could be interesting in a candidate set - where every candidate needs the information, i.e. icy roads, and so the protocol almost acts like a multicast routing protocol. Nevertheless a standalone ExOR would not be a very good solution. Combination with a Global Positioning System (GPS) or with something like a braided multi-path routing could improve its reliability. Knowledge about the topology of the network would help a lot and could even increase reliability of transmissions up to its physical limits. Combining ExOR with the perfect knowledge of the geographic topology might enhance the chance to find real good forwarding candidates. Access points or an on-board computing unit with an integrated GPS could supply this geographic information.
Why improving reliability when reliability is already excitingly good? Firstly, cars are changing the position in their "'plane"´ and are not standing still as Biswas and Morris assumed in their simulation. Maybe in a special case, a situation shows up where cars are relatively not moving among them but geographical they do, e.g. all of them are driving in the same direction with almost same velocity on the same highway-lane. However, everybody understands that there are too many scenarios not fulfilling these idealistic assumptions taken in

their simulation. In reality, conditions are neither stable nor constant and the networks are not only something like 50 times 50 square feet big, but networks are weird areas that do not meet mathematically proper forms.
As we know, that ExOR depends on its DRM and therefore on the topology of the network, a new question turns up. Is ExOR's benefit still big enough to strike out the incessant intensive computation of the link-state graph and DRM needed at big changes to the network?
An approach could be a classification of the network, e.g. classifying all cars moving in the same direction to the same class, in order to reduce the exhaustive computing. Maybe too big changes and node failure due to geographical position changes could be diminished.

At the beginning of VANET, there will be too few cars to benefit from ExOR compared to a routing with the best possible pre-determined path, but later on having a good saturation of cars using wireless; benefit will almost hit the roof. In the beginning it will be even very hard to keep alive a network somewhere out on highways, in hilly areas or simply were not a lot of cars are driving around.
Talking about weaknesses means also to have a look at the tricky candidate selection. Candidates will be selected as in theory, maybe with a slight change that choosing candidates will be selected from a classified set of nodes, but lets have a look at following situation. On a highway with a lot of tunnels and leading through semi-open area, such as a narrow and hilly topology, the transmission range of a 2 Mbps wireless will lessen from about 400 m, in an open topology, to 90 or even 40 meters. Now having computed the link-state graph, a node from the graph suddenly disappears, so this node will not be addressable. Worse is, in the same situation, if a node failure takes place when the prioritized candidate acknowledges the reception of the packet and every other candidate throws the packet away. Right at that moment the candidate, selected as the forwarder, reaches a tunnel, fails or is suddenly out of range of any network the packet will be lost. It can be too late until the original sender recognizes this. The best pre-determined path algorithm has the same problem, but our interest is how could be ExOR put in VANETs. Maybe it would make sense to have some backup paths that contain other nodes than the original path. Exactly this node failure demands for an additional backing, maybe the mentioned multi-path routing - called on temporarily - could support ExOR routing in hard environment.

As we see there still has to be done a lot of work with implementing, simulating and testing in order to deploy it commercially in VANETs. A combination of ExOR with one or more other protocols that strike ExOR's weaknesses out will surely result in an extremely good and stable routing.

First, I had to read the paper several times until I began to understand all the details and how it really works. I like the way they wrote the paper. The paper is - in my opinion - sometimes a bit confusing, because they easy understandable facts several times and very important and hard stuff only short. At the beginning I thought that the topic is very boring, but now I really like this topic and can also image to go deeper and further in this topic. Especially when combining ExOR with geographical methods. I hope that I could bring you their paper and my thoughts a bit closer and you like my work.

References

[MoBi03] Robert Morris and Sanjit Biswas (Hrsg.). Opportunistic Routing in MultiHop Wireless Networks. Technischer Bericht, Technischer Bericht, November 2003.

[WCEG+02] A. Woo, D. Culler, D. Estrin, D. Ganesan, B. Krishnamachari, S. Wicker and Qi Chen (Hrsg.). Complex behavior at scale: An experimental study of low-power wireless sensor networks. Technischer Bericht, UCLA/CSD-TR 02-0013, 2002.

List of Figures