DATA MANAGEMENT *in* VEHICULAR NETWORKS

Relevance-Aware Networking for Advanced Driver Assistance Systems

Vom Fachbereich Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von

TOBIAS MEUSER, M.SC.

Geboren am 13. Juli 1993 in Wetzlar

Vorsitz: Prof. Dr.-Ing. Anja Klein
Referent: Prof. Dr.-Ing. Ralf Steinmetz
Korreferent: Prof. Dr. Ioannis Stavrakakis

Tag der Einreichung: 22. Oktober 2019
Tag der Disputation: 16. Dezember 2019

Darmstadt 2019

Tobias Meuser, M.Sc.: *Data Management in Vehicular Networks,*
Relevance-Aware Networking for Advanced Driver Assistance Systems
Darmstadt, Technische Universität Darmstadt,
Jahr der Veröffentlichung der Dissertation auf TUprints: 2020
Tag der mündlichen Prüfung: 16. Dezember 2019

Dieses Dokument wird bereitgestellt von — This document is provided by
tuprints, E-Publishing-Service der Technischen Universität Darmstadt.
`http://tuprints.ulb.tu-darmstadt.de`
`tuprints@ulb.tu-darmstadt.de`

Bitte zitieren Sie dieses Dokument als: — Please cite this document as:
URN: `urn:nbn:de:tuda-tuprints-113780`
URL: `https://tuprints.ulb.tu-darmstadt.de/id/eprint/11378`

ABSTRACT

Future vehicles will exchange an increasing amount of data to increase their awareness beyond their local perception. This data is generated by the sensors of other vehicles, which share their local view of the environment. Compared to the data exchanged by today's vehicles, this data is much more fine-granular and, thus, changes more frequently, requiring much higher bandwidth to maintain an up-to-date view of the environment. The diverse level of accuracy or potential inaccuracy of vehicle-generated data, in conjunction with their increased bandwidth volume, poses considerable challenges for future vehicular networks.

The potential inaccuracy of data provided by other vehicles necessitates a validation, which requires knowledge about the measuring sensors. Besides, the higher bandwidth consumption requires a more accurate consideration of each vehicle's interest in data, as not everything can be exchanged. The paradigm of *Approximate Networks* is particularly well suited for the provisioning of fine-granular data, as it allows to trade network and computation resources with the availability and quality of data.

Our contributions in this thesis amount to developing mechanisms to apply the concept of approximate networks in the vehicular scenario. For this purpose, we first develop mechanisms for the assessment of data in these networks, which are the basis for our approach to approximate vehicular networks. As our *first contribution*, we propose an aggregation scheme to increase the data quality in the network. Our innovative aggregation scheme considers the heterogeneity of sensors and data-specific properties to adapt the influence of old measurements and increase the quality of the resulting aggregate. We then investigate the relevance of data to a specific vehicle as our *second contribution*, which relies on the prediction of the specific vehicle's future context. By combining the accuracy of the aggregate and its relevance, we determine the expected gain for a specific vehicle, the so-called impact. This impact is key for effective data prioritization and builds the foundation of our approximate vehicular network. As our *third contribution*, we design and implement an approximate vehicular network based on *Diverse Prioritization and Treatment*, aiming at improving network performance without increasing the resources consumed, as typically advocated under approximate networking. A probabilistic mechanism is proposed to properly modulate the redundancy of the messages in the network, leading to their increased overall availability to the interested vehicles without increasing the consumed resources.

Finally, we design and develop our VEHICLE.KOM platform that is used to assess the effectiveness of the developed mechanisms under varying environmental conditions. We show that our aggregation scheme drastically reduces the false aggregates and adapts its behavior to lifetime and accuracy effectively. In addition, we demonstrate the effectiveness of our approach to approximate vehicular networking, by showing a drastic increase in the network performance under dynamic network conditions, especially when considering cooperation between vehicles.

KURZFASSUNG

In den nächsten Jahren werden Fahrzeuge immer mehr Daten austauschen, um mögliche Gefahren auch außerhalb ihrer eigenen Sensorreichweite wahrnehmen zu können. Die so ausgetauschten Daten werden von anderen Fahrzeugen generiert, die ihre lokale Wahrnehmung mit den Fahrzeugen im Netzwerk teilen. Dadurch ist es möglich, deutlich feingranularer auf Veränderungen der Umwelt zu reagieren als es heute möglich ist. Allerdings hat dies zur Folge, dass deutlich mehr Kommunikationsbandbreite benötigt wird, um diese Veränderungen mit anderen Fahrzeugen zu teilen. Dabei stellt die mögliche Ungenauigkeit der ausgetauschten Daten in Kombination mit dem steigenden Bandbreitenbedarf eine große Herausforderung für zukünftige Fahrzeugnetzwerke dar.

Durch diese mögliche Ungenauigkeit der Daten sind Verfahren nötig, die Sensordaten von anderen Fahrzeugen validieren können, wozu Informationen über die messende Sensorik benötigt werden. Zusätzlich muss der steigende Bandbreitenbedarf kompensiert werden, was eine Analyse des Datenbedarfs eines einzelnen Fahrzeugs erfordert. In diesem Kontext eignet sich das Konzept der *Approximate Networks* besonders gut, da dieses eine Abwägung zwischen Ressourceneffizienz und der Verfügbarkeit/Qualität von Daten ermöglicht.

Dementsprechend tragen unsere Beiträge in dieser Arbeit zum Konzept der *Approximate Networks* in Fahrzeugnetzwerken bei. Dazu entwickeln wir zunächst Mechanismen zur Datenbewertung in Fahrzeugnetzwerken, welche dann als Grundlage für unseren Ansatz für unsere entwickelten Fahrzeugnetzwerke dienen. Als unseren ersten Beitrag entwickeln wir ein innovatives Aggregationsschema, welches die Datenqualität in Fahrzeugnetzwerken erhöht, indem es die Heterogenität von Sensoren in Kombination mit den Eigenschaften der generierten Daten berücksichtigt, um den Einfluss von älteren Messdaten auf das Aggregationsergebnis zu bestimmen. Dieses Gewicht wird so gewählt, dass für den jeweiligen Datentyp die Qualität des Aggregates erhöht wird. Im Anschluss untersuchen wir die Relevanz von Daten für ein spezifisches Fahrzeug als unseren zweiten Beitrag, wobei wir eine Prädiktion des Fahrzeugkontexts nutzen, um die Nützlichkeit eines Datums für das Fahrzeug zu bestimmen. Basierend auf der Genauigkeit der Daten und der fahrzeugspezifischen Relevanz bestimmen wir den Einfluss der generierten Daten auf ein spezifisches Fahrzeug. Diese Einfluss Metrik ist ein wichtiger Aspekt für eine effektive Priorisierung von Daten und bildet die Grundlage für unser Konzept von *Approximate Vehicular Networks*. Dieses designen und entwickeln wir als unseren dritten Beitrag basierend auf dem Konzept von *Diverse Prioritization and Treatment*. Mit diesem Konzept ist es uns möglich, die Netzwerkperformanz zu erhöhen, ohne dabei die Menge der verbrauchten Kommunikationsressourcen zu ändern. Dadurch folgt unser Konzept der Grundidee der *Approximate Networks*. In diesem Kontext schlagen wir ein wahrscheinlichkeitsbasiertes Verfahren vor, welches die Redundanz der Nachrichten im Netzwerk so anpasst, dass die Verfügbarkeit von Nachrichten für interessierte Fahrzeuge erhöht wird, ohne dabei jedoch die genutzten Kommunikationsressourcen zu verändern.

Wir nutzen dann unser VEHICLE.KOM Framework, um die entwickelten Verfahren in einer ausgiebigen Evaluation zu analysieren und zu bewerten., wobei wir verschiedene Umwelteinflüsse auf unsere Ansätze betrachten. Wir zeigen, dass unser Aggregationsschema die Datenqualität im Netzwerk durch Anpassung an die Genauigkeit und Langlebigkeit der Daten signifikant erhöht. Zusätzlich demonstrieren wir die Verbesserungen durch unseren Ansatz für *Approximate Vehicular Networks* in dynamischen Umgebungen, wobei besonders der Mehrwert von Kooperationen zwischen Fahrzeugen beleuchtet wird.

CONTENTS

1 INTRODUCTION

TODAY's vehicles become increasingly connected, which enables the exchange of road-traffic data with other vehicles. In addition to the currently available cellular communication via Long Term Evolution (LTE), decentralized communication via the Wifi-based 802.11p is expected to be deployed in the future. Today's vehicles already exchange coarse-grained road-traffic data, like accidents and jams, and provide maintenance information to the manufacturer [143]. These coarse-grained road-traffic data are provided to the driver by a central entity via the radio, radio-related technologies like Traffic Message Channel (TMC), or the cellular network. This central entity relies on reports of traffic participants (passengers and vehicular sensors) to verify the correctness of the provided data and distributes only validated data back to the vehicles. This validation of data entries is often performed statically without consideration of the quality of the provided data. Additionally, the validated data is provided to all vehicles within a certain region, without considering the individual context (e. g., location, future route) of the vehicle. For coarse-grained road-traffic data, this approach induces limited overhead due to the limited amount of shared data.

With the expected increase in the level of autonomy of future vehicles, road-traffic data of much finer granularity than currently available will be required [181]. Examples for these fine-granular data are minor changes in traffic flow, changes in lane-marking, and properties of the road, to which we refer to as road events in the following. Compared to high-importance data like traffic jams and accidents, these fine-granular data changes more frequently. Thus, they are also exchanged more frequently and can only be reported by the sensors of the vehicles driving on the street [87–89, 115]. This increase load renders the aforementioned strategies for validation and dissemination of these data impractical: For the validation, the amount of available data increases drastically, such that a validation process based on static parameters is hardly possible. Additionally, vehicle-specific properties like heterogeneous sensors and the diverse accuracy of the generated data complicate an automated validation of this data. For the dissemination, the available communication bandwidth will be insufficient to provide this amount of validated data to all vehicles in the network or even in a region. Additionally, these data is not relevant to all vehicles in a region, but generally only to a small subset of these. Consequently, the vehicles on the streets form a huge, context-aware sensor-actuator network, intending to provide maximum safety and comfort to the passengers of the vehicles given the available communication bandwidth of the communication network. In this thesis, we propose the concept of approximate vehicular networks to cope with the aforementioned challenges.

1.1 MOTIVATION FOR APPROXIMATE VEHICULAR NETWORKS

In accordance with approximate computing [25], approximate networking may be viewed as a networking paradigm that trades communication and computation resources with the availability and quality of data. While the performance of conventional networks is generally increased by increasing the available communication resources, approximate networks focus on adapting the performance to predefined communication resources by, e. g., decreasing the quality or quantity of the provided data to free bandwidth for the transmission of previously not transmitted data. Approximate vehicular networks are approximate networks tailored for the vehicular environment. In this environment, the data itself is commonly more important then the provider [15] which is generally unknown, as data is proactively generated by the vehicles in the network and shared among them. Due to the inaccuracy of the built-in sensors and environmental conditions, these data and the associated measurements are not always correct and thus introduce uncertainty. Today's networks commonly ignore this uncertainty and its influence on the data consumer, which is expected to become a major issue for future vehicular networks.

The data generated by vehicles is transmitted to other vehicles to improve their decision-making [192]. While some works follow a completely decentralized data management approach [109, 184, 187], others rely on server-based coordination to improve the performance of data management [45, 147]. Decentralized data management is commonly used to distribute data to vehicles close to the measurement location via Wifi-based communication technology to support, e. g., accident prevention [24, 86]. Although long-range data exchange is also possible using decentralized solutions with multi-hop relaying, the induced latency is generally very high [103], which renders decentralized data management approaches unsuitable for short-lived or high-importance data. For this purpose, the cellular network is generally used, as the transmission latency is almost unaffected by the transmission distance. The data exchange via the cellular network often relies on a centralized coordination unit (storage server, broker) to manage the data transmission to the vehicles. However, due to the limitations in bandwidth and the induced costs due to the usage of licensed frequency bands, the transmission of data via the cellular network is limited. Thus, data is filtered at the server to avoid unnecessary transmission to unconcerned vehicles [66, 128]. However, the state-of-the-art approaches for data exchange in vehicular networks are controlled by static attributes, like the distance between a vehicle and a data location [54, 187]. Such approaches are very inefficient, as the influence of data to a vehicle commonly does not depend on these static attributes. That is, the data in such networks should generally improve the vehicle's driving behavior, and the possible improvement depends strongly on a multitude of attributes, which consider the vehicle and the shared piece of data. For a vehicle, the context of the vehicle (e. g., vehicle type, passenger preferences, location) needs to be considered. For the piece of data itself, the available meta-information (e. g., measurement location, measurement date, accuracy, data type) determines the importance of that piece of data for the vehicle.

In addition to increasing the filtering of data at the server, different approaches towards hybrid vehicular networks have been proposed to further increase the efficiency of the bandwidth utilization by combining cellular and local communication [175, 201]. These approaches commonly organize vehicles in so-called clusters, such that all communication is performed by one so-called cluster-head. These clusters increase the communication efficiency, as each piece of data only needs to be transmitted once to each individual cluster. While these approaches perform well if the rate of topology changes is low, they become inefficient in highly dynamic networks, as vehicles regularly disconnect from their cluster head. This disconnect leads to a loss of data entries with potentially high impact, which drastically reduces the performance of cluster-based approaches in urban areas.

The influence of data on the behavior of the vehicles has barely been studied in the literature, although it is a pivotal aspect of the relevance of data for the receiving vehicles. Additionally, this influence is generally not deterministic, but is influenced by the uncertainty of these data-specific properties and the vehicle's context changes. In the related-work, this uncertainty has commonly been compensated for through drastic over-provisioning, but has rarely been considered as a pivotal property of the network. Uncertainty in this context refers to the network conditions, and the assessment of data itself, for both the dissemination and validation of data. To be efficient, future vehicular networks will require (i) the assessment of data and the determination of its influence on a vehicle and (ii) the utilization of uncertainty to increase the network performance. Uncertainty can be used to increase network performance by designing adaptive and robust communication mechanisms, as addressed in this thesis based on the paradigm of *approximate vehicular networks*.

1.2 RESEARCH CHALLENGES

The increasing amount of data shared in vehicular networks poses additional challenges to their dissemination, processing, and validation. The following research challenges are the basis for our approximate vehicular network.

Challenge: Providing high-quality measurements to data consumers.

In vehicular networks, the heterogeneity of the sensors regarding availability and quality influences the quality of the data provided to the vehicles or any relevant data consumer. While heterogeneity is not an issue when only local sensor measurements and specialized applications are used, future vehicular applications will be required to utilize data provided by other vehicles to increase traffic safety and driver comfort further. In these networks, vehicles will receive multiple measurements generated from diverse sources with diverse quality and at different measurement times. These potentially contradicting measurements need to be properly aggregated into high-quality measurement data to support effectively various vehicular applications relying on them. Such an aggregation or joint processing of diverse measurements is a major challenge, due to the aforementioned diverse quality and measurement times of the provided measurements.

Challenge: Influence of context on the relevance of data.

The data exchanged in a vehicular network is commonly context-sensitive, i. e., a produced piece of data is not relevant for all vehicles in the network. The relevance of data for a specific vehicle is strongly connected to its expected behavior and its current context, like location and type of the vehicle. A piece of data is considered relevant for a vehicle if this piece of data is expected to improve the vehicle's future behavior regarding safety or driver comfort. The assessment of relevance poses a significant challenge because of the uncertainty in future behavior and context of the vehicles, and the uncertainty of the influence of a piece of data on that behavior and the associated improvements. Thus, an appropriate measure for the influence of data to a data consumer needs to consider these two influence factors.

Challenge: Impact-aware dissemination considering privacy demands.

The data in a vehicular network has different usability to the data consumers, which depends on multiple factors like data quality, the context of the vehicle, and the importance of the data. As an example, a traffic jam in proximity is much more useful than a traffic sign far away. When exchanging data, especially via the cellular network, the available communication bandwidth is limited. Thus, appropriate prioritization of data entries in this network is necessary to utilize the available bandwidth efficiently. This may also include the cooperative reception of data by vehicles to further increase the efficiency of the network. In addition, the dissemination of context-sensitive data via the cellular network often requires the location of the vehicle, which compromises the privacy of the passengers. Thus, privacy-sensitive vehicles would not be able to receive context-sensitive data unless their location privacy is considered by the data dissemination mechanisms.

1.3 RESEARCH GOALS AND CONTRIBUTIONS

The main goal of this thesis is the modeling, design, and evaluation of our concept for approximate vehicular networks and the necessary assessment for data quality. These objectives are divided into the following research goals.

Research Goal 1: Aggregation scheme for data of diverse accuracy due to source heterogeneity or age.

When multiple sources provide measurements with diverse sensor quality, these measurements will potentially be contradicting. In this case, it is pivotal for the vehicle to determine the real value, so that the applications running on the vehicle are properly supported and not negatively impacted by false or low-quality measurements. For this purpose, we develop a model for inaccurate measurements that tracks the inaccuracy of the sensors through the aggregation [118]. We develop an aggregation scheme to produce high-quality data, which modulates the influence of old measurements on the aggregate. The proposed aggregation scheme considers data-specific properties - such as the quality of the data and its expected lifetime - and aims to maximize the quality of the resulting aggregates [121, 124].

Research Goal 2: Assessment of the usability of data for the data consumers.

While our aggregation scheme for data of diverse accuracy increases the overall quality of data available, this more accurate data may or may not be useful to certain vehicles. The usability of such data for a specific data consumer (vehicle) depends on the data's quality and type, as well as the context and expected behavior of the specific data consumer. To determine the usability of data for a data consumer, we propose an approach considering the future context of the vehicle and the future state of the measured road event [119]. We utilize statistical methods to predict both the future context and event state, as these quantities are generally unknown to the server assessing the usability. To quantify the usability of data for a vehicle, we propose a mechanism to determine the specific-impact of a message for a specific vehicle, which is a pivotal meta-information for efficient dissemination of data in vehicular networks [122].

Research Goal 3: Approximate vehicular networking: Improving efficiencies by exploiting uncertainties and trading resources.

In our vehicular network, a central server forwards the data to concerned vehicles based on the impact value prioritizing high-impact data. For this reason, data is typically received redundantly by all vehicles in proximity. When we consider cooperation between vehicles, the benefit provided through cooperation is low, as only known data is received. To increase the potential of cooperation, our approach for approximate vehicular networks relies on probabilistic mechanisms to coordinate the transmission from the server to the vehicles [122]. These probabilistic mechanisms reduce the utilized bandwidth for high-impact messages and, thus, frees communication resources, which can then be used to receive data that would not have been received previously. This concept constitutes our contribution to approximate vehicular networking, in which even high-impact data may be dropped to free network resources, which can then be used more efficiently.

Besides, the utilization of a central server requires the vehicles to share their context with this server, which compromises the privacy of the passengers. To alleviate this issue, we explicitly model privacy constraints in our approach for approximate vehicular networks, such that vehicles are able to protect their privacy by sharing an imprecise representation of their context [123, 125].

In this thesis, we focus on concepts for approximate vehicular networks relying on cellular or hybrid vehicular networks, i. e., vehicular networks with more than one available communication technology. For this purpose, we communicate using centralized and decentralized communication technologies. However, we do not perform research in decentralized communication protocols, as plenty of research is available in this field [90, 103, 136, 137].

In a real-world vehicular network, vehicles might not want to cooperate or provide false measurements to increase their own benefit or damage the network. Many works have analyzed the impact of malicious or non-cooperative vehicles or network participants in general [91, 94, 139, 186], thus, we do provide relevant related work and do not conduct own research in this area.

In this work, incentive mechanisms for the sharing of data between vehicles are not investigated, but possible incentives are a promising research direction to be investigated in future work. Especially, the pricing of data in these networks is still an open issue, which needs to be addressed to increase the willingness of vehicles to share their perceived data.

1.4 STRUCTURE OF THE THESIS

After this short introduction to this thesis, we describe the necessary background and previous works regarding data dissemination and assessment in vehicular networks in Chapter 2. In Chapter 3, we describe the properties of large-scale vehicular networks and analyze them to motivate the core contributions of this thesis. Based on this analysis, we propose our framework for data assessment in vehicular networks in Chapter 4, which is then utilized in Chapter 5 to increase the performance of the network by prioritizing data according to their influence to the vehicles. We then present our *Vehicle.KOM* framework in Chapter 6, which is tailored for the rapid development of data dissemination mechanisms for vehicular networks and the basis for our evaluation In Chapter 7, we then perform an in-depth evaluation of the influence of our data assessment for the data quality in vehicular networks, which is followed by the evaluation of our cooperative approximate vehicular networks. We then conclude this thesis in Chapter 8 by providing a summary of our core contributions. Finally, we provide an outlook on potential future work.

2 BACKGROUND AND RELATED WORK

In this chapter, we provide background information about communication technologies for data exchange in vehicular networks and discuss the current state-of-the-art regarding quality-centric vehicular networks. We first provide an insight into technologies used for data exchange in vehicular networks in Section 2.1. After that, we investigate the different types of communication in these networks in Section 2.2 and approaches to consider the quality of the network in Section 2.3. Then, we describe the current research state regarding Approximate Networking in Section 2.4. In Section 2.4, we describe state-of-the-art combining the assessment of data with vehicular networks by adapting the network to the quality of data. Finally, we conclude this chapter in Section 2.5.

2.1 COMMUNICATION TECHNOLOGIES FOR VEHICULAR DATA EXCHANGE

Vehicular networks are highly mobile and dynamic, posing additional challenges to the underlying communication technology [50, 68]. While many works focus on the dissemination of data, other research has been performed in the efficient processing of data in a vehicular network [111, 112]. Data dissemination in vehicular networks has been researched for almost three decades [161], always adapting to the current trends in network technology and methods. In the literature, different communication technologies have been used for the construction of vehicular networks, which can be divided into infrastructure-based and infrastructure-less communication technologies. Especially for infrastructure-less communication technologies, the high frequency of disconnects, and the potentially high relative speed of the vehicles is an issue [9].

2.1.1 *Infrastructure-less Communication*

Infrastructure-less communication technologies do not utilize any infrastructure to exchange data between two vehicles. This communication type can be used to either exchange data locally or to improve the efficiency of the infrastructure-based communication by transmitting it to only a single vehicle, which distributes the received data locally (offloading). In these networks, there are issues concerning the coordination of data transmission, the transmission range, and the mobility that can be compensated for. These parameters vary between the different communication technologies, which are presented in the following.

2.1.1.1 *Wifi-based Device-to-Device Communication*

Wifi-based communication in vehicular networks is commonly performed via 802.11p, which is an extension of the 802.11a standard. 802.11p is specially tailored for data transmission in vehicular environments. While its PHY-layer is based on 802.11a, 802.11p has additional adaptations to increase the transmission range, reduced channel width, better support for varying temperatures [82], and an adapted Distributed Coordination Function (DCF) [9]. Based on 802.11p, different standards evolved to support local communication between vehicles. In the United Stated, Dedicated Short-Range Communication (DSRC) utilizes 802.11p as a basis to communicate in an unlicensed spectrum between 5.850GHz and 5.925GHz [5]. In Europe, ITS-G5 follows a similar idea and is performed in the unlicensed 5GHz frequency band [61], which is almost similar.

Many different works analyze the performance of 802.11p based on simulative environments [43, 60, 179] and mathematical models [28, 174], which is justified by the lack of evaluation hardware. These works conclude that the delay of message transmission is comparably low, especially compared to infrastructure-based communication technologies like LTE [126]. These simulative works, however, confirm the drastically decreasing bandwidth and a drastically increasing message delay with an increasing number of vehicles in proximity. This bandwidth increase is, to a large degree, caused by the inefficiency of the utilized Decentralized Congestion Control (DCC) [19, 76, 79, 173]. To further improve the efficiency of 802.11p, different approaches to improve the DCC have been proposed [77, 116, 165]. However, 802.11p cannot provide any guarantees regarding latency due to the DCC [14].

2.1.1.2 *Cellular Device-to-Device Communication*

Although cellular communication, including 5G, is generally considered to be an infrastructure-based communication technology, it supports the possibility for in-band device-to-device communication [18, 169, 191]. In the literature, many works investigated the potential performance increase by utilizing device-to-device communication [55, 93, 141]. For vehicular networks, cellular-based device-to-device communication has been commonly used to increase the efficiency of multicast distribution of messages [23, 202], which is an important factor for the centralized data exchange in vehicular networks. In addition, its use for the exchange of local Machine-to-Machine (M2M) communication has been considered in [145]. Since the description of the standard, the automotive community investigates the possibility of utilizing this device-to-device communication to exchange data locally via 5G [67]. Compared to 802.11p-based data exchange, 5G-based device-to-device communication provides the possibility for coordination of data transmission by the cell tower [18, 99]. While this is not necessary for this type of communication, this coordination by the cell removes the necessity for DCC and, thus, increase the performance compared to traditional device-to-device communication [69].

2.1.1.3 *On the Usability of Communication Technologies in Vehicular Networks*

In conclusion, both technologies have their advantages and disadvantages. For 802.11p, the transmission is free of charge due to the use of an unlicensed spectrum, but the transmission is comparably inefficient due to the lack of a centralized coordination unit. This lack is compensated for by cellular-based device-to-device communication. However, this communication type commonly utilizes a licensed spectrum to perform data exchange, although the coexistence on the Wifi spectrum has also been considered [40, 168, 185, 195]. Especially the usage of the licensed spectrum might not be free of charge, while the usage of the unlicensed spectrum might interfere with other communication technologies. Yet, the direction of vehicular communication is not entirely clear, but will likely be determined within the next years in projects like [46].

2.1.2 *Infrastructure-based Communication*

In contrast to infrastructure-less communication, infrastructure-based communication relies on cell towers and the Internet to exchange data between vehicles. This is often referred to as the Internet of Vehicles (IoV) [8, 73, 189]. In IoV, vehicles can exchange arbitrary data between each other and with a so-called Vehicular Cloud, which provides all services required by fully-autonomous vehicles [73]. The data exchange with the Vehicular Cloud is then performed via the cellular network, which enables efficient data transmission over large distances through the utilization of the Internet. In the following, we provide a short overview of the available cellular communication technologies and discuss their suitability for vehicular networks.

2.1.2.1 *Universal Mobile Telecommunications System*

UMTS is an umbrella term for standards developed in the third generation of radio technologies. It provides Frequency Division Duplex (FDD) [1] and Time Division Duplex (TDD) [2], and later introduced High Speed Packet Access (HSPA) to further increase the transmission speed. A big advantage compared to stationary Wifi-networks is the increased stability under mobility [52]. Regarding 3G networks, most works investigated on the possibility of increasing the network performance through local (infrastructure-less) offloading of data [22, 30, 83]. That aside, the research focus on third-generation radio technologies is comparably limited due to the limited available bandwidth and the initial high costs of data transmission.

2.1.2.2 *Long Term Evolution*

LTE is still part of the third generation of radio technologies, although often considered to be part of the fourth generation [3]. Starting from LTE Release 10, Long Term Evolution-Advanced (LTE-A) is available, which is the first real standard of the fourth generation of radio technologies. With LTE-A, even higher bandwidths are available than with LTE, peaking at 1Gb/s in downlink and 500Mb/s in uplink [74]. This bandwidth increase is achieved through carrier aggregation, in which multiple smaller frequency bandwidth bands are virtually merged into one large frequency band. With the introduction of LTE-A, it is first possible to also perform carrier-controlled device-to-device communication using cellular communication technology [99].

For vehicular networks, both LTE and LTE-A have been considered. There are several applications scenarios for these technologies in vehicular networks [16]. European Telecommunications Standards Institute (ETSI) has performed an analysis for the suitability of LTE for the dissemination of highly periodic messages [62], like Cooperative Awareness Messages (CAMs) which are described in Section 2.1.3. They show that the performance regarding latency of CAMs drops quite severely if the number of messages in proximity is high. In contrast, the Decentralized Environment Notification Messages (DENMs), also introduced in Section 2.1.3, can be transmitted via the cellular network, as the frequency of messages might be much lower [62]. In this case, the backend-server needs to filter the messages and only share an aggregated value. This leads to better scalability and has been demonstrated in [144]. However, the potentially high bandwidth consumption might interfere with other traffic in the network, if all data is transmitted through the backend. Thus, solutions have been proposed to reduce the utilized bandwidth through probabilistic filtering [37, 84] or cluster-based communication [117, 148].

2.1.3 *Vehicular Data Types*

In vehicular networks, different messages have been standardized, which support the functionality and efficiency of the vehicles in the network. While many messages could be named here, we focus the Cooperative Awareness Message (CAM) and the Decentralized Environment Notification Message (DENM) in the following, as these messages are most influential for this thesis.

2.1.3.1 *Cooperative Awareness Message*

Cooperative Awareness Messages (CAMs) are disseminated in a vehicular network and provide "information of presence, positions as well as basic status of communicating ITS stations to neighbouring ITS stations that are located within a single hop distance" [64]. This message is periodically generated and frequently exchanged between all vehicles in proximity to provide awareness of the presence of other vehicles [35]. This frequent exchange of messages increases the load to the wireless channel, which has been studied by several works [36, 49, 59]. Due to their periodic nature, these CAMs can then be used to determine the vehicle's current neighborhood to coordinate, for example, offloading of the cellular connection between vehicles. In this thesis, we do consider CAMs only as periodic beacons to determine the neighborhood of a vehicle.

2.1.3.2 *Decentralized Environment Notification*

The purpose of Decentralized Environment Notification Messages (DENMs) is to alert road users of detected events and, thus, is an event-driven message [65]. Possible road events are an emergency breaking and an accident, but also less severe events like road adhesion or strong wind [65]. DENMs are generally disseminated in an area, which can be rectangular, circular, or elliptical [63]. In this thesis, we consider the dissemination of DENM-like messages for the dissemination of road events. However, compared to the specification provided by ETSI [65], we employ a more accurate concept for the necessity for the transmission of a DENM to a vehicle.

2.2 DATA EXCHANGE IN VEHICULAR NETWORKS

In this section, we present the state-of-the-art for data-exchange in vehicular networks. While today's vehicle networking capabilities can exchange information about free parking spaces [95] or jams [78], future data will become much more fine-granular [89]. The communication of this data via vehicular networks is commonly context-sensitive due to the context-sensitivity of the active vehicular applications [172]. Examples for such applications are, amount others, collision warnings [166] and traffic conditions notification [182]. In the following, we first investigate infrastructure-less data exchange supporting vehicular applications, followed by infrastructure-based data exchange. Then, we describe the possibilities for hybrid dissemination of data, which is more efficient than relying on a single communication technology.

2.2.1 *Infrastructure-less Data Exchange*

Vehicular Ad-hoc Networks (VANETs) are a special type of Mobile Ad-hoc Networks (MANETs), in which vehicles form a self-organizing and decentrally controlled network [10, 180]. This general idea is similar to Peer-to-Peer (P2P) networks, in which nodes decentrally coordinate themselves [163]. In VANETs, the mobility of nodes is much higher than in MANETs, which increases the frequency of topology changes [10, 103] and decreases the lifetime of routing paths drastically [20]. Via these networks, different applications are possible, like alert generations, vehicle maintenance, communicate services, and security services [167]. We divide the available VANET routing protocols into four categories, which are topology-based, broadcast, geographic, and information-centric protocols.

2.2.1.1 *Topology-based Protocols*

Topology-based protocols generally suffer from frequent changes in the network topology [129]. There are two types of topology-based protocols, which are proactive and reactive protocols. Proactive data routing approaches maintain the network topology even if no payload is transmitted by periodically probing the network. Due to the much higher mobility and the necessary overhead for route maintenance, proactive protocols designed for MANETs [42, 72, 142, 157] do not perform well in the vehicular setting. Thus, special proactive routing approaches are tailored for VANETs, which focus on the compensation for topology changes [132, 171]. However, even these protocols suffer from the high frequency of topological changes and induce a lot of control traffic.

Compared to that, reactive routing is slower, as the route finding is only performed after a vehicle wants to transmit data [57]. Different works proposed adaptations to common MANET protocols to adapt them for the usage in vehicular networks [6, 53]. However, the management overhead remains an issue for these protocols.

2.2.1.2 *Broadcast Protocols*

For certain data entries of high importance, a broadcast of data in the network is required, i. e., data is disseminated to all vehicles in the network until the lifetime is expired [21, 38]. For these protocols, the high number of rebroadcasts, especially of vehicles in proximity, is an issue and reduces the available bandwidth. This problem is known as the broadcast storm problem, which is addressed by several works in the literature [113, 159, 170, 203, 204]. These approaches can be utilized well for data of high-importance, but generally might lead to channel congestion if the number of exchanged messages increases.

2.2.1.3 *Geographic Protocols*

Geographic routing protocols aim at routing data to a certain location [57]. For this purpose, many protocols have been proposed that deliver data to a certain location using the store-carry-forward paradigm [54, 109, 184, 187, 199]. In general, these approaches consider the road topology and the movement of vehicles to select vehicles that are most suitable as data carriers. However, as these approaches rely only on VANETs, the data dissemination is very slow, which is not suitable for all use cases.

2.2.1.4 *Information-Centric Protocols*

While the geographic protocols already introduce context to the dissemination of messages, this context might not be efficient in capturing the required data of the vehicles. For this purpose, information-centric approaches have been investigated [15]. The information-centric protocols enable the request of certain data entries using information-specific properties, which is considered to be more accurate than IP-based networking. These approaches partially face similar issues compared to broadcast protocols, as the requests for data are often broadcasted in the network. For the efficient dissemination of data, different approaches have been designed [13, 177, 178, 193], which drastically outperform IP-based networks. These protocols can capture the interest of vehicles well, but still introduce a potentially high delay to the message delivery if data is to be transmitted over multiple hops in the network.

2.2.2 *Cloud-Supported Provisioning of Data*

While infrastructure-less communication is well suited for the dissemination of local data, it performs poorly if data shall be disseminated to distant areas. In this case, infrastructure-based communication is well suited, which can transmit data independent of the distance between the sender and the receiver, but relies on cell coverage.

The data shared between them is often information-centric [15], i. e., only vehicles query individual pieces of information instead of information from a specific host.

2.2.2.1 *Geocast Protocols*

Geocast is a location-dependent type of multicast, in which vehicles in a certain area receive data. It is frequently used in practice as a simple possibility to consider the location of nodes in the dissemination and reduce the load on the cellular network. However, most geocast protocols have been developed for infrastructure-less communication, as the developed approaches consider additional information-specific properties like age and type. These protocols are presented in the following.

2.2.2.2 *Information-Centric Protocols*

For cloud-based communication, the Publish/Subscribe (Pub/Sub) paradigm is well suited to support vehicular communication [66, 128]. In the Pub/Sub paradigm, vehicles express their interest in data with certain properties using *subscriptions*. These subscriptions are shared with a so-called broker, which is responsible for the dissemination of messages. When a vehicle wants to share a DENM, it creates a *notification* containing the available data and sends it to the broker. The broker then forwards the received notification to the vehicles with matching subscriptions.

Several approaches have been developed, which consider the location, age, and type of data for the matching of subscriptions to notifications [45, 56]. In these approaches, the broker has only forwarding capabilities, which is not necessarily true for all cloud-based solutions. Some approaches consider a central server, which manages the data and can perform aggregation and filtering to reduce the overall network traffic [31, 34]. Through the centralized management of the cloud, these approaches reduce the network traffic compared to pure forwarding-based brokers and increase the quality of the provided data. For an efficient usage of cloud-based approaches, the data consumption needs to be reduced, which we investigate in the next section.

2.2.3 *Offloading in Heterogeneous Networks*

When multiple communication technologies are available, there is the possibility to use the concept of transitions [12] to adapt the communication technology or dissemination strategy to the current environment. One possible dissemination strategy is offloading, which can be used to increase the efficiency of the utilized cellular network resource [147, 153]. There is the possibility to perform offloading with and without the additional infrastructure [147]. While several works have investigated on the usage of so-called Road Side Units (RSUs) (mobile access points) [27, 41, 44, 114, 155, 156] and achieved good results regarding the efficiency of the offloading, the deployment of these RSUs is still unclear. Thus, we focus on the approaches without additional infrastructure in the following. While some approaches use opportunistic networks to disseminate data locally [102], the majority create clusters, i. e., groups of vehicles, which perform their transmission in a coordinated manner. That is, one or more vehicles are elected as a so-called cluster head, which is responsible for the transmission and reception of data via the cellular network. In the literature, clustering approaches for MANETs [80, 106, 151] have been proposed. Similarly to the difference between routing approaches for MANETs and VANETs, the clustering approaches for MANETs are not directly applicable to VANETs due to the lack of mobility support. For VANET clustering approaches, there are three main directions for the election of the cluster head: (i) reducing the number of disconnects of cluster members from the cluster head [26, 96, 130, 196], (ii) increasing the connectivity of the cluster by selecting the best-connected vehicle [200], and (iii) a combination of the two [175, 201]. In addition to clustering approaches that actively transmit control information, there are also hybrid approaches that are passively coordinating the transmission of data [98, 117]. While VANET clustering approaches generally reduce the number of topology changes, these approaches still have issues with changes in the network topology, i. e., if a member of a cluster detaches from its cluster head. In this case, the member of the cluster is without connection until a reclustering is initiated [47]. This happens after a timeout to prevent too frequent reclustering caused by packet loss on the local communication channel. To reduce the impact of disconnects to the members of the cluster, several approaches have been introduced [108, 202].

2.3 NETWORK QUALITY ASSESSMENT IN DISTRIBUTED NETWORKS

In addition to the networking aspects, the assessment of data quality is a pivotal aspect of this thesis, which later influences the transmission of data between vehicles. Data quality is a pivotal aspect and has commonly been referred to in the literature, often without defining good or bad data quality like in [17, 75]. For this purpose, we describe previous works aiming at defining data quality for different scenarios. We start with data-independent network quality parameters in Section 2.3.1, followed by the data-dependent network quality parameters in Section 2.3.2.

2.3.1 *Quality of Service Assessment*

Regarding context-independent network quality assessment, many works have focused on increasing the QoS, which they generally connect to the share of transmitted data [104, 107, 135, 162, 188]. Examples for considered QoS parameters are the packet loss probability and the average packet delay, which can estimate the behavior of an application to the data or lack of data. Thus, these works consider the influence of data transmission and the age of data but do not investigate the effects of outdated data.

2.3.2 *Data-Quality Assessment*

While many works do not focus on the assessment of data quality and utilize less precise definitions [138], the definition of data-dependent network quality parameters like the accuracy, relevance, and freshness of data has become a hot topic in recent literature [71, 134, 140, 190, 198]. These approaches rate the quality of data considering the reputation of the providing entity, which depends on the accuracy of the provided data. These works, however, mainly focus on the assessment of the measurement itself, without considering the usability of a piece of data for the underlying application, which is reflected in metrics like the relevance and freshness. To assess the usability of data, the usage of machine learning has been proposed [75]. However, these approaches have flaws due to the unpredictability of the output of most machine learning approaches. In contrast, type-based prioritization for resource-constraint environments has been used to increase network performance [105]. In contrast, other works used statistical methods to investigate the planned trajectories of the vehicles to assess the requirement of sharing data [11, 51]. In addition to that, the influence of age on data on applications has also been considered [92, 197]. However, these works generally assume fixed restrictions to the age of data, while the restrictions generally vary depending on the context of the vehicle and the application.

2.4 APPROXIMATE NETWORKS

In a typical (digital) communications environment, the received symbol in the presence of noise is a good approximation of the transmitted symbol, as it will be either the same symbol as the transmitted one or a neighboring signal/symbol to the true transmitted one in the signal constellation. Such a neighboring symbol differs from the transmitted symbol only in one or more (if the noise level is higher) of the Least Significant Bits (LSBs) it represents [160]. This is the case because signals next to the transmitted one in the signal constellation correspond to symbols that differ from the original in the LSBs; the top Most Significant Bits (MSBs) would be the last to be affected as the noise level increases. One can say that in general (digital) communications is *Approximate Communications* in the sense that typically errors occur and, thus, the received signal/symbol is an approximation of the transmitted one. By increasing the available resources, the approximation will be improving, and beyond a threshold, the approximation will be practically always perfect (i.e., no approximation). As, in general, the better the approximation, the higher the QoS provided by the communication system, one can trade off resources for QoS, approximation, or accuracy. The latter is in line with the recently coined concept of *Approximate Computing*, under which computational accuracy can be traded off for (energy/processor/memory) resources [25].

One can extend the aforementioned concepts to a networking environment and define similarly as *Approximate Networking* to be a networking environment in which the availability or quality of data is traded off for networking resources [32, 146]. These concepts are generally applied in resource-constraint environments, in which it is undesired or infeasible to increase network performance by adding additional communication resources. According to Betzel [32], different methods like compression [48], relaxed synchronization [127], and prediction are used to relieve the load on the network in such approximate networks. For Approximate Vehicular Networks, all of these methods are possible, but their applicability depends strongly on the considered set of vehicular applications. Different works have proposed the utilization of loss-based compression to increase the efficiency of vehicular networks [176, 194]. Similarly, relaxed synchronization has already been considered for vehicular networks for data that is not immediately required by the vehicles [100, 101]. Prediction also has been considered in vehicular networks [39]. However, a concept for Approximate Networking in vehicular networks has not yet been considered fully, which is partially justified by the generally assumed high requirements to data quality in these networks. However, these requirements do not always hold, and the limited predictability of the wireless communication link prevents a deterministic behavior of vehicular networks. Thus, it is beneficial to consider this uncertainty in the applications, such that they can deal with this uncertainty.

2.5 SUMMARY AND IDENTIFIED RESEARCH GAP

In this work, we investigate the concept of Approximate Vehicular Networks and the possible improvement compared to state-of-the-art methods. Approximate Vehicular Networks face additional challenges due to the high safety requirements of vehicular applications [97]. In this thesis, we investigate this potential conflict and develop robust and efficient communication mechanisms to enable the usage of Approximate Vehicular Networks. For this purpose, we investigate the quality of data and methods for efficient dissemination of these data, which is considered to be pivotal [16]. The data quality is then used in the form of an impact score to prioritize high-impact data. This prioritization increases the benefit of the network provided to each vehicle, but reduces the benefit of this cooperation between vehicles, as all vehicles in proximity receive the same data. While approaches in the literature commonly form clusters to coordinate the transmission and prevent this redundant transmission of data [26, 130, 196], the frequent topology changes may lead to a loss of data. This loss might also affect high-impact data, which might severely decrease the performance of an individual vehicle. To utilize cooperation in this network, we propose our concept Approximate Vehicular Networks, which adapts its properties to the impact of the transmitted data. For high-impact data, the network focuses on a highly robust and close-to-deterministic transmission, such that the negative influence of a loss of this data is prevented. For low-impact data entries, the network utilizes the bandwidth more efficiently, i. e., reduces the number of receiving vehicles. This adaptive behavior of the network increases the benefit through cooperation, while simultaneously being very robust to topology changes and messages loss for high-impact messages. With these contributions, we allow for more efficient and robust vehicular networks, that drastically outperform deterministic networks regarding the efficiency of communication.

3 LARGE-SCALE VEHICULAR COMMUNICATION NETWORKS

In this chapter, we analyze the properties of a large-scale vehicular communication network and outline the necessity for intelligent mechanisms for data aggregation and dissemination as proposed later in this thesis. In our description of issues of large-scale vehicular communication networks Section 3.1. First, we describe the components participating in this network, their interaction, and the influence of the environment. Based on the description of the issues of these networks, we describe the necessity for our developed approaches regarding approximate vehicular networking in Section 3.2. In approximate networking, reducing the data provided to a consumer, who does not need them, allows us to enhance the data availability/quality to other consumers who need them, without necessarily consuming more resources. This is achieved through Diverse Prioritization and Treatment (DPT), which filters unnecessary data to utilize the freed bandwidth for the transmission of necessary data.

3.1 ISSUES OF LARGE-SCALE VEHICULAR COMMUNICATION NETWORKS

A vehicle receives messages containing the local perception of distant vehicles to improve safety and comfort of its passengers. Based on these remote perceptions, a vehicle can react to traffic anomalies, like traffic jams early by, for example, detouring or decelerating. This message exchange is influenced by several factors, such as: (i) the environment, (ii) the vehicles in the network, (iii) entities supporting the communication between vehicles, and (iv) the available communication infrastructure. In the following, we analyze each of these factors and their influence on the network.

3.1.1 *Influence of the Vehicular Environment*

The vehicular environment is captured through a graph-based representation of the road network, in which each road can be modified through the appearance of road events, like accidents, bumps, and traffic jams. The local environment of each vehicle is measured by the vehicular sensors, which update the model of the environment stored in each vehicle. If the environment matches the model stored in the vehicle, i. e., there are no anomalies, generally no message exchange with other vehicles is necessary. If the stored model of the environment differs from the local perception of the vehicle, the vehicle shares messages with other (concerned) vehicles stating that the environment has changed. We call these changes of the environment *road events*. The receiving vehicles then update their model of the environment based on this message. While it is possible to always exchange the full perception of each vehicle, this is neither necessary nor reasonable given the limitations in bandwidth. Thus, the environment has a major influence on the vehicular network, as it influences the number of messages generated. Depending on the frequency and type of road events, different reactions may be triggered by the vehicles, e. g., change of planned path or driving behavior.

A road event is a change of the environment that will potentially affect the driving performance of the vehicles, i. e., by reducing either traffic safety or driver comfort. Examples of road events are bumps, traffic jams, and traffic signs. A road event is characterized by several attributes, such as the date of occurrence, the impacted location, a lifetime estimate, a value specifying or providing some key information about the event and others as needed. It is evident that not all of these attributes are possible to be determined by the vehicles themselves, but some of them may be inferred from historic knowledge or through aggregation. For instance, the appearance date of a road event can be estimated using the first observation of the event by a vehicle. Similarly, the lifetime can only be estimated using the lifetime of past events of similar value/-type. The value of an event can either be continuous (road temperature, traffic speed) or discrete (glace, traffic jam). Continuous variables in our model are approximated by discrete ones by dividing the value range of a continuous variable into a finite number of buckets. The resulting approximation error depends on the number of buckets and diminishes to zero for a very high numbers of buckets.

To exchange a road event, it is packed into a message, containing all attributes associated with the specific event. The resulting message is clearly context-sensitive, as it is only relevant in a certain (limited) area around the measurement location. The size of this area depends on multiple factors and may also depend on each individual vehicle: Depending on the active applications of this vehicle, the relevance of a message to the specific vehicle might vary. As an example, the size of the area that is relevant to vehicular path-planning applications can be fairly large, depending on the path and detour options available.

3.1.2 *Influence of the Vehicles*

The aforementioned road events can be detected and measured by vehicles in the proximity of the event. As the available resources and sensors of a vehicle can be fairly diverse, it is important that the heterogeneity of the vehicles be taken into consideration. This heterogeneity applies to multiple components of the vehicle, like sensor and networking and computational resources.

The sensor heterogeneity of vehicles has two different dimensions, (i) the availability of sensors, and (ii) the accuracy of the equipped sensors. According to ISO-5725, the accuracy of a measurement method is described through the terms "trueness" and "precision". "'Trueness' refers to the closeness [...] between the arithmetic mean of a large number of test results and the true or accepted reference value. 'Precision' refers to the closeness [...] between test results." [85] If a vehicular onboard sensor produces measurements of low trueness, the sensor can be considered to be broken. However, the sensors might have low precision, i. e., produce measurements with a high standard deviation. For instance, the accuracy of a measurement of a certain road event carried out by diverse sensor technologies can be fairly different. A common example is the comparison between a lidar and a camera. Although these two sensors aim at capturing the same type of road events, a lidar generally achieves higher accuracy, as it is less dependent on the weather conditions and the daylight. Similarly, the accuracy of sensors of the same technology may differ between two vehicles, as the quality of the built-in sensor equipment may be quite different for various reasons (e.g., type of vehicle or brand dependent). Thus, the different accuracy of shared measurements is a pivotal aspect and needs to be considered.

Similarly to the sensor equipment, the available network and computational resources may vary between different vehicles. As described in Section 2.1, there are different communication technologies available, like Long Term Evolution (LTE) and 802.11p. In the future, this heterogeneity might increase with the introduction of 5G to vehicular networks. While LTE and Long Term Evolution-Advanced (LTE-A) are the very commonly used for cellular communication in today's vehicular network [3, 4], there is still a lot of discussion about Vehicle to Vehicle (V2V) communication, which can either be performed via 802.11p or the cellular-based LTE-Device-to-Device (LTE-D2D)/5G. While it cannot be predicted which technology will become standard, there will most likely be one uniform technology for local communication and one uniform technology for cellular communication available, which provides additional possibilities for efficient message dissemination in future vehicular networks. In contrast, the computational resources available to the vehicles might vary greatly, especially comparing future automated vehicles with conventional vehicles. Thus, the processing of data should generally be computationally lightweight, as computational expensive operations need to be performed either at a central server (to relief the vehicles) or on the vehicles with sufficient computation resources. This poses additional challenges to our vehicular network, as the selection of suitable processing components depends on multiple factors, like the load on the individual components and time-criticality of the processed data. In this thesis, we assume that the available computation resources are sufficient to perform the validation of data at the vehicles, but also provide the possibility to shift the validation to the server. Thus, we focus on the heterogeneity of sensors and communication technologies.

3.1.3 *Influence of Supporting Entities*

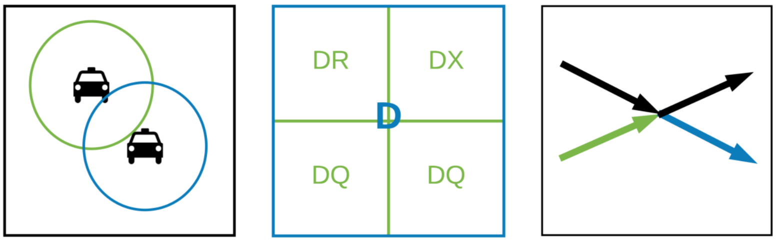
The heterogeneity regarding communication technology is a pivotal aspect of vehicular networks, as the available technologies compensate for the weaknesses of each other. However, especially the communication via the cellular network relies on additional infrastructure like cell towers, and a central server to manage the transmission process. While we do not investigate closer on the role of the cell tower, we describe the influence of the server and other support entities for vehicular networks in the following.

A central server can have different roles in the network: it can either focus purely on forwarding or provide additional services like persistent storage and processing. In the first case, the server is used as a Publish/Subscribe (Pub/Sub) broker, which disseminates messages to vehicles based on previously performed subscriptions and the vehicles' context. Thus, it only needs to monitor the context of the vehicles and forward incoming messages without any additional processing. The resource requirements of such a broker are much lower than the resource requirements of a server with storage and processing capabilities, but force the vehicles to manage the data in a decentralized manner after the reception. In the second case, the server has additional storage and processing capabilities, which it uses to manage data in a centralized manner. As an example, the server could only provide already validated messages to the vehicles to reduce the processing required by them. Thus, the vehicles rely much more on this server compared to the first case, but the quality of the shared messages is generally superior. While both approaches have their advantages and drawbacks, we generally assume a resource-less broker for the dissemination of messages to the vehicles, such that the costs for the server and its maintenance are reduced. However, the server needs to coordinate the dissemination of data to concerned vehicles.

For this purpose, the vehicles actively update their current location at the server to receive the context-sensitive messages, such that the server can provide the relevant data directly to them. However, other supporting entities have also been proposed in the literature, the so-called Road Side Units (RSUs). These RSUs are basically Wifi hotspots at the side of the road, which are capable of storing and (potentially) processing data. Due to their fixed location and short range, they can provide context-sensitive data and notify all vehicles in their proximity. Thus, the server would not necessarily need to monitor the location of the vehicles, but could also provide the data to the RSUs only. While the concept of RSUs has been frequently used in the literature, it faces some practical issues due to the potentially high deployment and maintenance costs, due to its required dense deployment to achieve a good network coverage and performance. For that reason, we will not consider RSUs in this thesis and focus on the communication of road properties between vehicles either via direct V2V communication or via the cellular network and a central server.

3.1.4 *Influence of Communication*

Communication, both vehicle-to-vehicle and vehicle-to-server, heavily influences the data available to the vehicles and, thus, the performance of the vehicular communication network. An important aspect of vehicular communication is its content-centrality [15], i. e., traditional host-based networks are generally considered inefficient in this context. For this purpose, we rely on the Pub/Sub paradigm with context-awareness and transitions, as introduced in Bypass.KOM [149], to distributed messages in the network. As LTE is a quasi-standard for mobile communication, we assume LTE as the underlying communication technology. The usage of LTE induces costs for the vehicle manufacturer or the owner. Thus, LTE should be used cautiously to keep these costs low and provide bandwidth to other applications. In fully automated vehicles, bandwidth consuming applications - especially video-streaming and gaming - should not be influenced by the exchange of road events. While this may change with the introduction of 5G, the general limitations (limited bandwidth, energy consumption, costs) will still be present, although at other levels of magnitude. However, the past has shown that the available resources are generally used to extend existing services or accommodate new ones, leading soon to a new resource shortage; that could be the case, for instance, by exchanging fine-granular sensor data between vehicles.



(a) Area-based dissemination.

(b) Geohash-based dissemination.

(c) Road-based dissemination.

Figure 1: Various data dissemination schemes.

For context-aware Pub/Sub, different methods are available to incorporate context-awareness into the system. In this thesis, we consider the location and time as possible influence factors to the dissemination of data. Possible schemes for context-aware data dissemination are depicted by Figure 1. They differ in the accurateness of their filtering and their computational overhead. The *area-based* approach is a commonly used and easy-to-implement approach, in which each vehicle shares its location with the server, and the server provides data that are inside a certain (data-dependent) area around the vehicle's location. Thus, the filtering is computationally cheap, but in certain cases such a filtering could be fairly inaccurate inducing substantial overhead by transmitting data to non-concerned vehicles. An example of such behavior can appear in a highway that is close to a small village. Information about a traffic jam in the village will most likely not be of interest to the vehicles in the highway, as they will most likely never encounter the traffic jam. The same holds for the *Geohash-based* approach, which relies on the Geohash-mechanism [133] to provide data to the vehicles, and is a special case of the area-based approach. A Geohash is a sequence of characters with a maximum of 11 characters, which can express each location on earth with a precision of up to $7.4\mathrm{cm}$. The region associated with a Geohash of length n contains the regions associated with all Geohashes of length $n + 1$, that start with that Geohash of length n. The advantage of Geohash-based compared to traditional *area-based* filtering is its natural support of different transmission ranges through Geohash's hierarchical structure. Thus, the filtering is even less complex and, thus, faster compared to area-based filtering.

Besides these universal area-based filter mechanisms, there are also filtering methods available which are tailored for vehicular networks: These approaches use *road-based* dissemination and are generally more computationally expensive, but also more accurate, i.e., the amount of unnecessarily transferred data is reduced. Under this approach, data can be provided more accurately to certain road segments, which are generally around $200\mathrm{m}$ of size. This dissemination scheme utilizes a natural property of the road network, i.e., the partially predefined movement of the vehicles on the roads. That is, the length of the shortest path towards the data location can be utilized to determine if data should be shared. However, the road-based approach still does not consider the relevance of a road event in the dissemination. This causes unnecessary transmissions to uninterested vehicles like in our previously mentioned example with the village near the highway.

In this section, we described the properties and issues of large-scale vehicular networks: the high heterogeneity of the vehicles, especially regarding their sensing capabilities, and the dissemination of messages in the network towards ensuring that vehicles receive all relevant data and do not receive irrelevant data. In the next section, we analyze the properties of these networks and motivate our work and contribution to approximate vehicular networks.

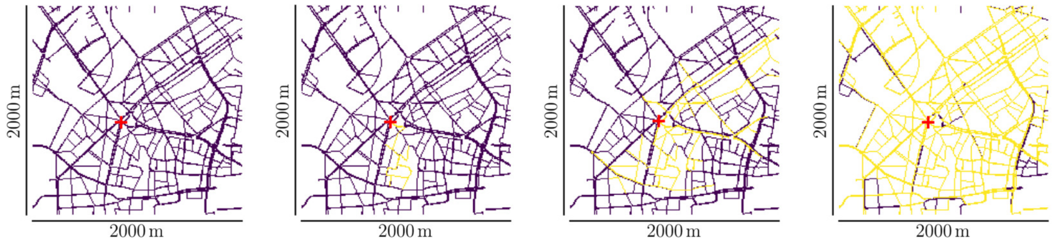
3.2 APPROACHES FOR LARGE-SCALE VEHICULAR NETWORKS

In this section, we provide a detailed insight into the consequences of the issues of large-scale vehicular communication networks described previously. For that purpose, we first discuss the influences of road properties and their measurements inaccuracy to the data dissemination in the network. Based on the insights gained there, we discuss the issue of context-sensitive data dissemination and the link to data quality. Finally, we conclude this section with a motivation of the necessity of approximate vehicular networks to efficiently exchange data between future connected vehicles.

3.2.1 *Aggregation of Measurements with Diverse Quality*

Vehicles can measure different types of road events, which have a different influence on the future behavior of the vehicles and, consequently, are of different importance in the data dissemination. The measurements of these road events are influenced by multiple factors, like weather, measurement conditions, and sensor quality. While the measurement conditions are generally similar for all vehicles in the area, the quality of the equipped sensors, as well as their availability, might vary heavily among vehicles. Thus, the measurements that are shared with other vehicles are of diverse quality, which complicates the interpretation of the provided data. In most situations, the provided data may even be contradicting regarding the current state of the road.

This contradiction can either be handled at the application layer or directly in the network: data from different participants (vehicles, servers) of the network may be aggregated to obtain the true state of the road. To perform this aggregation efficiently and correctly, it is pivotal that meta-information, like sensor quality and the associated accuracy of the provided data, be available to the aggregating unit, such that the data can be weighted accordingly. An example is the availability of two contradicting data entries, one with very high accuracy and the other with very low accuracy. It is intuitive that the high accuracy data entry generally should have a higher influence on the final aggregation result, but this entry might be very old and, thus, the real state of the road might have changed in the meantime. To resolve this conflict effectively, a data-centric approach is necessary, to aggregate the data entries considering their individual (event-dependent and sensor-dependent) properties.



(a) $p \geqslant 100\%$. (b) $p \geqslant 10\%$ (c) $p \geqslant 1\%$ (d) $p \geqslant 0.1\%$

Figure 2: Road segments (marked yellow) for which at least p of the vehicles, driving on it, will encounter the event (marked with red cross).

3.2.2 *Accurate Modeling of Vehicle's Interest*

Figure 2 displays cases of different percentages of vehicles encounter a specific event (marked with a red cross). The number of yellow road segment depends on the parameter p, which indicates if a segment should be colored yellow or purple. A segment is colored yellow if at least p of the vehicles, that are located at the specific segment when the event spawns, will encounter the event while it is active. If less than p vehicles located at the specific segment encounter the event, the color of the segment is purple.

When we analyze the number of yellow segments in the road network, it is evident that the number of yellow segments increases with decreasing p. In addition, we can observe that segments that are part of main roads or highways become yellow even for large p, while small roads are only considered if p is very low or they are in proximity to the event. If we assume that an event is only relevant for a vehicle if the event is encountered by the vehicle, we can use this visualization to determine where data about the event might be required. To this end, we can utilize this visualization to determine the vehicles interested in the reception of messages.

In general, vehicles on a yellow segment are more likely to be interested in the event than vehicles on the purple edges. Thus, only vehicles on yellow edges should be considered for receiving the event. However, as mentioned previously, the number of yellow segments, and thus the expected bandwidth used for transmitting the event to all vehicles on yellow edges, depends on the parameter p. In Figure 2a, $p = 100\%$, i. e., the event is only transmitted to vehicles who certainly require the event, which prevents unnecessary transmission of the event. This reduces the necessary bandwidth drastically, but negatively influences the dissemination of the event to interested vehicles which are distant from the event. That is, as the necessity for a transmission of the event to a distant vehicle cannot be assured. Thus, the number of vehicles that receive the event *and* are encountering it is comparably low. In contrast, Figure 2d almost ensures that every vehicle, that might be encountering the event, receives it. As the number of yellow segments in this scenario is very high, the bandwidth consumption is likewise, i. e., the event is transmitted to many vehicles Thus, most of the receiving vehicles are unlikely to encounter the event, which increases the amount of unnecessarily used bandwidth. Figure 2b and Figure 2c achieve a certain balance between these two extreme cases.

Especially when we analyze the dissemination for Figure 2b and Figure 2c, it is evident that the assumption of area-based approaches does not accurately capture the data requirements of the vehicles. While these approaches assume that the distance of a vehicle to the event estimates the interest of a vehicle, we can clearly observe that the distance to an event is not necessarily expressive to estimate the interest. As mentioned previously, the structure of the road network, especially regarding main roads and highways, has a very high influence on the possibility of a vehicle being interest in a certain event. Due to the wrong estimation the vehicles' interest by area-based approaches, unnecessary bandwidth is consumed, which could be used to notify vehicles that are more likely to encounter the event. Thus, efficient dissemination of messages requires an accurate model for the relevance of a message to a vehicle, which considers the road topology and event-specific properties like the lifetime.

3.2.3 *Necessity for Approximate Vehicular Networks*

The more accurate assessment of the relevance of an event for a specific vehicle already constitutes a first step in the direction of approximate vehicular networks, in which the performance of the network is increased without increasing the available communication resources. However, in a general network, not only the total bandwidth is restricted, but in which also different types of events are disseminated. The available bandwidth needs to be shared between these event types, such that the network performance is maximized. For this purpose, not only the relevance of the event and the accuracy of the measurement are important, but further the impact that a certain piece of data has on the vehicular network. As an example, events of high impact, like accidents, should generally have more bandwidth available than events of low impact, like small changes in the traffic flow.

When an event is disseminated, it generally cannot decrease the performance of the vehicle, i. e., a transmission can never decrease the performance of the vehicle, unless other (more important) events cannot be transmitted due to that transmission. In combination with the limited bandwidth a vehicle can utilize, this opens very interesting possibilities: Vehicles may filter events with low impact to receive high-impact data, i. e., the vehicle may prioritize high-impact events. While this approach is common for networks without cooperation between vehicles, the prioritization of high-impact events decreases the potential of cooperation. That is, as all vehicles, that could exchange messages via Wifi, will generally receive the same events redundantly, i. e., the performance gain of cooperation would be low. Additionally, coordinated mechanisms for the cooperative reception of messages suffer from the frequent topology changes in a vehicular network.

For this purpose, we employ the concept of approximate vehicular networks, in which network and communication resources are traded with data availability and quality. Specifically, we focus on approximate vehicular networks that rely on probabilistic mechanisms to modulate the redundancy to improve the performance of the network. That is, we decrease the redundancy for high-impact events, with the potential consequence of missing some events, to free bandwidth for the reception of previously not received events. This seems to be counter-intuitive for vehicular networks, as events might have a high impact to the receiving vehicle. However, our concept for approximate vehicular network performs deterministically for very high-impact events, while becoming increasingly probabilistic with decreasing impact of an event. This increases the flexibility and bandwidth efficiency of vehicular networks drastically, as the removed determinism of the network offers new possibilities to the data processing and transmission in this network.

4 DATA QUALITY ASSESSMENT FOR VEHICULAR NETWORKS

Based on our findings in the state-of-the-art in Chapter 2, in this chapter we analyze the vehicle-generated data shared in a vehicular network and outline possible influence factors to data quality and impact. We propose a holistic approach to determine the impact of data in a vehicular network, considering the participating entities in the generation and dissemination of data: (i) the measuring sensors, (ii) the context of the receiving vehicle, and (iii) the data itself. This impact is an abstract measure of the increase (or decrease) in performance of the vehicle's behavior, like an increase/decrease in travel time, fuel consumption, or driver comfort. It is an important foundation for our proposed approach towards approximate vehicular networks in Chapter 5, as it provides an assessment of the influence of data, which is utilized to adapt the proposed probabilistic mechanisms in order to maximize efficiency.

In the following, we start with a description of the scenario in Section 4.1. In Section 4.2, we describe our innovative aggregation scheme for improving the quality of data from different sources (vehicles) considering data-specific properties to reduce the amount of false data. In Section 4.3, we propose a mechanism to determine the relevance of a piece of data to a specific vehicle. Our approach predicts the vehicle's future context and the state of the road event contained in the data to determine the necessity to transfer a particular piece of data to a particular vehicle. We then propose our holistic approach for assessing the specific-impact of a piece of data for a vehicle, which combines our findings of Section 4.2 and Section 4.3 to determine the specific-impact for a vehicle in Section 4.4. The specific-impact is then used in Chapter 5 to improve the efficiency of the vehicular network.

4.1 SCENARIO DESCRIPTION

Figure 3 displays the considered scenario for the data assessment in vehicular networks. In this scenario, vehicles on the right (vehicles 1 and 2), which are in the proximity of an event, measure it and generate a message $\vec{d}$ containing this measurement $\vec{m}$. Due to the potential inaccuracy of the measuring sensors, the measurements generated by these vehicles might be contradicting. The generated measurements are then shared with the vehicles on the left (vehicles 3 and 4), which are distant from the event. However, only vehicle 3 needs to receive the measurements, as vehicle 4 takes an exit and will never encounter the event. Thus, the goal of this chapter is to determine decide on the correct aggregate for vehicle 3 based on possibly contradicting measurements, and to determine which vehicles need to be notified with the measurements of the event.

For the aggregation, not only the measurement $\vec{m}$ itself, but also a description of the data type w is provided. Additionally, the sender includes meta-information about the measurement conditions in the provided message as shown in Equation 1, like the

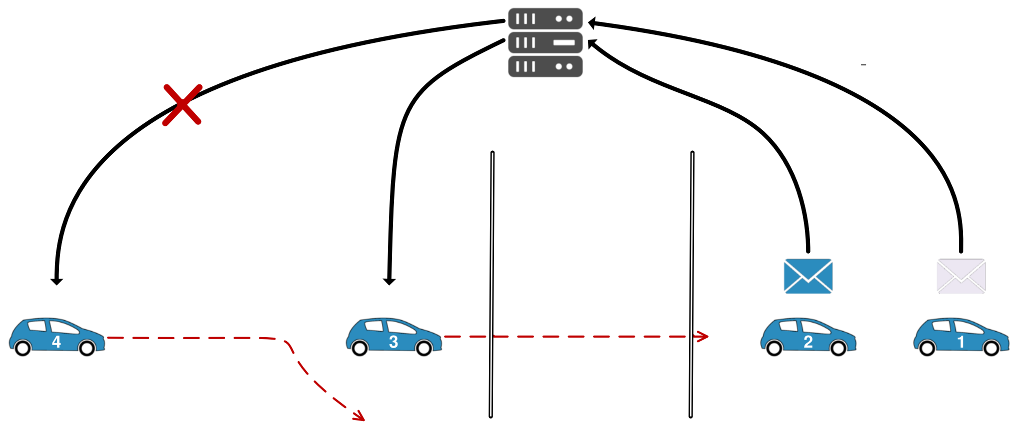


Figure 3: Visualization of the scenario for data assessment in vehicular networks.

measurement location $\vec{l}$ and time t to increase the usability of data for the receiving vehicle.

$$\vec{d} = \begin{pmatrix} \vec{m} \\ w \\ t \\ \vec{l} \end{pmatrix} \tag{1}$$

Based on these data, the receiving vehicle aims at determining the true value of the underlying variable, which is only described by the received measurements. For this purpose, the vehicle aggregates the received data. In Section 4.2, we investigate the efficient handling of possibly inaccurate sensor data and propose an aggregation scheme for sensor data to reduce the number of false aggregates in the vehicular network. In previous work, we have analyzed the necessity for a correct aggregate before encountering the event, like in the case of vehicular path-planning [120].

It should be noted that the amount of improvement of the vehicle's decision-making is affected by the limited network resources. For this purpose, it is pivotal to assess the relevance of an event for a vehicle (see Section 4.3) before consuming network resources for transmitting it to the vehicle. An event is considered relevant to a vehicle if it influences the decision-making of that vehicle. In this work, we assume that the decision-making is influenced only if the vehicle is in the proximity of the event while the event is active.

To decide on which vehicles to notify, the relevance of an event to a specific vehicle needs to be determined. This relevance accounts for the uncertainty regarding the future path of the vehicle and the lifetime of the event. Although the paths might be known to the vehicle itself, disclosing themto the server is a severe intrusion into the passengers' privacy. Additionally, the lifetime of an event is considered to be unpredictable. Thus, statistical methods are required to predict the relevance based on the probability that the vehicle passes by the event location and the event is active. To determine this probability, the future path of a vehicle and the lifetime of the event will need to be predicted first.

Based on the quality of the measurement and the relevance of the event to the receiving vehicle, we design a model for the impact of a measurement on a vehicle in Section 4.4. This impact will be critical for the effective handling of certain measurement data, as it will determine the usefulness of such data to a vehicle by balancing properly the quality and relevance of such data.

4.2 sensor-specific influence on data impact

To assess the impact of a piece of data (referred to also as message), the quality of the sensor generating the measurement $\vec{m}$ is taken into consideration, along with the complete representation of a measurement in $\vec{d}$. While the quality of the measuring sensor has no essential importance for today's vehicular applications, it is a pivotal aspect if multiple (possibly contradicting) values of the same underlying variable are sensed from multiple vehicles with different sensor setups. Each measurement might have a different accuracy, which needs to be considered in the aggregation of these measurements. In this case, the vehicle needs to decide on the true state of the underlying variable, considering that (i) some of the measurements might be erroneous and (ii) the underlying variable might have changed between the measurements. To provide the necessary meta-information for this decision to the receiving vehicles, we first employ a model for inaccurate measurements in Section 4.2.1. Based on this model, we describe basic methods for the aggregation of measurements in Section 4.2.2, which considers measurements with the same detection times and measurements with different detection times. The latter increases the complexity of the aggregation process, as the underlying variable might have changed between two measurements. As existing aggregation schemes do not appropriately consider this factor, we propose our innovative aggregation scheme in Section 4.2.3, which considers the accuracy of measurements and the lifetime of events to adapt the influence of old measurements on the aggregate. With this approach, the possibility of false measurements and a change of the underlying variable are considered in the aggregation, so that false aggregates are reduced.

4.2.1 *Model for Inaccurate Measurements*

As motivated in Chapter 3, the high heterogeneity of the available sensors and their quality leads to measurements of different accuracy. This creates additional issues for the interpretation of messages, especially when several (contradicting) messages have been received. Thus, the meta-information about the message accuracy needs to be preserved throughout the dissemination of the message and in all processing steps.

For that purpose, a measurement $\vec{m}$ is represented as a vector with n_w entries (states), one for each possible state of the underlying variable. Notice that continuous variables are divided into buckets before, such that a sufficient level of granularity is reached. In this state probability vector representing a measurement, each entry describes the (conditional) probability that the underlying variable is in a certain state, given the specific physical measurement by the sensor, as shown in Equation 2. For simplicity, we do not show the dependence of this probability on the measurement in the notation.

$$\vec{m} = \begin{pmatrix} p_1 \\ \vdots \\ p_{n_w} \end{pmatrix} \tag{2}$$

for which holds

$$\sum_{\psi=1}^{n_w} p_\psi = 1$$

and

$$p_\psi \geqslant 0, \forall \psi \in \{1, \ldots, n_w\}$$

The properties of a measurement $\vec{m}$ strongly depend on the quality of the installed sensor and the environmental conditions. Since the quality of a measurement is also affected by the capabilities of the specific sensor, the standard deviation of the measurements generated by the specific sensor is encoded in the measurement $\vec{m}$; this standard deviation is recorded in the sensor's datasheet. In general, multiple vehicles that use multiple sensors to detect a road event can provide more accurate measurements. The standard deviation $\sigma_{\vec{m}}$ is defined according to Equation 3 with $\overline{\psi}$ being the expected value of a measurement. Notice that $\overline{\psi}$ is not necessarily a real state, but might also be between two states.

$$\sigma_{\vec{m}} = \sum_{\psi=1}^{n_w} p_\psi \cdot \left(\psi - \overline{\psi}\right)^2 \tag{3}$$

In case of a very inaccurate measurement, the probabilities are similar, as the sensor cannot exclude any states. In this case, all probabilities could be $p_\psi = {}^1/_{n_w}$ and $\sigma_{\vec{m}}$ would become very high. For an optimal measurement, $\exists \psi \mid p_\psi = 1 \wedge p_i = 0, \forall i \neq \psi$. In this case, the standard deviation $\sigma_{\vec{m}} = 0$ and the expected value corresponds to the only state with a non-zero probability. The accuracy of the measuring sensors is considered in the aggregation of measurements using the aforementioned representation of a measurement.

4.2.2 *On the Aggregation of Inaccurate Measurements*

In general, multiple vehicles measure the same underlying event and share their measurements with others. Due to the diverse sensor quality, these measurements might be very diverse and lead to contradicting decisions as to what the true state of the underlying variable is. In such a case, a decision on the real state of the event needs to be made by using (or properly aggregating) these contradicting measurements. In the following, we first introduce the aggregation of measurements that have been measured at the same time. After that, we consider the aggregation of measurements that have been measured at different times.

4.2.2.1 *Aggregation of Measurements with the Same Measurement Time*

Consider the merging of two measurements, $\vec{m}_1$ and $\vec{m}_2$, that have been measured at the same time and location, and are of the same type. For this purpose, conditional probabilities are used. Due to the spatio-temporal similarity of the measurements, the vehicles have observed the same underlying event, i. e., the underlying variable is in the same state for both measurements. Equation 4 displays the probability that the underlying variable is in a certain state ψ_1.

$$P(\psi = \psi_1) = P(\psi_{\vec{m}_1} = \psi_1 \wedge \psi_{\vec{m}_2} = \psi_1 \mid \psi_{\vec{m}_1} = \psi_{\vec{m}_2}) \tag{4}$$

Notice that the measurement of vehicle 1 is independent of the measurement of vehicle 2. Thus, the aggregate $\vec{m}_a$ can be calculated according to Equation 5, where p_i refers to a probability that ψ_i is the correct state in $\vec{m}_1$ and q_i refers to a probability that ψ_i is the correct state in $\vec{m}_2$.

$$\vec{m}_a = \frac{\begin{pmatrix} p_1 \cdot q_1 \\ \vdots \\ p_{n_w} \cdot q_{n_w} \end{pmatrix}}{\vec{m}_1 \cdot \vec{m}_2} \tag{5}$$

Thus, a vehicle or the server can aggregate data considering their individual accuracy level, which is reflected in the probability vector representing a measurement. Notice that this does not include the false reporting of data by malicious or non-cooperative vehicles. We performed some separate analyses and showed that malicious vehicles can be detected and removed from the vehicular network [131].

This aggregation assumes that the measurements are describing the same underlying event, i. e., are performed at the exact same time, which is not commonly happening in vehicular networks. This issue is addressed by modeling the aging of measurements in a vehicular network, which enables the aggregation of measurements with different detection times.

4.2.2.2 *Aggregation of Measurements with Diverse Measurement Times*

If data is not measured at the same time, the measurements may observe a different state of the underlying variable, i. e., the previous conditional probability cannot be applied directly. To aggregate two measurements of different times, any change change of the underlying variable over the duration between the measurements needs to be considered. A discrete-time model is utilized, in which data is aged by 1s in each discrete time instant.

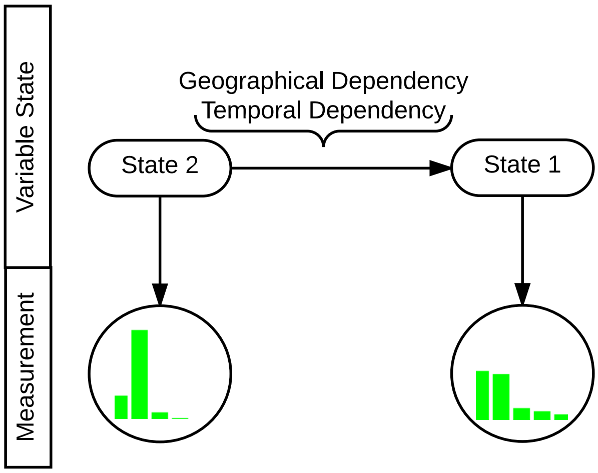


Figure 4: Sample HMM utilized for the short-term aging of data. Taken from [118].

To predict the change of the underlying variable, HMM, as depicted by Figure 4, can be used, which models both the aging and sensing of data in the vehicular network. Every event type is modeled according to an HMM with its own underlying Markov Chain, such that prediction of value changes is possible. The parameters of the Markov Chain can be estimated based on measurements gathered by the vehicles. In order to apply the aging to the probability vector, the transition matrix $\mathbb{T}(t)$ of the HMM is required. This transition matrix $\mathbb{T}(t)$ captures the possible state transitions from every state over a certain time t. Each entry $\mathbb{T}_{ij}(t)$ in the transition matrix is equal to the probability that the event switches to j after t, if it is previously in i. Thus, the aged probability vector $\vec{m}(t)$ can be retrieved by multiplying transition matrix $\mathbb{T}(t)$ and the original vector $\vec{m}$ as shown in Equation 6.

$$\vec{m}(t) = \mathbb{T}(t) \cdot \vec{m} \tag{6}$$

According to the associativity of matrix multiplications, we can obtain $\mathbb{T}(t)$ by multiplying the transition matrix for one timestep $\mathbb{T}(1)$ t-times by itself. To aggregate two measurements taken at different times, the future state of the older measurement needs to be estimated, i. e., it needs to be aged by the difference of measurement times to the newer measurement. The aged measurement can then be aggregated with the newer measurement using Equation 5, as the two measurements then describe the underlying variable at the same time.

When we extend this approach to more than two measurements by aggregating additional measurement with the aggregate, we can observe that is really robust to false measurements, as all measurements are considered until they exceed their lifetime. However, this also reduces the adaptability to a change of the environment due to the influence of old measurements.

4.2.3 *Proposed Adaptive Aggregation Scheme for Inaccurate Measurements*

The aggregation of data using Markov Chains performs well if the data is not changing frequently, as it adapts slowly to environmental changes. After an environmental change, the majority of the measurements in a vehicle's cache have been generated before the event change. As these measurements decrease in influence over time and the number of fresh measurements increases, the aggregate does adapt to the change eventually. However, it generally takes a long time for the aggregate to adapt to this change, which decreases the performance of the approach in this setting.

To alleviate this issue, we propose our adaptive aggregation scheme, which considers the required robustness and adaptability of an event type in the aggregation to reduce false aggregates. Robustness is defined as the ability of the aggregation scheme to alleviate false measurements, while adaptability is defined as the ability to adapt quickly to changes of the environment. To increase the robustness of an aggregation scheme, generally more (potentially old) measurements may be considered in the aggregation process, such that the aggregate is less prone to false measurements. In contrast, the adaptability of an aggregation scheme is increased if only a few (or no) old measurements are considered, such that an adaptation to an environmental change can be performed immediately. It is evident that the requirements for robustness and adaptability are contrary, i. e., a tradeoff between these two is necessary for the design of our aggregation scheme. This trade-off depends on the average accuracy $\overline{p}_a$ of the provided measurements and the expected lifetime T of the event. If the data provided by the vehicles is always error-free, the system is intrinsically robust and does not require the consideration of additional (past) measurements, i. e., it can adapt to updates immediately. If the data type has a very long lifetime, the system is not required to react to an environmental change and, thus, can be very robust.

4.2.3.1 *Robustness vs. Adaptability*

To find this tradeoff between robustness and adaptability, the expected number of false aggregates per timestep n_{total} needs to be reduced, as described in Equation 7. The parameter t is the time window, after which an adaptation of the aggregate is performed given measurements in a different state.

$$\min n_{total}(t) \tag{7}$$

$n_{total}(t)$ is shaped by the two possible causes for false aggregates, (i) the number $n_{\text{rob}}(t)$ of false aggregates due to an adaptation to a series of false measurements (missing robustness) and (ii) the number $n_{\text{ada}}(t)$ of false aggregates due to a late adaptation to a changing environment (missing adaptability), i. e., $n_{\text{total}}(t) = n_{\text{rob}}(t) + n_{\text{ada}}(t)$. We now provide a definition for $n_{\text{rob}}(t, r)$ and $n_{\text{ada}}(t)$, which we then utilize to derive the time window t_{opt}, for which the number of false aggregates is minimized.

To analyze the number $n_{\text{rob}}(t, r)$ of false aggregates due to robustness, the average accuracy $\overline{p}_a$ of a sensor is analyzed, which refers to the probability that the correct state is the state with the highest probability. Based on $\overline{p}_a$, the average number of false aggregations per timestep $n_{\text{rob}}(t, r)$ due to missing robustness can be derived using

Equation 8. It refers to the probability that all measurements since a time t before the current time are correct, and an adaptation to a false state would be performed in this timestep. For this purpose, the number of messages during that period t is estimated based on the average message rate r. This message rate r is assumed to be constant, which is a necessary assumption for our optimization problem, which is addressed later.

$$n_{rob}(t, r) = (1 - \overline{p}_a)^{\frac{t}{r}} \tag{8}$$

Similarly, the average number of false aggregations per timestep $n_{ada}(n)$ due to missing adaptability can be derived using Equation 9. In utilizes the expected lifetime T to estimate the change that is expected over a certain period t, i. e., there is a chance of $1/T$ for a change of the environment per timestep, for which the aggregation scheme needs t to adapt. Notice that the changes in different timesteps are not independent of each other, but a change in timestep 2 requires that the variable has not changed in timestep 1.

$$n_{ada}(t) = \sum_{i=0}^{t} \left[\left(\prod_{j=0}^{i} \frac{T-j}{T} \right) \frac{1}{T} \right] \tag{9}$$

However, for $t \ll T$, Equation 9 can be approximated using the much simpler Equation 10, which is used in the following. $t \ll T$ is generally true, as an adaptation needs to be performed much faster than the change of the underlying value, otherwise the aggregation will always deliver wrong results. If T is itself very low, then the size of the timestep is inappropriate and needs to be adapted for that event, such that t can be chosen much smaller than T.

$$n_{ada}(t) \approx \frac{t}{T} \tag{10}$$

n_{total} is a combination of n_{rob} and n_{ada}, which are generally equally important. As n_{total} should be minimized, the partial derivative of n_{total} with respect to t needs to be 0 as shown in Equation 11.

$$\frac{\partial n_{total}}{\partial t} = (1 - \overline{p}_a)^{\frac{t}{r}} \cdot \ln(1 - \overline{p}_a) + \frac{1}{T} = 0 \tag{11}$$

After some transformations, the optimal value for the time until adaptation t_{opt} can be obtained as shown in Equation 12, which states after which time an adaptation shall be performed given the current rate of incoming messages r. Due to the exponentially decreasing $n_{rob}(t)$ and the linearly increasing $n_{ada}(t)$, the derived solution is always a minimum.

$$t_{opt} = \frac{\ln\left[-\frac{1}{\ln(1-\overline{p}_a)\cdot T}\right]}{\ln(1-\overline{p}_a)} \cdot r \tag{12}$$

One way to consider the time interval t_{opt} in the aggregation is to adapt the aggregate only if all measurements over this time interval are similar. However, there are two main issues with that approach, (i) for data types with low average accuracy, the aggregate would almost never change to a new value, as the low sensor accuracy makes a series of correct measurement improbable, and (ii) time between two messages is generally not constant. To alleviate this issue, weighting function $f_w(t)$ is used, which weights messages according to their age. The purpose of the weighting function $f_w(t)$ is to adapt the robustness and the adaptability of the aggregation. This function modulates the aggregation such that a value change of the aggregate is performed if the same message arrives for a time t_{opt} with a constant rate r, but an aggregation scheme using this function can still handle varying message rates and false measurements. The function $f_w(t)$ depends on the type w of the data, as it requires the data type dependent t_{opt} and T. The idea is to utilize the expected lifetime T of an event to reduce the weight of a measurement in the aggregation process accordingly. Thus, the weight of a fresh measurement is 1 (Equation 13), while the weight of an outdated measurement (higher age than the expected lifetime) is 0 (Equation 14). The latter also reduces the fluctuation of the aggregate when measurements are invalidated, as their influence degrades slowly to 0 previous to their invalidation.

$$f_w(0) = 1 \tag{13}$$

$$f_w(T) = 0 \tag{14}$$

While these two special cases of $f_w(t)$ are evident, the behavior of $f_w(t)$ between 0 and T needs to be determined. For this purpose, an exponential function with three configuration variables, a, b, and d is utilized, as shown in Equation 15, i. e., an exponential function that can be scaled along the x-axis and y-axis and translated along the y-axis.

$$f_w(t) = a \cdot e^{bt} + d \tag{15}$$

Based on Equation 13 and Equation 14, the parameters a and d can be derived to obtain the family of functions shown in Equation 16. The derivation of this family of functions is shown in Section A.1.

$$f_w(0) = a + d = 1$$

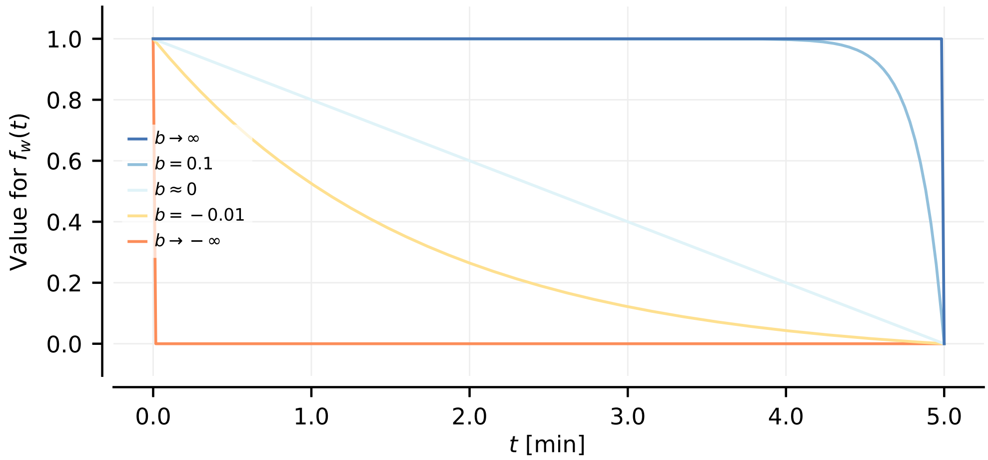


Figure 5: Aging function f_w for different parameters b with $T = 5$min. Adapted from [121].

$$f_w(T) = a \cdot e^{bT} + d = 0$$

$$f_w(t) = \frac{e^{bt} - e^{bT}}{1 - e^{bT}} \tag{16}$$

This family of functions depending on the parameter b as shown in Figure 5, and each function assigns different weights to old measurements. On the y-axis, the weight of a data of a certain age is displayed. For high values of b, the weight of data is relatively constant over time and drops drastically to 0 at the end of the expected lifetime. A high value of b would be used for very inaccurate or constant data types, in which old data can improve the performance of the system. For low values of b, the weight of data decreases drastically with increasing age. Thus, low values of b are suitable for frequently changing or accurately measured data types. As the exact value of b depends on the utilized aging function, we first describe the aging function based on the weight metric.

4.2.3.2 *Weight-based Long-Term Aging*

In order to use the weighting function in our aggregation process, the influence of a certain measurement to the aggregate over time needs to be determined. For this purpose, the probability vector $\vec{p}_0$ with $|\vec{p}_0| = |\vec{m}|$ is used as shown in Equation 17, which has no influence on the previously presented aggregation process, and thus should be equal to a measurement aged by its expected lifetime T. $\vec{p}_0$ has no influence to the aggregate, as every probability in $\vec{p}_0$ is equal, and the aggregation via conditional probability will always return the other input as result of the aggregation.

$$\vec{p}_0 = \begin{pmatrix} \frac{1}{n_w} \\ \vdots \\ \frac{1}{n_w} \end{pmatrix} \tag{17}$$

To obtain $\vec{p}_0$ after the expected lifetime of the event, the aging of any measurement $\vec{m}$ with the transformed transition matrix $\mathbb{T}^*(T)$ needs to output $\vec{p}_0$. For this purpose, the original transition matrix $\mathbb{T}(t)$ needs to be transformed to $\mathbb{T}^*(t)$ such that aging any probability vector with $\mathbb{T}^*(T)$ always leads to $\vec{p}_0$. This transition matrix always distributes the probability of each state equally to all the other states, which leads to $\vec{p}_0$, and is shown in Equation 18.

$$\mathbb{T}^*(T) = \begin{pmatrix} \frac{1}{n_w} & \cdots & \frac{1}{n_w} \\ \vdots & . & \vdots \\ \frac{1}{n_w} & \cdots & \frac{1}{n_w} \end{pmatrix} \tag{18}$$

Between 0 and T, the transition matrix is adapted, i.e., $\mathbb{T}^*$ is combined with the original transition matrix $\mathbb{T}(t)$ weighted by $f_w(t)$. Thus, every entry $\mathbb{T}^*_{mn}$, where m refers to the column and n refers to the row, is calculated as shown in Equation 19.

$$\mathbb{T}^*_{mn}(t) = \mathbb{T}_{mn}(t) \cdot f_w(t) + \frac{1}{n_w} \cdot (1 - f_w(t)) \tag{19}$$

Based on the modified transition matrix $\mathbb{T}^*(t)$, data is aged considering the parameter b. b can now be determined such that the aggregation performs an adaptation after t_{opt}, given a rate of incoming messages r.

4.2.3.3 *Determining the Aging Function* $f_w(t)$

As shown in the optimization problem that minimizes the number of false aggregates, an adaptation should be performed after t_{opt}. Thus, we need to choose $f_w(t)$ such that an adaptation to a new value is performed after t_{opt}. For this purpose, our aggregation scheme simulates the incoming messages based on the message rate r for the expected lifetime of a message T. Older messages do not need to be simulated, as they are invalidated after exceeding their expected lifetime. In the time between T and t_{opt}, measurements of a certain state ψ_i are incoming, which then changes in the time between t_{opt} and 0 to a state ψ_j. The messages themselves are generated based on the average accuracy $\overline{p}_a$, that is given to the system, and assuming a Gaussian distribution of noise. Our aggregation scheme starts with $b = 0$ and observes the behavior of the aggregation. For the optimal b, an adaptation of the aggregate value is performed *exactly* with the last incoming measurement. If not, there are two possibilities: (i) there is no change in the value at all, i. e., the adaptation is performed too late and, thus, b is too high (ii) there is a change at or prior to the second last measurement, i. e., the adaptation is performed too early and, thus, b is too low. If b is not optimal, our aggregation scheme adapts b (reduce if it is too high, increase if it is too low) and repeats the process, until a value for b is found that matches the adaptation criteria. However, the determination of b still relies on the initial state ψ_i and the destination state ψ_j, which may have a major influence to the final value of b. As neither i nor j are known with certainty, our aggregation scheme needs to determine b for every possible combination of i and j. To determine the overall value for b, the determined values for b are weighted according to their transition probability from i to j, which is derived from the underlying Markov Chain. This enables us to optimally aggregate data to achieve data of higher quality and increased usability for the receiving vehicle.

In this section, we proposed our innovative aggregation scheme, which dynamically adapts to the data that is being aggregated. That is, it considers both the accuracy of the aggregated data and an estimate of the lifetime to adapt the influence of old measurements in the aggregation process. In Section 7.2, we describe the performance of our aggregation scheme under varying environmental conditions. We show that our aggregation scheme reduces false aggregates drastically and adapts its robustness and adaptability to the requirements given through the data properties. While this section focused on the measurement $\vec{m}$ itself, the next section analyzes the influence of the consumer context and its connection to the meta-information like the location $\vec{l}$ and the measurement time t.

4.3 RELEVANCE OF DATA FOR THE RECEIVING VEHICLES

Data is not important for all vehicles in the network, as some vehicles may not require certain pieces of data due to their current and future context. Context-sensitive data is filtered by comparing the location of the data $\vec{l}$ with the current location $\vec{l_v}$ of the vehicle v. For this purpose, several methods, like area-based or region-based filtering [63], are available, which have already been described in detail in Section 3.1.4. However, these methods commonly assume that the linear distance is the only metric capturing the interest of a vehicle. This assumption is flawed in many cases, as a message is often only required if the vehicle is actually traversing the location of the message (compare Section 3.2.2). An example is an accident in a village which is close to a highway. Although the accident is really close to the highway, it is most likely not encountered by many vehicles, as the majority of vehicles generally stays on the highway. Thus, it is not reasonable to provide the data to all vehicles on the highway, but only to the ones that leave it and are, thus, much more likely to encounter the accident.

In this section, we propose the utilization of the future context of the vehicles and future event state to determine the relevance of an event to a vehicle. This relevance needs to be determined by the server disseminating the event, such that the relevance can be used to improve the network efficiency. To determine the relevance, the distributing server needs to not only consider the current location of the vehicle and the location of the event but also needs to consider the probability of the vehicle encountering the event. This probability can be calculated based on two factors: (i) the probability that the vehicle will traverse the event location at some point in the future, and (ii) the probability that the event is still active when the vehicle traverses the event location. To predict the probability of the vehicle traversing the event location, the server utilizes the planned path of the vehicle. However, the planned path needs to be known to the server disseminating the data in the network, which is not always the case due to privacy considerations. If the planned path is unknown, the server will need to predict the planned path based on historic knowledge about the traffic flow. The server considers the probability that the vehicle encounters an event in conjunction with the probability that the event is active by developing a set of recursive equations. These equations describe the probability that the vehicle traverses along a certain road segment s_j, given it is currently located at the end of road segment s_i, and that the event is still active when the vehicle arrives at the end of s_j. For this purpose, the server also predicts the lifetime of the event based on historic data about the average event lifetime.

While we focus on the definition of relevance based on events at fixed locations in this thesis, we also analyzed the relevance of data for other applications like cooperative maneuver coordination [33], in which location data of possible cooperation partners are disseminated to concerned vehicles.

4.3.1 *Relevance Assessment Assuming Knowledge of the Vehicle's Path*

If the server knows the path $\phi = \{s_1, \dots, s_n\}$ consisting of road segments that a vehicle plans to traverse, it can easily determine if the vehicle will, at some point in time, traverse the segment s_e containing the event. As discussed previously, the event may only be relevant if $s_e \in \phi$, i. e., the vehicle traverses this segment. If $s_e \notin \phi$, the relevance is 0. However, even if the vehicle traverses the event location, it is not guaranteed that the vehicle encounters the event, as the event might have turned inactive in the meantime. Thus, the server needs to check if the event is still active by the time the vehicle arrives, which is generally unknown to both the vehicle and the server. To assess the state of the event by the time the vehicle passes, the server utilizes historic knowledge to predict the lifetime of the event. Based on this historic knowledge, the prediction of the event state is based on the distribution function $p_w(t)$, which depends on the event type w. This function returns the probability that the event is still active after a certain time t, which is monotonically decreasing. In the following, we assume that the lifetime t is exponentially distributed. Thus, p_w is the complementary cumulative distribution function of this exponential distribution based on the expected lifetime T as shown in Equation 20.

$$p_w(t) = e^{-\frac{t}{T}} \tag{20}$$

Notice that Equation 20 determines the probability that the lifetime of the event exceeds t. To determine the probability of the event being active when the vehicle encounters it, the server utilizes the travel time $t(\phi, s_e)$ of the vehicle. $t(\phi, s_e)$ returns the time to get to the end of the segment containing the event s_e along the path ϕ starting at the end of the first segment in the path. $t(\phi, s_e)$ returns ∞ if $s_e \notin \phi$. The relevance $R(s_v, s_e)$ is defined as the probability that the vehicle encounters the event located at the segment s_e, thus, the relevance for a vehicle driving on segment s_v is defined as shown in Equation 21.

$$R(s_v, s_e) = p_w(t(\phi, s_e)) \tag{21}$$

4.3.2 *Relevance Assessment Assuming No Knowledge of the Vehicle's Path*

If the server has no knowledge of the path a vehicle plans to take, it can only rely on historic knowledge to predict the path the vehicle is likely to follow. For this purpose, a set of recursive equations is developed, defining the probability of a vehicle encountering the event, given the current road network S, the vehicle's current location s_v, and the event location s_e. In the road network graph, each end of each segment is represented as an edge, and the connection between segments (i. e., intersections) are represented as edges in the graph, with their respective transition probability. This probability p_{ij} is the joint probability that (i) the vehicle moves from s_i to s_j, and (ii) that the event is still active after the vehicle has moved. Thus, p_{ij} is shaped by spatial and temporal dimensions. For the *spatial* dimension, the server uses the path probability p_{ij}^l, i. e., the fraction of vehicles at the end of s_i to drive over s_j. We assume that p_{ij}^l does only depend on the adjacent segments s_i and s_j, thus, it does not capture the possibility of multi-edge dependencies like paths, but is much easier to store, determine, and process. For the *temporal* dimension, the server uses the probability that the event is still active when the vehicle arrives at the end of s_j based on Equation 20. The travel time along every road segment can be considered separately, due to the memoryless property of the exponential function lifetime function, i. e., $p_w(t + \Delta) = p_w(t) \cdot p_w(\Delta)$. Thus, the server uses the travel time $t(s_1, s_2)$ between two segments s_1 and s_2, which is the time required to travel from the end of s_1 to the end of s_2. The probability p_{ij} can be calculated according to Equation 22, as the spatial and the temporal dimension are independent.

$$p_{ij} = p_{ij}^l \cdot p_w(t(s_1, s_2)) \tag{22}$$

While the event to traverse from s_j to s_k requires the vehicle being in s_j, it does not account for how the vehicle got there. Thus, the pair of probabilities p_{ij} and p_{jk} are independent of each other. Consequently, the probability to get from s_i to a non-adjacent segment s_k over the connecting segment s_j, given that there is only one valid path between s_i and s_k, can be calculated according to Equation 23.

$$p_{ik} = p_{ij} \cdot p_{jk} \tag{23}$$

Algorithm 1 : Determining encounter probability for every segment s_i and the event edge s_e

Result : $p(i,e), \forall s_i \in S$

$p(i,j,e) \leftarrow 0, \forall s_i, s_j \in S \mid s_j \in \text{neighbors}(s_i)$;
$p(e,j,e) \leftarrow 1, \forall s_j \in \text{neighbors}(s_e)$;
$n \leftarrow 0$ **while** $n < |S|$ **do**
  for $s_i, s_j \mid s_j \in \textit{neighbors}(s_i) \wedge s_i \neq s_e$ **do**
    $p(i,j,e) = p_{i,j} \cdot \sum_{s_k \in \text{neighbors}(s_j)} p(j,k,e)$;
  end
end
for $s_i \in S$ **do**
  $p(i,e) = \sum_{s_j \in \text{neighbors}(s_i)} p(i,j,e)$;
end

If there are multiple paths connecting s_i and s_k, one over s_{j_1} and one over s_{j_2}, the probability of traversing s_k when starting at s_i can be calculated using Equation 24. This combination of probabilities is possible as two paths are excluding each other, a vehicle can take either one or the other.

$$p_{ik} = p_{ij_1} \cdot p_{j_1k} + p_{ij_2} \cdot p_{j_2k} \tag{24}$$

Let $p(i,e)$ denote the probability of a vehicle at road segment s_i to encounter the event located at s_e. Then, this probability can be defined as a recursive function using the probability of the neighbors to encounter that event as shown in Equation 25.

$$p(i,e) = \begin{cases} \sum_{s_j \in \text{neighbors}(s_i)} p_{ij} \cdot p(j,e) & i \neq e \\ 1 & i = e \end{cases} \tag{25}$$

The solution of this set of recursive equations can only be estimated due to the presence of cycles in the road network and the size of the road network. To estimate the solution for the recursive equations, we develop Algorithm 1, which is very similar to Bellman-Ford shortest paths algorithm [29, 70] and updates the probabilities each round, such that the probabilities considering paths with maximum of n hops are determined correctly.

The server needs to execute Algorithm 1 once per road event that should be transmitted. While the worst-case runtime can be very large for large-scale road networks, the relevant probabilities can be determined much faster terminating the algorithm if the expected improvement of the result is very small. That is, if there are no changes performed to probabilities above a predefined threshold in one round, i. e., it is likely that the induced error by early termination is small. The relevance score $R(s_v, s_e)$ for a specific vehicle v is equal to the probability $p(i,e)$ for the segment s_v associated with the current location of the vehicle and is shown in Equation 26.

$$R(s_v, s_e) = p(s_v, s_e) \tag{26}$$

In this section, we proposed our novel approach towards assessing the relevance of data for a specific vehicle. We consider a piece of data to be relevant if the vehicle is likely to encounter it on its planned path. As the planned path is generally unknown to the server, the server predicts the future path of the vehicle and utilizes the predicted paths and their respective probabilities to estimate the probability of the vehicle to encounter the event. This encounter probability also considers the requirement that the event is still active by the time the vehicle arrives at the event location, which is achieved by predicting the lifetime of the event based on an exponential function. In Section 7.3, we evaluate our relevance assessment to increase the efficiency of the message dissemination. We provide a detailed discussion about the influence factors on the performance of this relevance-based dissemination approach and highlight the situations in which our approach provides the largest benefit to the network.

4.4 MODEL FOR THE IMPACT OF MESSAGES ON RECEIVING VEHICLES

While relevance captures the influence of the vehicle's context on the message dissemination, it does not capture the influence of the measurement accuracy. Similarly, the measurement accuracy does not consider the relevance of a message to a vehicle. To determine the importance of a message to a vehicle, a holistic metric is required, which considers both the relevance, the measurement accuracy, and the possible impact of the sensed event. As an example, a traffic jam is generally of higher importance than a traffic sign, which should be reflected in the impact of the associated measurements.

In this section, we describe the modeling of the base-impact and specific-impact of a message carrying some road event for a vehicle. The base-impact is defined in Definition 4.4.1 and the specific-impact is defined in Definition 4.4.2.

Definition 4.4.1: Base-Impact (event type dependent)

The base-impact of a message is a metric that captures the maximum possible gain (or reduction in costs) for a vehicle when receiving a message containing a road event. The base-impact is generally without unit and captures the maximum impact that could be provided by a specific message without considering its contents, but only its type.

Definition 4.4.2: Specific-Impact (associated with a vehicle)

The specific-impact of a message is a metric that captures the expected gain (or reduction in costs) for a vehicle when receiving this message. This specific-impact is generally without unit and considers data-specific properties like accuracy and the vehicle's context. It is to be used for comparing the gain provided by messages in order to prioritize/filter them accordingly under limited bandwidth conditions.

While the base-impact is determined based on the type of event contained in the message, the specific-impact considers the relevance of a message to a vehicle (Section 4.3) as well as the accuracy of the measurement (see Section 4.2). The accuracy influences the usability of the message for the vehicular applications, while the relevance influences the probability that a vehicular application requires that message. To determine the specific-impact of a message, a state-dependent modifier is utilized to consider the influence of an event state on a vehicular application, as shown Section 4.4.1. After that, we discuss the implications of this derivation of the specific-impact in Section 4.4.2.

4.4.1 *Derivation of the Specific-Impact Considering the Measurement Accuracy*

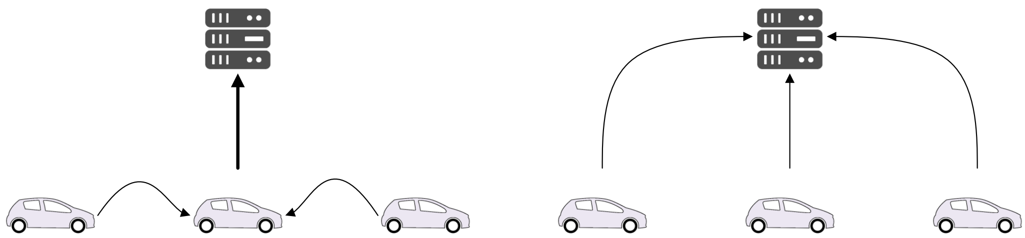
In the following, we assume that the influence of application-specific behavior is limited or known in advance. Thus, for every possible state ψ of the event associated with the data type w, an impact value $\mu(\psi, w)$ is assigned that captures the influence of this event state ψ to the application. The base-impact for a data type w equals the maximum specific-impact value of the states of the data type (i. e., of the states of the contained road event of type w). An example of a data type with multiple states is the event jam. It can have multiple states like no jam, stop-and-go traffic, jam, and complete closure. While these states belong to the same event, it is evident that their impact on the system differs drastically: Stop-and-go traffic might not require a detour, but complete closure almost always requires a detour due to the unpredictability of the closure duration. It is, however, important that these states have the same base-impact for our networking approach described in Chapter 5, as otherwise messages stating the disappearance of an event would not be transmitted to the vehicles that received the initial notification that the event was active.

To consider the accuracy of the measurements contained in the provided data in the definition of specific-impact, we utilize the benefit $\beta(\vec{m}, w)$ of the contained measurement $\vec{m}$, as shown in Equation 27. There, p_i refers to the (conditional) probability associated with state ψ_i, given the measurement $\vec{m}$; for simplicity, we do not show in the notation the dependence of this probability on the measurement. The product operator refers to the scalar product of the two vectors.

$$\beta(\vec{m}, w) = \begin{pmatrix} p_0 \\ \vdots \\ p_{n_w} \end{pmatrix} \cdot \begin{pmatrix} \mu(\psi_0, w) \\ \vdots \\ \mu(\psi_{n_w}, w) \end{pmatrix} \tag{27}$$

The benefit $\beta(\vec{m}, w)$ only describes the influence of the measurement but does not consider the relevance for a certain vehicle. This relevance depends, as mentioned previously, on the probability that the vehicle requires the message. The specific-impact of a message for a certain vehicle v is shown in Equation 28.

$$\mu(\vec{m}, s_e, w, v) = \beta(\vec{m}, w) \cdot R(s_v, s_e) = \begin{pmatrix} p_0 \\ \vdots \\ p_{n_w} \end{pmatrix} \cdot \begin{pmatrix} \mu(\psi_0, w) \\ \vdots \\ \mu(\psi_{n_w}, w) \end{pmatrix} \cdot R(s_v, s_e) \tag{28}$$



(a) Client-side aggregation. (b) Server-side aggregation.

Figure 6: Possibilities of aggregation in our vehicular network (showing the flow of aggregated data with arrows).

The advantage of a definition of the specific-impact using Equation 28 is its simplicity and comprehensibility, which enables the prediction of the network behavior. However, it does not always capture the exact influence of the message to the vehicle, as the vehicle's applications might handle the message using the additional knowledge available at the vehicle like the future path. For example, a jam might only be detoured if a shorter path is available. If no shorter path towards the planned destination is available, the value of the jam notification decreases for the vehicle. We investigated on this aspect in previous works [120], in which we showed that the consideration of application behavior can further increase the performance of the network. However, this puts additional load to the server, thus we focus on the specific-impact definition as shown in Equation 28 in this thesis.

4.4.2 *Implications of the Usage of the Specific-Impact for Prioritization of Messages*

The specific-impact of a message might be high for wrong measurements, which erroneously detect high-impact states. These wrong messages would be distributed to the vehicles and potentially never be revoked if the bandwidth is insufficient to transmit low-impact messages. Although this error can be alleviated by the receiving vehicles through invalidation of the message upon the expiration of the Time to Live (TTL), this effect can potentially decrease the system performance through unnecessary actions like detours of vehicles. To prevent this behavior, the vehicles can pre-aggregate measurements locally and send those aggregates to the server, which again can aggregate these measurements with other measurements.

The possible realizations of this process are displayed in Figure 6. Figure 6a displays the client-side aggregation, in which only aggregates are distributed to other vehicles. While the dissemination area of low-impact messages is generally limited, the vehicles in proximity will generally still receive this aggregate and can consider it in their decision-making. Thus, wrong measurements are not distributed in the whole vehicular network, as they are previously merged with the knowledge already available to the vehicle. The client-side aggregation, however, relies on the vehicles in proximity to share aggregated data, which is not always possible due to low vehicle traffic density or a low deployment rate of the necessary sensors. Figure 6b addresses this issue by shifting the aggregation process to the server. Thus, vehicles only share their raw measurements (not shown in Figure 6b), which are then processed by the server and eventually distributed (not shown in Figure 6b) to the concerned vehicles. Due to the centralized management of data, the errors induced by decentralized data management, due to the imperfect knowledge of the vehicles, are avoided. In this case, all network knowledge is always collected at a central, fail-proof entity. This, however, has the disadvantage of high resource demand in terms of computational power, which cannot always be assumed. Thus, the aggregation is generally performed by the vehicles, and can be performed by the server if the server resources are sufficient. In general, the most influential data types might be managed by the server to prevent false data with high-impact in the network.

In this chapter, we investigated potential influence factors for the importance of a message in a vehicular network and developed methods to assess this importance based on data-specific properties and considering the receiving vehicle's context. These insights are key to the design of our approximate vehicular networks in Chapter 5, as data can be prioritized and filtered based on their importance for the network.

5 APPROXIMATE VEHICULAR NETWORKS

Using the specific-impact definition presented in the previous chapter, we present our innovative approach for approximate vehicular networks, which are described by Definition 5.0.1 based on the definition of approximate computing [25].

Definition 5.0.1: Approximate Vehicular Networks

Approximate networking is a networking paradigm that trades communication and computation resources with the availability or quality of data. Approximate vehicular networks are approximate networks tailored for the vehicular environment.

In traditional networks, the availability/quality of data is commonly increased by increasing the amount of available resources. In contrast, approximate networks focus on resource-constrained environments, in which an increase of available networking resources is either undesired or infeasible. In these networks, some data may not be possible to become available to certain consumers or their quality might be low.

As an alternative to increasing the amount of available resources, we can enhance Quality of Service (QoS) in an approximate networking environment through what we will refer to in general as *Diverse Prioritization and Treatment (DPT)*. The Unequal Error Protection (UEP) scheme [160] described in Section 2.4 is basically a special instance of the aforementioned DPT framework, in which unequal error protection to bits that are of diverse importance to an application is provided, without consuming any additional resources. By applying the UEP scheme, the QoS to a specific application will be improved without increasing the amount of resources consumed. Similarly, the quality or availability of data could be enhanced in an approximate networking environment by applying the concept of DPT, without increasing the resources consumed. The DPT framework can be stochastic and not limited to deterministic prioritization and treatment only. The main focus of this chapter is to show the potential benefits of Approximate Vehicular Networking through the development and study of a DPT mechanism that enhances data availability and quality in vehicular networks without increasing the amount of resources consumed.

When a vehicle subscribes to receive data from a central server, it generally prioritizes high-impact messages to increase its communication efficiency. However, this leads to redundant transmission of a message to all vehicles with similar context, as every vehicle receives all messages ordered by their priority. When these vehicles cooperate, this redundancy is unnecessary and can be reduced to increase the efficiency of the network. Our proposed mechanism adapts this redundancy probabilistically to reduce the utilized communication resources, that can then be used to receive lower-priority messages, that would previously have not been received.

In Section 5.1, we describe our system including the required assumptions. In Section 5.2, we analyze approximate vehicular networks without cooperation, that is assuming that vehicles do not share any messages with vehicles in proximity. This scenario will be used as a baseline to assess the improvements through cooperation. In Section 5.3, we propose our probabilistic approach to approximate vehicular networking using a utility-based game to modulate the vehicle's subscription such that the benefit of the local sharing of messages is increased. In Section 5.4, we propose mechanisms to determine the necessary parameters of our system. In Section 5.5, we describe possible mechanisms to consider the specific-impact for a vehicle in the dissemination of messages. We conclude this chapter with the description with the analysis of our cooperative approximate vehicular networks in Section 5.6.

5.1 SCENARIO DESCRIPTION

In the following, we describe the proposed approach to approximate vehicular networking. As also prescribed in the previous chapter, such an approximate vehicular network will be a context-aware vehicular network, in which a central server provides context-sensitive messages to interested vehicles. The vehicles may cooperate to improve the communication efficiency, by sharing messages received by the server with their neighborhood via Wifi-based communication technology, which is considered to be free of cost.

Our approximate vehicular network is slotted in time. In each timeslot, a vehicle is willing to receive up to A bits on average via the cellular network. This does not mean that this bandwidth may not be temporarily exceeded, but only that the vehicle needs to receive at most A bits per timeslot *on average*. Since the available bandwidth A is assumed to be limited, a vehicle aims to receive high-impact per bit messages and might drop low-impact per bit messages if the bandwidth is insufficient to receive everything. The size $a(\vec{d})$ of every message $\vec{d}$ and its base-impact are known and thus, the average base-impact per received bit $\mu_B(\vec{d})$. We use the normalized metric of the base-impact per bit as the bandwidth is a limiting factor, and thus, the base-impact per utilized bandwidth is pivotal for our approach.

The available messages are divided to n_μ buckets, the so-called impact-levels. As the specific-impact of future messages is not known, these impact-levels are based on the base-impact of the available message types. A message is assigned to an impact-level i if the base-impact per bit of the message is similar to the impact assigned to that impact-level ($\mu_B(\vec{d}) \approx \mu_{B\,i}$). Each impact-level is generally associated with one specific event type, such that the base-impact per bit of that message type equals the base-impact $\mu_{B\,i}$ of the impact-level i. The server then considers the specific-impact of a message by filtering messages based on their specific-impact before considering the transmission to the vehicles. This specific-impact is based on the vehicles, i. e., the context-based filtering of messages is performed at the server. In Section 5.5.2, we describe the server-side filtering in detail.

For each impact-level i, a vehicle expects a load of $a_{0,i}$ bits for receiving all messages, with an average specific-impact per bit $\mu_{0,i}$ if it shares its exact location with the server. At the beginning of each timeslot, the vehicles announce the impact-levels they are interested in, such that its bandwidth restrictions are fulfilled on average. This announcement must not be changed over the course of a timeslot. During the timeslot, the vehicles update their context at the server, such that the server can calculate the specific-impact of a message to provide proper messages to the vehicles. If a vehicle is not willing to share its exact context to protect its privacy, it is also possible to share an inaccurate representation of this context (i. e., location in the discussion in this thesis) as discussed in Section 5.1.1. In the following, we focus our attention on one vehicle in the network, the so-called tagged vehicle.

5.1.1 *Privacy-based Context Sharing*

As mentioned previously, the server calculates the specific-impact of a message for every vehicle to consider this specific-impact in the dissemination. As the specific-impact is context-dependent, a tagged vehicle needs to constantly share its context with a central server. However, this constant sharing of context compromises the privacy of its passengers and might be undesired. Thus, the vehicles can preserve the privacy of their passengers by sharing an inaccurate representation of the context, such that privacy-sensitive vehicles can participate in the network. This concept is known as obfuscation [58] and its privacy protection depends on the degree of the inaccuracy of the context. For this purpose, different privacy-levels $\phi \in \Phi$ with $\Phi \subset \mathbb{N}_0$ are available, which differ in the amount of inaccuracy introduced on purpose to the context. The privacy level $\phi = 0$ does not add any privacy and, thus, corresponds to the privacy-insensitive case. Each vehicle can then select its privacy-level such that its privacy demands are satisfied. We focus on adding imprecision to the location of the vehicle, as it provides the most insights about the person's everyday life. Instead of providing the exact location, the vehicle only provides a (circular) area (imprecision area), in which it is certainly located (uniformly distributed). The radius r_ϕ of the imprecision area varies depending on the privacy-level. While a large imprecision area complicates tracking of vehicles along their route, it also degrades the performance of the dissemination of context-sensitive messages: The server shares a message with a vehicle based on the highest possible specific-impact for *any* possible location of the vehicle, which is necessary to guarantee the reception of a context-sensitive message. Thus, large imprecision areas naturally induce a higher load compared to the small imprecision areas, as more messages need to be considered to assure the reception of all messages. For an impact-level i and the privacy level ϕ, the bandwidth consumption $a_{\phi,i}$ depends on the bandwidth consumption $a_{0,i}$ of the privacy-insensitive case, which is adapted using the privacy adaptation factor $\rho_{\phi,i}$.

The base-impact $\mu_{B\,i}$ of impact-level i is independent of relevance, but a metric for a tagged vehicle's performance needs to take relevance into account, which is captured by the specific-impact. As the specific-impact considers the current context of the vehicle, which is (potentially) unknown by the server, the average (over all possible locations of the tagged vehicle) specific-impact per bit is used, which is assumed to be independent of the current context of the vehicle. Thus, we derive for a tagged vehicle the average specific-impact per bit $\mu_{\phi,i}$ for the impact-level i and the privacy-level ϕ using the average specific-impact per bit $\mu_{0,i}$ for the privacy level 0, which does not add any privacy. The privacy adaptation factor $\rho_{\phi,i}$ captures the increase in bandwidth consumption and the decrease in the average specific-impact per bit for the the impact-level i caused by the privacy-level ϕ compared to the non-privacy case. That is, as messages with a smaller dissemination scope (or, more context-sensitive) are more affected by privacy than than messages with a large dissemination scope (or context-insensitive). In Section 5.4, we describe methods to determine $\rho_{\phi,i}$.

To account for the additional bandwidth consumption, we introduce the variable $a_{\phi,i} \geqslant a_{0,i}$, which captures the load in bits per time interval for subscribing to impact-level i if the vehicle has a privacy-level ϕ and is shown in Equation 29.

$$a_{\phi,i} = a_{0,i} \cdot \rho_{\phi,i} \tag{29}$$

Similarly, the average specific-impact per bit $\mu_{\phi,i} \leqslant \mu_{0,i}$ for impact-level i and privacy-level ϕ is given by Equation 30.

$$\mu_{\phi,i} = \frac{\mu_{0,i}}{\rho_{\phi,i}} \tag{30}$$

A property of the privacy-dependent specific-impact and bandwidth is the fact that the total specific-impact of messages in an impact-level is independent of the privacy-level, as the specific-impact associated with the vehicle is independent of the privacy-level. This is also shown in Equation 31.

$$\mu_{\phi,i} \cdot a_{\phi,i} = \frac{\mu_{0,i}}{\rho_{\phi,i}} \cdot a_{0,i} \cdot \rho_{\phi,i} = \mu_{0,i} \cdot a_{0,i} \tag{31}$$

5.1.2 *Announcement of Impact-Levels of Interest*

In addition to the context-sharing with the server, the tagged vehicle announces interest in a set of impact-levels to receive messages from a server. When considering a setting with limited cellular bandwidth, vehicles generally aim at receiving messages with the highest specific-impact per bit. To achieve that, each vehicle could communicate its available bandwidth to the server, which then could share the messages with the highest specific-impact per bit with the vehicles. This approach would maximize the communication efficiency of each individual vehicle.

However, with this approach, the server delivers the same messages to all vehicles with the same context. While this is no issue for non-cooperative networks, it drastically reduces the possible benefits of cooperation between vehicles. To increase the benefits of cooperation, the similarity between the messages received by vehicles in communication range needs to be reduced. For this purpose, we generalize the aforementioned prioritization strategy using a probabilistic strategy. In this probabilistic strategy, each vehicle has a certain probability of subscribing to a certain impact-level. For the non-cooperative case, all probabilities for impact-levels with a high base-impact per bit (and thus high average specific-impact per bit) will be 1, until the available bandwidth is insufficient. However, for the cooperative case, the probabilistic strategy enables the modulation of the redundancy for high-impact-levels, such that the similarity of messages received by vehicles in proximity is reduced.

The impact-levels a tagged vehicle of privacy level ϕ_t subscribes to are expressed through a set of probabilities $p_{\phi_t,i}$, with one probability per impact-level i; the vector containing these probabilities for all i's is denoted by $\vec{p}_{\phi_t}$, and is referred to as the vehicle's strategy. The strategy of the tagged vehicle $\vec{p}_{\phi_t}$ depends on the current level of privacy ϕ_t. Depending on the probability $p_{\phi_t,i}$ to receive messages of impact-level i and the expected total load $a_{\phi_t,i}$ of an impact-level i in the next timeslot, the vehicle calculates the expected traffic load for each impact-level i. To fulfill the bandwidth restrictions, the vehicle needs to choose its strategy $\vec{p}_{\phi_t}$ such that the sum of the expected traffic loads for all impact-levels is smaller or equal to the available bandwidth A. If the vehicle detects that its average bandwidth usage is lower than the available bandwidth A, the vehicle may share the information about the unused bandwidth with the server. In this case, the server may adapt its dissemination using this knowledge as discussed in Section 5.5.2.

Based on the strategy $\vec{p}_{\phi_t}$, there are two possibilities for the tagged vehicle to announce its impact-levels of interest based on a probabilistic strategy, (i) determine its impact-levels of interest probabilistically in the vehicle itself and provide only the final set of impact-levels, or (ii) transfer the probabilities stating the impact-levels of interest in certain impact-levels to the server. In the first case, the vehicle communicates to the server its derived interest in certain impact-levels and receives all messages assigned to these levels. It may coordinate this announcement of impact-levels of interest locally with other vehicles in proximity, but we do not consider this approach in this thesis due to the higher vulnerability to malicious nodes. In the second case, the vehicles may announce their impact-levels of interest probabilistically, i. e., they announce that they want to receive a certain fraction of messages of a certain impact-level. This enables the server to provide different messages to neighboring vehicles, but increases the size of a subscription and does not provide any benefit for privacy-sensitive vehicles. Thus, we focus on the first option in this thesis.

5.1.3 *Distribution of Messages*

The central server uses the context representation and the impact-levels of interest provided by the vehicles to actively push incoming road events to the interested vehicles. A vehicle is considered to be interested in receiving a message iff (i) the vehicle has subscribed to the impact-level associated with the message, and (ii) the server considers the message transmission to the vehicle by taking the specific-impact of the message for the potential receiver into account. The description of the dissemination of messages based on their specific-impact is described in detail in Section 5.5.1. Notice that the specific-impact considers the context of the vehicle, i. e., prevents messages from being flooded in the network.

5.1.4 *Utility Metric for Approximate Vehicular Networks*

The utility of the tagged vehicle is an appropriate metric for the influence of the developed approaches to our approximate vehicular network. The utility is defined as the sum of specific-impact values of the set of messages D_{rcv} received by that vehicle. Consequently, the utility is defined as shown in Equation 32. $a(\vec{d})$ describes the bandwidth consumption for the message $\vec{d}$, and $\mu(\vec{d})$ the specific-impact per used bandwidth.

$$u = \sum_{\vec{d} \in D_{snt}} a(\vec{d}) \cdot \mu(\vec{d}) \tag{32}$$

In order to use the utility to optimize our approximate network, the expected utility for a vehicle is required. As mentioned previously, we assume that the subscription strategy $\vec{p}_{\phi_t}$ of the tagged vehicle is probabilistic, as well as the (unknown) strategies of the vehicles in proximity. In this vector, each probability $p_{\phi_t,i}$ refers to the probability that the vehicle subscribes to a certain impact-level i. Thus, this probability also captures the share of messages of impact-level i received by the vehicle via the cellular network. However, the utility is based on the probability of the tagged vehicle to receive a message of a particular impact-level via *any* communication channel. Each vehicle has two options to receive a message, (i) via the cellular network (the only possibility under the non-cooperative case), and (ii) via Wifi from at least one vehicle in its neighborhood (possible only under the cooperative case). For the reception via the cellular network, the tagged vehicle with privacy level ϕ_t can influence the reception probability p_i^{cel} to receive a message of impact-level i by adapting its strategy $\vec{p}_{\phi_t}$, with which it receives messages via the cellular communication channel. For this case, also the bandwidth restrictions apply. Notice that p_i^{cel} might potentially differ from $p_{\phi,i}$, as messages might get lost. For the reception via Wifi, the tagged vehicle cannot influence the probability p_i^{wif} to receive the message of impact-level i from its neighbors, as this purely relies on their willingness to share data and their respective probabilities to receive a message via the cellular communication channel. A vehicle receives a message if it is received via either communication channel, which is captured by the expected utility of the tagged vehicle as shown in Equation 33.

$$\overline{u}(p_0^{cel}, \dots, p_{n_\mu}^{cel}, p_0^{wif}, \dots, p_{n_\mu}^{wif}) = \sum_{i=1}^{n_\mu} a_{\phi,i} \cdot \mu_{\phi,i} \cdot \left[1 - (1 - p_i^{cel}) \cdot (1 - p_i^{wif})\right] \tag{33}$$

Notice that the restriction on the average cellular bandwidth needs to be fulfilled by the tagged vehicle as shown in Equation 34.

$$\sum_{i=1}^{n_\mu} a_{\phi,i} \cdot p_i^{cel} \leqslant A \tag{34}$$

Furthermore, the constraints for each probability p_i^{cel} and p_i^{wif} need to be fulfilled as shown in Equation 35.

$$0 \leqslant p_i^{cel} \leqslant 1, \forall i \in \{1, \dots, n_\mu\} \text{ and } 0 \leqslant p_i^{wif} \leqslant 1, \forall i \in \{1, \dots, n_\mu\} \tag{35}$$

In the following, we develop strategies for the determination of the probabilities $p_i^{cel}, \forall i \in \{1, \dots, n_\mu\}$ of the tagged vehicle for the reception of cellular messages, such that the overall utility is maximized. We first consider the case of a vehicular network without cooperation between vehicles in Section 5.2, followed by the case of a cooperative network in Section 5.3.

5.2 NON-COOPERATIVE APPROXIMATE VEHICULAR NETWORKS

In vehicular networks without cooperation, vehicles cannot rely on others to provide messages to them, i. e., $p^{wif} = 0$. In this thesis, we assume that there is no message drop during the transmission of messages from the server to the vehicle and, consequently, $p_i^{cel} = p_{\phi_t,i}$ for the privacy level ϕ_t of the tagged vehicle. That is, the probabilistic strategy (to be optimized below) coincides with the set of probabilities of receiving messages through the cellular network. This reliable transmission can be achieved with reliable transport layer protocols like Transmission Control Protocol (TCP) and Multipath TCP (MPTCP), or protocols achieving reliability on the application layer like Quick UDP Internet Connections (QUIC). With that, the expected utility can be expressed using Equation 36. Notice that Equation 36 can be obtained from Equation 33 if $p_i^{cel} = p_{\phi_t,i}$ and $p^{wif} = 0$.

$$\overline{u}(\vec{p}_{\phi_t}, \phi_t) = \sum_{i=1}^{n_\mu} a_{\phi_t,i} \cdot \mu_{\phi_t,i} \cdot p_{\phi_t,i} \tag{36}$$

To maximize the expected utility, the partial derivatives of the expected utility with respect to each probability $p_{\phi_t,i}$ need to set equal to 0. In order to do that, the bandwidth is included as a side condition into the term, as otherwise the utility would grow until $p_{\phi_t,i} = 1, \forall i \in \{1, \dots, n_\mu\}$. Thus, the bandwidth condition from Equation 34 is utilized to express the probability $p_{\phi_t,m}$ as a combination of the others considering $p_i^{cel} = p_{\phi_t,i}$ as shown in Equation 37.

$$p_{\phi_t,m} \leqslant \frac{A - \sum_{i=1 \mid i \neq m}^{n_\mu} a_{\phi_t,i} \cdot p_{\phi_t,i}}{a_{\phi_t,m}} \tag{37}$$

As less bandwidth consumption cannot increase the utility due to the assumption of positive-only values of the specific-impact, an optimal strategy always utilizes the available bandwidth fully (if possible). Thus, the inequality presented in Equation 37 is assumed to be an equality. If this equality is inserted into the expected utility function, Equation 38 is obtained.

$$\overline{u}(\vec{p}_{\phi_t}, \phi_t) = \sum_{i=1 \mid i \neq m}^{n_\mu} a_{\phi_t,i} \cdot (\mu_{\phi_t,i} - \mu_{\phi_t,m}) \cdot p_{\phi_t,i} + \mu_{\phi_t,m} \cdot A \tag{38}$$

Based on Equation 38, the expected utility depends on each probability $p_{\phi_t,i} \mid i \neq m$. As there is only a linear dependency of the expected utility from the probabilities, the partial derivatives of the expected utility with respect to any probability $p_{\phi_t,i} \mid i \neq m$ are always non-zero. Thus, the average utility is maximized at the boundaries of the eligible values of the variables we are trying to determine, which are the probabilities. Being probabilities, their values are limited between 0 and 1. Notice that the values of probabilities maximizing the average utility may not be the maximum possible, so that any other constraint that is in place be met; this is the bandwidth constraint in Equation 34. As a result, it turns out that the optimal solution for the networks without cooperation is achieved by setting all probabilities to 1, until the bandwidth is insufficient, starting from the impact-level with the highest expected specific-impact per bit $\mu_{\phi_t,i}$. The probability that would exceed the bandwidth when being set to 1 is then chosen such that the bandwidth is utilized completely.

5.3 COOPERATIVE APPROXIMATE VEHICULAR NETWORKS

In cooperative approximate vehicular networks, vehicles may also rely on their neighbors to receive messages of interest to them, besides getting them directly by the server via the cellular network. In the previously discussed non-cooperative case, vehicles will follow the same optimal (probabilistic) strategy derived in the previous section and will receive all messages possible (subject to cellular bandwidth capacity limits), starting from the ones of the highest impact to the vehicle; we will also refer to such a strategy as a selfish or greedy-local strategy. As a result and since a given message will very likely be of the same specific-impact level for all vehicles around the same location, it is expected that vehicles in the same neighborhood will pretty much receive the same messages and will miss out (due to bandwidth restrictions) again pretty much the same messages. Consequently, if the vehicles cooperated and exchanged messages in order to receive some messages from other vehicles (or to save cellular bandwidth), such cooperation would bring almost no benefit, due to the similarity of messages of the vehicles in the same neighborhood. In other words, cooperation is beneficiary only if the vehicles in the same neighborhood do not receive exclusively all the highest impact level messages, but decide to leave out some high impact messages and get some lowest impact instead. This way, the set of messages available to the entire neighborhood will be larger and the messages in the vehicles would be more diverse, compared to that under the non-cooperative strategy, Section 5.2.

Based on the above it is clear that in a cooperative environment vehicles would need to adopt a strategy different than the selfish or greedy-local one. That is, vehicles should not receive the highest impact messages which would maximize their utility if no cooperation is available, but opt for receiving also some messages of lower impact and reducing the resulting total utility, aiming at more than compensating for this utility reduction by fetching missing (high impact) messages from the neighbors. In fact, the average over all vehicles utility in a cooperative network is expected to be higher than that under no cooperation (greedy-local strategy).

One possibility to coordinate the reception of messages is through explicit cooperation using cluster-based approaches, in which one selected vehicle, the so-called cluster-head, coordinates the reception of messages. While this approach is very efficient regarding the additional gain of cooperation, it decreases in performance if the topology changes frequently, as happening in vehicular networks. Thus, we develop a non-cooperative game, in which the vehicles in the neighborhood adapt their strategy without explicitly coordinating with the other vehicles. It is clear that the vehicles in a neighborhood choose their strategy $\vec{p_\phi}$, aiming at maximizing their utility in the cooperative networking environment. In the following, we investigate the utility of the tagged vehicle, i. e., pick one vehicle and derive the optimal strategy for this vehicle.

5.3.1 *Game-Theoretic Model*

In contrast to the approach without cooperation between vehicles, the utility of the tagged vehicle in the cooperative use-case depends not only on its own strategy but also on the strategies of the vehicles in proximity. These vehicles can forward messages to the tagged vehicle and, thus, increase its utility. However, the strategies of the neighboring vehicles are generally unknown to the tagged vehicle and, again, depend on the strategies of their respective neighbors. Thus, the global optimal solution would require knowledge of the full vehicular network, which cannot be obtained due to complexity reasons.

To address this issue, we develop a local algorithm by assuming a common strategy for all vehicles of the same privacy-level in the neighborhood. This assumption is not far from realistic as the neighborhoods are (partially) overlapping. It is evident that this assumption is a simplification, but converts the problem to be solved to a problem only dependent on the current neighborhood of the tagged vehicle.

To determine the optimal strategy for the tagged vehicle, we again employ a probabilistic behavior, i. e., we use the probability $p_{\phi,i}$ to receive a certain fraction of messages of impact-level i given its current privacy-level ϕ. As discussed previously, we assume that there is no message drop during the transmission of messages from the server to the vehicle. Thus, the faction of received messages of impact-level i is precisely the probability for the tagged vehicle to subscribe to this impact-level i, i. e., $p_i^{cel} = p_{\phi_t,i}$. In contrast to the approach without vehicular cooperation, the tagged vehicles might receive a message from one of the neighbors. As already mentioned, every vehicle is willingly sharing every message that it receives via the cellular channel with its one-hop neighborhood. Given the number of vehicles n_ϕ with privacy-level ϕ in Wifi communication range, we can determine the probability to receive a message via Wifi as shown in Equation 39. It is equal to the probability that any of the neighbors has received the message via the cellular channel, i. e., *not* none of the neighbors has received the message. Notice that the probability to receive a message from *one* specific vehicle via Wifi is equal to the probability that the tagged vehicle receives it itself via the cellular network due to the assumption of equal strategies. Additionally, notice that n_ϕ includes the tagged vehicle, i. e., the number of neighbors with the same privacy-level is $n_{\phi_t} - 1$.

$$p_i^{wif} = 1 - (1 - p_{\phi_t,i})^{n_{\phi_t}-1} \cdot \prod_{\phi \in \Phi \setminus \{\phi_t\}} (1 - p_{\phi,i})^{n_\phi} \tag{39}$$

Based on p_i^{cel} and p_i^{wif}, we can use Equation 33 to describe the utility of the tagged vehicle based on its neighborhood and its own strategy as shown in Equation 40.

$$\overline{u}(\vec{p}_1, \dots, \vec{p}_{|\Phi|}, \phi_t) = \sum_{i=1}^{n_\mu} a_{\phi_t,i} \cdot \mu_{\phi_t,i} \cdot \left[1 - \prod_{\phi \in \Phi} (1 - p_{\phi,i})^{n_\phi}\right] \tag{40}$$

with the constraints

$$\sum_{i=1}^{n_\mu} a_{\phi,i} \cdot p_{\phi,i} \leqslant A, \forall \phi \in \Phi \tag{41}$$

and

$$p_{\phi,i} \in [0,1], \forall \phi \in \Phi, \forall i \in \{1, \dots, n_\mu\} \tag{42}$$

5.3.2 *Deriving the Optimal Strategy*

The optimal strategy can be derived by finding the set of probabilities $p_{\phi_t,i}$, for which the utility is maximized. To maximize the utility, we use a method which is aligned with the KarushKuhnTucker conditions. The KarushKuhnTucker conditions can be used to derive the input values (in this case $p_{\phi_t,i}$) for which a function (in this case $\overline{u}(\vec{p}_1, \dots, \vec{p}_{|\Phi|}, \phi_t)$) is maximized. Our method devides the set of privacy-levels, for a given impact-level i, into the set of privacy-levels $\Phi^+(i)$, containing the privacy levels under which the tagged vehicle may subscribe to impact-level i (subscribing or not will depend on the derived strategy), and the set of privacy-levels $\Phi^-(i)$, containing the privacy levels under which the tagged vehicle will never subscribe to impact-level i. With that, the cases in which invalid values of $p_{\phi_t,i}$ maximize the expected utility are handled similarly as with the KarushKuhnTucker conditions. The upper bound of $p_{\phi_t,i}$ do not need to be considered as shown later in the derivations.

Similar to the non-cooperative case, we assume that the bandwidth available through Equation 41 is fully utilized; that is, the left hand side is equal to the right hand side. Based on that, any probability $p_{\phi_t,m}$ can be replaced by a combination of all other probabilities as shown in Equation 43.

$$p_{\phi_t,m} = \frac{A - \sum_{i=1 \mid i \neq m}^{n_\mu} a_{\phi_t,i} \cdot p_{\phi_t,i}}{a_{\phi_t,m}} \tag{43}$$

Based on this equation, an increase (decrease) in any of these probabilities would require the decrease (increase) of some other one(s). Notice that all probabilities $p_{\phi_t,i}$ are independently selected (subject to the constraints in Equation 42) except one of them, say $p_{\phi_t,m}$, due to Equation 43. Then, we will try to derive the values of the probabilities $p_{\phi_t,i}$ that maximize the expected utility by considering the partial derivatives of the expected utility with respect to $p_{\phi_t,l}$, for all l except m. As it is only possible to decrease $p_{\phi_t,m}$ if it is above 0, it is necessary that $\phi_t \in \Phi^+(m)$. Otherwise, the derivative of the expected utility with respect to any probability $p_{\phi_t,l}$ for $l \neq m$ would not necessarily be 0, as the optimal value for this probability might be outside of the allowed value range, which would lead to a wrong value of $p_{\phi,l}$. Additionally, the probability $p_{\phi,m}$ needs to be greater than 0 for all privacy-levels, to provide the necessary coordination between them. Otherwise, the impact-level m could not be used as a reference, would increase the complexity of Theorem 1, which is necessary to obtain the optimal strategy. This requirement cannot always be assured and limits the applicability of our solution to a subset of the possible solutions. In Section 5.3.3, we discuss possibilities to alleviate this requirement and obtain all possible solutions.

Until required, $p_{\phi,m}$ is kept in the equation to increase readability and only consider its dependency from all other probabilities $p_{\phi,i}, \forall i \neq m$ when calculating derivatives. The requirements of the probabilities from Equation 42 are not considered in the calculation but are considered later in Section 5.3.3 by changing the impact-levels to which a privacy-level may subscribe to. This ensures the detection of the utility-optimal solution and enables us to choose slightly less optimal solutions to reduce the reliance on the vehicles in proximity. The utility-optimal strategy is derived for considering every impact-level i individually. For this purpose, we calculate the partial derivatives of the utility with respect to the probability to subscribe to any impact-level l any privacy-level $\phi \in \Phi^+(l)$ as shown in Equation 44. As $p_{\phi_t,m}$ is used in Equation 43 to account for the bandwidth dependency, all other two probabilities $p_{\phi,i}$ and $p_{\phi,j}$ are independent if $i \neq j$ and $i \neq m$ and $j \neq m$. Thus, we only need to consider the summands containing the probability $p_{\phi_t,l}$ and $p_{\phi_t,m}$ in the derivative with respect to $p_{\phi,l}$, as the derivative of all other summands are 0. As we are searching for optimal points, the derivative of the expected utility is set to 0 to find the turning points of the utility function. We provide an analysis of these turning points showing that they maximize the utility function (and, thus, yield the optimal points) in Section 5.6.1.

$$\begin{aligned}\frac{\partial \overline{u}}{\partial p_{\phi_t,l}} &= \mu_{\phi_t,l} \cdot a_{\phi_t,l} \cdot n_{\phi_t} \cdot (1-p_{\phi_t,l})^{n_{\phi_t}-1} \cdot \prod_{\phi \in \Phi \setminus \{\phi_t\}} (1-p_{\phi,l})^{n_\phi} \\ &+ \mu_{\phi_t,m} \cdot a_{\phi_t,m} \cdot \left(\frac{\partial p_{\phi_t,m}}{\partial p_{\phi_t,l}}\right) n_{\phi_t} \cdot (1-p_{\phi_t,m})^{n_{\phi_t}-1} \cdot \prod_{\phi \in \Phi \setminus \{\phi_t\}} (1-p_{\phi,m})^{n_\phi} \stackrel{!}{=} 0\end{aligned} \tag{44}$$

Notice that the derivative of $p_{\phi_t,m}$ with respect to $p_{\phi_t,l}$ appears in Equation 44 since $p_{\phi_t,m}$ depends on $p_{\phi_t,l}$ as explained earlier; this derivative can be calculated based on Equation 43 as follows.

$$\frac{\partial p_{\phi_t,m}}{\partial p_{\phi_t,l}} = -\frac{a_{\phi_t,l}}{a_{\phi_t,m}} \tag{45}$$

After some transformations and the insertion of the derivative of $p_{\phi_t,m}$ with respect to $p_{\phi_t,l}$, we obtain Equation 46.

$$\begin{aligned}\mu_{\phi_t,l} \cdot a_{\phi_t,l} \cdot n_{\phi_t} \cdot (1-p_{\phi_t,l})^{n_{\phi_t}-1} \cdot P_l(\Phi \setminus \{\phi_t\}) \\ = \mu_{\phi_t,m} \cdot a_{\phi_t,m} \cdot \frac{a_{\phi_t,l}}{a_{\phi_t,m}} \cdot n_{\phi_t} \cdot (1-p_{\phi_t,m})^{n_{\phi_t}-1} \cdot P_m(\Phi \setminus \{\phi_t\})\end{aligned} \tag{46}$$

with

$$P_j(\Phi) = \prod_{\phi \in \Phi} (1-p_{\phi,j})^{n_\phi}$$

This equation can be simplified considering that (i) $\mu_{\phi_t,i} = {}^{\mu_{0,i}}/_{\rho_{\phi_t,i}}$ as discussed in Section 5.1.1, and (ii) $a_{\phi_t,l}$ and n_{ϕ_t} are contained on both sides of the equation. This leads to Equation 47, which needs to be fulfilled for the optimal solution.

$$\left(\frac{1-p_{\phi_t,l}}{1-p_{\phi_t,m}}\right)^{n_{\phi_t}-1} = \left(\frac{\mu_{0,m}}{\mu_{0,l}}\right) \cdot \left(\frac{\rho_{\phi_t,l}}{\rho_{\phi_t,m}}\right) \cdot \frac{P_m(\Phi \setminus \{\phi_t\})}{P_l(\Phi \setminus \{\phi_t\})} \tag{47}$$

In this equation, not only the probability $p_{\phi_t,l}$ of the impact-level l of the tagged vehicle is contained, but also the probabilities $p_{\phi,l}, \forall \phi \in \Phi \setminus \{\phi_t\}$ of all other privacy-levels. To be able to calculate $p_{\phi_t,l}$, for all l except m, we need to represent $p_{\phi,l}, \forall \phi \in \Phi \setminus \{\phi_t\}$ as $p_{\phi_t,l}$. According to Theorem 1, Equation 47 can be represented as Equation 48. In this equation, we use the auxiliary variables $\Phi^+(l)$, which is a set containing all privacy-levels with $p_{\phi,l} > 0$, and $\Phi^-(l)$, which is a set containing all privacy-levels with $p_{\phi,l} = 0$. It is evident, that any ϕ can only be contained in either $\Phi^+(l)$ or $\Phi^-(l)$, i. e., $\Phi^+(l) \oplus \Phi^-(l) = \Phi$ and $\Phi^+(l) \cap \Phi^-(l) = \emptyset$.

$$\left(\frac{1-p_{\phi_t,l}}{1-p_{\phi_t,m}}\right)^{n^+(l)} = \left(\frac{\mu_{0,m}}{\mu_{0,l}}\right) \cdot \left(\frac{\rho_{\phi_t,l}}{\rho_{\phi_t,m}}\right) \cdot \prod_{\phi \in \Phi^+(l) \setminus \{\phi_t\}} \left(\frac{\rho_{\phi,l} \cdot \rho_{\phi_t,m}}{\rho_{\phi,m} \cdot \rho_{\phi_t,l}}\right)^{n_\phi} \cdot P_m(\Phi^-(l)) \tag{48}$$

with

$$n^+(l) = \sum_{\phi \in \Phi^+(l)} n_\phi - 1$$

Notice that Equation 48 depends on the probability $p_{\phi,m}$ of all privacy-levels $\phi \in \Phi^-(l)$. Additionally, $n^+(l)$ is the total number of neighbors that may subscribe to the impact-level l.

Theorem 1. *We can represent Equation 47 as Equation 48.*

Proof. We use full-induction to show that Equation 47 can be represented as Equation 48. For this purpose, we utilize the partial derivatives of the utility with respect to the probabilities of all privacy-level $\phi \in \Phi^+(l)$ that are in the neighborhood of the tagged vehicle.

For the base-case, we set $\Phi = \{\phi_t\}$. Based on Equation 47, we observe that $P_m(\Phi \setminus \{\phi_t\}) = 1$ and $P_l(\Phi \setminus \{\phi_t\}) = 1$, as Φ contains only ϕ_t. Additionally $n^+(l) = n_{\phi_t} - 1$ and $\phi_t \in \Phi^+(l)$ as the derivative with respect to $p_{\phi_t,l}$ has only been considered if $\phi_t \in \Phi^+(l)$. Thus, we can immediately represent Equation 47 as Equation 48 for the base-case.

For the induction step, we introduce additional auxiliary variables. These are $\Phi^+_+(l) \subseteq \Phi^+(l)$ and $\Phi^+_-(l) \subseteq \Phi^+(l)$ with $\phi \in \Phi^+_+(l) \oplus \Phi^+_-(l), \forall \phi \in \Phi^+(l) \setminus \{\phi_t\}$. $\Phi^+_+(l)$ contains all privacy-levels ϕ, for which the derivative of the utility with respect to $p_{\phi,l}$ has already been included in the calculation of ϕ_t. Similarly, $\Phi^+_-(l)$ contains all privacy-levels that have not been included. Notice that the privacy-level after which the utility has been derived is *not* included in $\Phi^+_+(l)$, but always considered separately. In each induction step, the probability $p_{\phi_n,l}$ of one additional privacy-level ϕ_n is inserted into all equations that have not yet been considered, i. e., that are part of $\Phi^+_-(l)$.

Based on $\Phi^+_-(l)$ and $\Phi^+_+(l)$, we can introduce Equation 49 as a combination of Equation 47 and Equation 48, which captures the intermediate result, in which only a subset of the probabilities has already been inserted. $n^+_+(l)$ is defined similar to $n^+(l)$, but only considers privacy-levels in $\Phi^+_+(l)$. For all privacy-levels in $\Phi^+_+(l)$, the representation from Equation 48 is used, while Equation 47 is used for all privacy-levels in $\Phi^+_-(l)$, which still needs to be handled. Notice that $n^+_+(l)$ contains the -1 that would otherwise be missing, as the original exponent is $n_{\phi_t} - 1$.

$$\begin{aligned}
&\left(\frac{1 - p_{\phi_t,l}}{1 - p_{\phi_t,m}}\right)^{n^+_+(l) + n_{\phi_t}} = \\
&\left(\frac{\mu_{0,m}}{\mu_{0,l}}\right) \cdot \left(\frac{\rho_{\phi_t,l}}{\rho_{\phi_t,m}}\right) \cdot \prod_{\phi \in \Phi^+_+(l)} \left(\frac{\rho_{\phi,l} \cdot \rho_{\phi_t,m}}{\rho_{\phi,m} \cdot \rho_{\phi_t,l}}\right)^{n_\phi} \cdot \frac{P_m((\Phi^-(l) \cup \Phi^+_-(l))\} \setminus \{\phi_t\})}{P_l(\Phi^+_-(l) \setminus \{\phi_t\})}
\end{aligned} \tag{49}$$

In the following, we aim at including the privacy-level ϕ_n, which is currently in $\Phi^+_-(l)$, into $\Phi^+_+(l)$. For this purpose, we utilize a similar equation as Equation 49, just with the difference that the utility has been derived with respect to $p_{\phi_n,l}$ instead of $p_{\phi_t,l}$. The remainder of the equation remains similar except for the difference in adaptation factors ($\rho_{\phi_n,m}$ vs. $\rho_{\phi_t,m}$ and $\rho_{\phi_n,l}$ vs. $\rho_{\phi_t,l}$), which leads to Equation 50.

$$\left(\frac{1-p_{\phi_n,l}}{1-p_{\phi_n,m}}\right)^{n_+^\pm(l)+n_{\phi_n}} = \\ \left(\frac{\mu_{0,m}}{\mu_{0,l}}\right)\cdot\left(\frac{\rho_{\phi_n,l}}{\rho_{\phi_n,m}}\right)\cdot\prod_{\phi\in\Phi_+^\pm(l)}\left(\frac{\rho_{\phi,l}\cdot\rho_{\phi_n,m}}{\rho_{\phi,m}\cdot\rho_{\phi_n,l}}\right)^{n_\phi}\cdot\frac{P_m((\Phi^-(l)\cup\Phi_-^\pm(l))\setminus\{\phi_n\})}{P_l(\Phi_-^\pm(l)\setminus\{\phi_n\})} \tag{50}$$

We now insert the probability $p_{\phi_n,l}$ associated with the privacy-level ϕ_n into Equation 49. This probability is contained in $P_l(\dots)$ of Equation 49 respectively, which is adjusted such that $p_{\phi_n,l}$ can be considered separately. Additionally, the probabilities $p_{\phi_t,l}$ and $p_{\phi_t,m}$ are contained in $P_l(\dots)$ and $P_m(\dots)$ of Equation 50 respectively, which are also adjusted such that these probabilities can be considered separately. After some transforms, we obtain Equation 51. While $p_{\phi_n,l}$ is the replaced variable and, thus, not contained in this equation, $p_{\phi_n,m}$ could be canceled immediately by adjusting $P_m(\dots)$.

$$\left(\frac{1-p_{\phi_t,l}}{1-p_{\phi_t,m}}\right)^{n_+^\pm(l)+n_{\phi_t}-\dfrac{n_{\phi_n}\cdot n_{\phi_t}}{n_+^\pm(l)+n_{\phi_n}}} = \left(\frac{\mu_{0,m}}{\mu_{0,l}}\right)^{1-\dfrac{n_{\phi_n}}{n_+^\pm(l)+n_{\phi_n}}} \\ \cdot\left(\frac{\rho_{\phi_t,l}}{\rho_{\phi_t,m}}\right)^{1-\sum_{\phi\in\Phi_+^\pm(l)}n_\phi}\cdot\prod_{\phi\in\Phi_+^\pm(l)}\left(\frac{\rho_{\phi,l}}{\rho_{\phi,m}}\right)^{n_\phi\cdot\left(1-\dfrac{n_{\phi_n}}{n_+^\pm(l)+n_{\phi_n}}\right)} \\ \cdot\left(\frac{P_m(\{\Phi^-(l)\cup\Phi_-^\pm(l)\}\setminus\{\phi_t,\phi_n\})}{P_l(\Phi_-^\pm(l)\setminus\{\phi_t,\phi_n\})}\right)^{1-\dfrac{n_{\phi_n}}{n_+^\pm(l)+n_{\phi_n}}} \\ \cdot\left(\frac{\rho_{\phi_n,l}}{\rho_{\phi_n,m}}\right)^{\left(\sum_{\phi\in\Phi_+^\pm(l)}n_\phi-1\right)\cdot\dfrac{n_{\phi_n}}{n_+^\pm(l)+n_{\phi_n}}} \tag{51}$$

These exponents can be transformed now according to Equation 52, Equation 53, and Equation 54. Notice that $\sum_{\phi\in\Phi_+^\pm(l)}n_\phi - 1 = n_+^\pm(l)$.

$$n_+^\pm(l)+n_{\phi_t}-\frac{n_{\phi_n}\cdot n_{\phi_t}}{n_+^\pm(l)+n_{\phi_n}} = \frac{n_+^\pm(l)\cdot\left(n_+^\pm(l)+n_{\phi_n}+n_{\phi_t}\right)}{n_+^\pm(l)+n_{\phi_n}} \tag{52}$$

$$1-\frac{n_{\phi_n}}{n_+^\pm(l)+n_{\phi_n}} = \frac{n_+^\pm(l)}{n_+^\pm(l)+n_{\phi_n}} \tag{53}$$

$$\left(\sum_{\phi\in\Phi_+^\pm(l)}n_\phi-1\right)\cdot\frac{n_{\phi_n}}{n_+^\pm(l)+n_{\phi_n}} = \frac{n_+^\pm(l)\cdot n_{\phi_n}}{n_+^\pm(l)+n_{\phi_n}} \tag{54}$$

We can observe that all these exponents share $n_+^\pm(l)/(n_+^\pm(l)+n_{\phi_n})$ common factor, which we eliminate in the following to transform Equation 51 to Equation 55.

$$\left(\frac{1-p_{\phi_t,l}}{1-p_{\phi_t,m}}\right)^{n_+^\pm(l)+n_{\phi_n}+n_{\phi_t}} = \left(\frac{\mu_{0,m}}{\mu_{0,l}}\right)\cdot\left(\frac{\rho_{\phi_t,l}}{\rho_{\phi_t,m}}\right)^{-(n_+^\pm(l)+n_{\phi_n})}\cdot$$
$$\prod_{\phi\in\Phi_+^\pm(l)}\left(\frac{\rho_{\phi,l}}{\rho_{\phi,m}}\right)^{n_\phi}\cdot\frac{P_m(\{\Phi^-(l)\cup\Phi_-^\pm(l)\}\setminus\{\phi_t,\phi_n\})}{P_l(\Phi_-^\pm(l)\setminus\{\phi_t,\phi_n\})}\cdot\left(\frac{\rho_{\phi_n,l}}{\rho_{\phi_n,m}}\right)^{n_{\phi_n}} \tag{55}$$

With some minor transformations, like the grouping of the adaptation factors $\rho_{\ldots}$, we can then obtain Equation 56.

$$\left(\frac{1-p_{\phi_t,l}}{1-p_{\phi_t,m}}\right)^{n_+^\pm(l)+n_{\phi_n}+n_{\phi_t}} = \left(\frac{\mu_{0,m}}{\mu_{0,l}}\right)\cdot\left(\frac{\rho_{\phi_t,l}}{\rho_{\phi_t,m}}\right)\cdot$$
$$\prod_{\phi\in\Phi_+^\pm(l)}\left(\frac{\rho_{\phi,l}\cdot\rho_{\phi_t,m}}{\rho_{\phi,m}\cdot\rho_{\phi_t,l}}\right)^{n_\phi}\cdot\left(\frac{\rho_{\phi_n,l}\cdot\rho_{\phi_t,m}}{\rho_{\phi_n,m}\cdot\rho_{\phi_t,l}}\right)^{n_{\phi_n}}\cdot\frac{P_m((\Phi^-(l)\cup\Phi_-^\pm(l))\setminus\{\phi_t,\phi_n\})}{P_l(\Phi_-^\pm(l)\setminus\{\phi_t,\phi_n\})} \tag{56}$$

If we add ϕ_n into $\Phi_+^\pm(l)$ and remove it from $\Phi_-^\pm(l)$ in Equation 49, we end up exactly with Equation 56. Additionally, if we set $\Phi_+^\pm(l) = \{\Phi^+(l)\setminus\{\phi_t\}\}$ and $\Phi_-^\pm(l) = \emptyset$ by considering all privacy-levels, we obtain Equation 48. Thus, we successfully showed that the derivative of the utility function can be simplified as described above. □

According to Equation 48, we can calculate $p_{\phi_t,l}$ only based on the other probabilities of ϕ_t through the auxiliary variable $p_{\phi_t,m}$ and the other privacy-levels ϕ for which $p_{\phi,l} = 0$. To determine the value of $p_{\phi_t,l}$, we need to replace $p_{\phi_t,m}$ with all other probabilities $p_{\phi_t,i} \mid i \neq m$ as shown in Equation 57.

$$\left(\frac{1-p_{\phi_t,l}}{1-\left(\frac{A}{a_{\phi_t,m}}-\sum_{i=1\,\mid\,i\neq m}^{n_\mu}\frac{a_{\phi_t,i}\cdot p_{\phi_t,i}}{a_{\phi_t,m}}\right)}\right)^{n^+(l)}$$
$$= \left(\frac{\mu_{0,m}}{\mu_{0,l}}\right)\cdot\left(\frac{\rho_{\phi_t,l}}{\rho_{\phi_t,m}}\right)\cdot\prod_{\phi\in\Phi^+(l)\setminus\{\phi_t\}}\left(\frac{\rho_{\phi,l}\cdot\rho_{\phi_t,m}}{\rho_{\phi,m}\cdot\rho_{\phi_t,l}}\right)^{n_\phi}\cdot P_m(\Phi^-(l)) \tag{57}$$

After taking the $n^+(l)$-th root of the equation, and some manipulations, we obtain Equation 58. This equation still contains the dependency on other privacy-levels encapsulated in $P_m(\Phi^-(l))$, which increases the complexity of the solution. For now, we assume that $P_m(\Phi^-(l))$ is constant, i. e., the strategy associated with the other privacy-levels is already known. The implications or this assumption are discussed and addressed later for the general case. Notice also that no such assumption is needed in certain cases, such as when $\Phi^+(i) = \Phi$ or $\Phi^+(i) = \emptyset$ for all $i \in \{1,\ldots,n_\mu\}$, when there is only one privacy level, etc. In this case, Equation 58 is a linear equation and, thus, can be simplified.

$$1 - p_{\phi_t,l} = \left[1 - \left(\frac{A}{a_{\phi_t,m}} - \sum_{i=1 \mid i \neq m}^{n_\mu} \frac{a_{\phi_t,i} \cdot p_{\phi_t,i}}{a_{\phi_t,m}}\right)\right] \cdot \Lambda_l \tag{58}$$

with

$$\Lambda_l = \sqrt[n^+(l)]{\left(\frac{\mu_{0,m} \cdot \rho_{\phi_t,l}}{\mu_{0,l} \cdot \rho_{\phi_t,m}}\right) \cdot \prod_{\phi \in \Phi^+(l) \setminus \{\phi_t\}} \left(\frac{\rho_{\phi,l} \cdot \rho_{\phi_t,m}}{\rho_{\phi,m} \cdot \rho_{\phi_t,l}}\right)^{n_\phi} \cdot P_m(\Phi^-(l))} \tag{59}$$

Notice that since $P_m(\Phi^-(l))$ only considers the privacy levels ϕ for which $p_{\phi,l}$ is zero, Λ_l depends only on system inputs and $p_{\phi,m}$ for the privacy levels ϕ for which $p_{\phi,l}$ is zero. Also notice that ϕ_t is never included in $\Phi^-(l)$.

Equation 58 still depends on $p_{\phi_t,i}$, which needs to be resolved to obtain the optimal value for $p_{\phi_t,l}$. To remove this dependency, we first investigate the ratio between any pair $p_{\phi_t,i}$ and $p_{\phi_t,j}$, with the assumption that $i \notin \Phi^-(i)$ and $j \notin \Phi^-(j)$, i. e., $p_{\phi_t,i} \neq 0$ and $p_{\phi_t,j} \neq 0$. If $i \in \Phi^-(i)$, then $p_{\phi_t,i}$ does not appear in any of the other equation as $p_{\phi_t,i} = 0$. By considering Equation 58 for $l = i$ and $l = j$, and taking the ratio of the resulting equations, the following is obtained.

$$\frac{1 - p_{\phi_t,i}}{1 - p_{\phi_t,j}} = \frac{\Lambda_i}{\Lambda_j} \Rightarrow p_{\phi_t,i} = \Lambda_i \left(\frac{p_{\phi_t,j} - 1}{\Lambda_j}\right) + 1 \tag{60}$$

We insert Equation 60 into Equation 58 for every $i \neq m$, which leads to Equation 61. Notice that Equation 60 can also calculate the dependency if $i = j$, which is 1.

$$1 - p_{\phi_t,l} = \left[1 - \left(\frac{A}{a_{\phi_t,m}} - \sum_{i=1 \mid i \neq m \,\wedge\, \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \left[\Lambda_i \left(\frac{p_{\phi_t,l} - 1}{\Lambda_l}\right) + 1\right]}{a_{\phi_t,m}}\right)\right] \cdot \Lambda_l \tag{61}$$

This equation can be simplified to Equation 62. The complete derivation of this equation is shown in the appendix in Section A.2.

$$p_{\phi_t,l} = \frac{\left[A - \sum_{i=1 \mid \phi_t \notin \Phi^-(i)}^{n_\mu} a_{\phi_t,i}\right] \cdot \Lambda_l}{\sum_{i=1 \mid \phi_t \notin \Phi^-(i)}^{n_\mu} a_{\phi_t,i} \cdot \Lambda_i} + 1 \tag{62}$$

The optimal probability for the tagged vehicle to receive a message of impact-level l, given its privacy-level ϕ_t, depends on (i) the total available bandwidth A, (ii) the bandwidth consumption $a_{\phi_t,i}$ associated with the reception of messages of impact level i by the tagged vehicle of privacy level ϕ_t, and (iii) the expected specific-impacts $\mu_{\phi_t,l}$ and $\mu_{\phi_t,m}$ associated with the privacy-level of the tagged vehicle ϕ_t and the strategies associated with the other privacy-levels encapsulated in Λ_i. Notice that Λ_i only depends on the strategies associated with the privacy-levels contained in $\Phi^-(i)$. Additionally, any probability $p_{\phi_t,l}$ is always smaller than 1 as long as the available bandwidth is smaller than the bandwidth required to receive all messages. In this case, all probabilities $p_{\phi_t,l} = 1,\ \forall l \in \{1, \dots, n_\mu\}$, i. e., all messages are being received by the tagged vehicle. On the contrary, $p_{\phi_t,l}$ might also be smaller than 0 in certain situations. In this case, there is no valid solution available with the current definition of $\Phi^-(i)$ and $\Phi^+(i)$. To obtain the optimal strategy, all possible $\Phi^-(i)$ and $\Phi^+(i)$ are evaluated as described in Section 5.3.3.

Based on Equation 62, we can derive the strategy $\vec{p}_{\phi_t}$ using the strategy of all other vehicles. As these strategies are generally unknown as no explicit coordination is assumed, the vehicle needs to also calculate the strategy of the other vehicles and find their optimal strategy for every privacy-level. For this purpose, the strategy for every privacy-level is determined based on Heuristic 2. This heuristic determines the optimal strategy for each privacy-level based on the strategies of all other privacy-levels in a round-robin fashion: While the strategy of a certain privacy-level is determined, the strategies of the other privacy-levels are assumed to be fixed. This behavior is repeated until no change (within ϵ) to the strategy of any privacy-level is observed, which terminates the heuristic and the optimal strategy is determined. While we cannot prove that our heuristic terminates, there are some cases in which the heuristic can be easily shown to terminate:

1. There is no inter-dependency between privacy-levels, i. e., $\Phi^+(i) = \Phi, \forall i \in \{0, \dots, n_\mu\}$. In this case, the privacy-levels are calculated independent of each other, as each privacy-level can predict the strategy of the others.

2. There is no circular dependency, i. e., if privacy-level ϕ_1 depends on ϕ_2 as $\phi_1 \in \Phi^+(i)$ and $\phi_2 \in \Phi^-(i)$, and $\nexists j \mid \phi_1 \in \Phi^-(j) \wedge \phi_2 \in \Phi^+(j)$, i. e., ϕ_2 does not depend on ϕ_1. In this case, we can solve ϕ_2 first, and then utilize the final strategy of ϕ_2 to determine the strategy of ϕ_1.

As we cannot prove that our heuristic terminates in the general case, the iterations need to be terminated at some point and the goodness of the found strategy be checked. In our simulative analysis, we observed that the heuristic terminates yielding the correct strategies, which we attribute to the following three factors:

1. If the calculation of an impact-level i depends on any other privacy-level ϕ_o for which $\phi_o \in \Phi^-(i)$, then $p_{\phi_o,m}$ is considered in the calculation of $p_{\phi_t,l}$. $p_{\phi_o,m}$ depends on all probabilities $p_{\phi_o,i} \mid i \neq m$, i. e., possible errors of $p_{\phi_o,i}$ may compensate each other.

2. The deviation from the optimal solution influences Λ_l and Λ_i in Heuristic 62, where Λ_l is in the nominator and Λ_i is in the denominator, i. e., they potentially balance the error of the other.

3. $\exists\, j, \varphi_o \mid n^+(j) > n_{\varphi_o}$, in which case the error is reduced, as the overall influence of the error is determined by the relation between these parameters.

Heuristic 2 : Determination of the optimal strategy for all privacy-levels. **recal(...)** recalculates $p_{\varphi_t,i}$ based on the current values of $p_{\varphi,i}$. ϵ is the infinitesimal.

Result : $p_{\varphi,i}, \forall \varphi \in \Phi, i \in \{1, \ldots, n_\mu\}$
$p_{\varphi,i} \leftarrow 0, \forall \varphi \in \Phi, i \in \{1, \ldots, n_\mu\}$;
$c \leftarrow \infty$;
for $i \leftarrow 1$; $c > \epsilon$; $i \leftarrow (i \bmod |\Phi|) + 1$ **do**
  $\text{temp}_j \leftarrow p_{i,j}, \forall j \in \{1, \ldots, n_\mu\}$;
  $\text{recal}(p_{i,j}), \forall j \in \{1, \ldots, n_\mu\}$;
  $c \leftarrow \sum_{j=1}^{n_\mu} |\text{temp}_j - p_{i,j}|$;
end
return $p_{\varphi,i}, \forall \varphi \in \Phi, i \in \{1, \ldots, n_\mu\}$;

5.3.3 *Approximating the Optimal Strategy*

As already mentioned in the previous section, our approach utilizes the variables $\Phi^+(i)$ and $\Phi^-(i)$ for every impact-level i. This strategy is necessary as the bandwidth of an privacy-level might not be able to achieve the optimal solution due to the restrictions to probabilities ($p_{\varphi,i} \in [0, 1], \forall \varphi \in \Phi, i \in \{1, \ldots, n_\mu\}$). For this purpose, we analyze every possible combination of $\Phi^+(i)$ and $\Phi^-(i)$ and calculate the solution for each possible combination according to the previous chapter. As assumed by Theorem 1, it is necessary that there is one impact-level m for which $\Phi^+(m) = \Phi$, as the associated probability $p_{\varphi_t,m}$ might be used in the calculation of the strategies of the other privacy levels according to Equation 59. [1]

As for every privacy-level φ and every impact-level i either $\varphi \in \Phi^+(i)$ or $\varphi \in \Phi^-(i)$ holds, there are in the worst-case $2^{|\Phi| \cdot n_\mu}$ combinations. This leads to a worst-case complexity of our approach of $\mathcal{O}(2^{|\Phi| \cdot n_\mu})$, i. e., scales exponentially with the number of privacy-levels and impact-levels. An exponentially growing complexity of the solution is generally an issue. However, for this specific problem, the growth in complexity is generally low, as the number of privacy-levels and impact-levels is rather limited. In order to reduce the computational complexity if the computational resources are low, we can exclude solutions beforehand, such that the number of required calculations decreases. Solutions that can be excluded are:

1 While it cannot always be assumed that there is such an impact-level available, we could introduce a virtual impact-level $n_\mu + 1$, which as a $a_{n_\mu+1} = \epsilon$ and $\mu_{n_\mu+1} = 1/\epsilon$, with ϵ being the infinitesimal. This impact-level is assumed to be context-insensitive, i. e., $a_{\varphi,n_\mu+1} = a_{n_\mu+1}, \forall \varphi \in \Phi$ and $\mu_{\varphi,n_\mu+1} = \mu_{n_\mu+1}, \forall \varphi \in \Phi$. Due to the high expected specific-impact and the context insensitivity, this impact-level is always subscribed to by every vehicle but does not consume a relevant amount of bandwidth when being subscribed to. The obtained solution is then still an approximation, but the approximation is very close to the optimal value and depends on the value for ϵ used in the actual implementation.

1. Solutions in which $\exists i \in \{1, \ldots, n_\mu\} \mid \Phi^-(i) \neq \emptyset \wedge \Phi^+(i) \neq \emptyset$, i. e., we exclude all solutions in which the privacy-levels do not subscribe to the same impact-levels. This reduces the computational complexity to $\mathcal{O}(2^{n_\mu})$, as each impact-level can be either active or passive.

2. Solutions in which $\exists i \mid \phi \in \Phi^-(i) \wedge \phi \notin \Phi^-(i+1)$, i. e., a vehicle needs to subscribe to all impact-levels above the impact-level with the lowest index (ordered by the impact $\mu_{0,i}$). Removing these solutions also reduces the possibility of implicit coordination, but reduces the computational complexity to $\mathcal{O}(n_\mu^{|\Phi|})$, as there are n_μ possibilities per privacy-level which need to be combined with the possibilities of all other privacy-levels.

3. Solutions in which any of the conditions are fulfilled. In this case, the number of calculations is linear with the number of impact-levels, i. e., the computationally complexity reduces to $\mathcal{O}(n_\mu)$.

While all three possibilities greatly reduce the computational overhead induced by our approach, its performance also drops. In general, our approach can be refined with increasing computation time, starting with the solutions from 3, followed by 1 and 2 starting with the less computational expensive, and finally calculating the remaining combinations. That way, the calculation can be aborted at any time and still provide a reasonable result.

5.4 DETERMINING THE PRIVACY ADAPTATION FACTOR

The privacy adaptation factor $\rho_{\phi,i}$ determines the imprecision area which affects the context-based dissemination of messages. For a given impact-level i, the relation between the resulting context-based dissemination scope (e. g., an area) of a given event and the original dissemination scope (under no privacy concerns) is captured by the privacy adaptation factor $\rho_{\phi,i}$ defined to be the ratio of these dissemination scopes (e. g., areas). Thus, the negative influence of location privacy to the system is strongly correlated with the associated distribution mechanism. As the vehicle provides an area in which it is certainly located with a uniform distribution, the only way to ensure the reception of messages is by providing all messages that are relevant to *any* vehicle in this area. For context-sensitive data, location privacy thus impacts the area in which data is considered to be relevant. With increasing size of this area, the load for the vehicles increases drastically. In this section, we describe the influence of location privacy to the data distribution mechanisms described in this thesis and propose mechanisms to determine $\rho_{\phi,i}$ for each described data distribution mechanism.

5.4.1 *Area-based Dissemination*

Consider an event of impact-level i and a circular dissemination area of radius η_i. Any vehicle whose exact location is within this area will be receiving this event. For a non-privacy-sensitive vehicle, the radius $\kappa_{\phi,0}$ of the dissemination area $\kappa_{\phi,0} = \eta_i$. Suppose now that a tagged such vehicle would like to hide its exact location within a circular imprecision area of radius r_ϕ. Since the maximum displacement of the exact location of this vehicle is r_ϕ, the reception of the event by such a vehicle will be guaranteed as long as the radius η_i of the original circular dissemination area is extended by r_ϕ. The latter ensures reception of the event by a vehicle whose exact location is on the periphery of the original circular dissemination area of radius η_i and on the periphery of its imprecision area with radius r_ϕ. As data relevant to any vehicle in the area needs to be provided, $\kappa_{\phi,i} = \kappa_{\phi,0} + r_\phi = \eta_i + r_\phi$, where r_ϕ is the size of the area hiding the vehicle's real location. The adaptation factor is shown in Equation 63. It assumes that the area in which the vehicle hides is of the same geometrical form as the message dissemination area. In this equation, ζ is the relation between the actual size of the interest area compared to the size of the maximum circle fitting inside this area, which is $\zeta = {}^4/_\pi$ for squares and $\zeta = 1$ for circles.

$$\rho_{\phi,i} = \frac{\zeta \cdot \pi \cdot \kappa_{\phi,i}^2}{\zeta \cdot \pi \cdot r_i^2} = \frac{\eta_i^2 + 2 \cdot \eta_i \cdot r_\phi + r_\phi^2}{\eta_i^2} = 1 + 2 \cdot \frac{r_\phi}{\eta_i} + \left(\frac{r_\phi}{\eta_i}\right)^2 = \left(\frac{r_\phi}{\eta_i} + 1\right)^2 \quad (63)$$

5.4.2 *Road-based Dissemination*

For *road-based dissemination*, the adaptation factor is more context-sensitive than for the *area-based dissemination*. That is, as the events are not necessarily uniformly distributed. In general, a road segment has at least one incoming road segment. Thus, the number of road segments the vehicles retrieves data from seems to grow exponentially with the length of the maximum path. However, this exponential growth is only true for very short paths, as the possibility of segments that appear in more than one path increases. As every road segment occupies a certain space, and the maximum length of a path limits the available area, the scaling in the number of road segments is not exponential, but only squared. Thus, the adaptation factor $\rho_{\phi,i}$ can be estimated using the length of the maximum path (in meters) and utilizing Equation 63.

It is evident that this upper-bound does not accurately describe the additional overhead induced through privacy, which reduces the performance of our approach. To more accurately capture it, the vehicle can determine the adaptation factor $\rho_{\phi,i}$ for its current context by determining the segments contained in the imprecision area and determining the relevant road segments for each of these segments. The adaptation factor $\rho_{\phi,i}$ is then the ratio between the number of segments for the privacy-level ϕ and the number of road segments for no privacy-sensitivity.

5.4.3 *Relevance-based Dissemination*

The adaptation factor $\rho_{\phi,i}$ for the road-based dissemination can be determined similarly to the *road-based dissemination*. For this purpose, either the length of the longest path in conjunction with Equation 63 can be utilized to estimate $\rho_{\phi,i}$, or the number of road segments with and without privacy are compared to obtain $\rho_{\phi,i}$.

5.5 CONSIDERATION OF THE SPECIFIC-IMPACT IN THE DISSEMINATION OF MESSAGES

In this section, we describe the consideration of the specific-impact in the message dissemination. As described in the previous sections, vehicles subscribe to messages based on the base-impact. While the strategy of the vehicle is determined by estimating the average specific-impact of a message in each impact-level, mechanisms for the consideration of the specific-impact of each message need to be proposed.

There are multiple options to address this issue: (i) let every impact-level address a range of specific-impact values, (ii) create a new impact state for every possible specific-impact value, or (iii) reduce the influence of lower-impact messages by reducing the probability of them being transmitted on the server-side. The first option induces inaccuracy into the system, as the specific-impact value of a message in an impact-level can only be estimated. Thus, the found solution is suboptimal. To address this issue, the second option removes that inaccuracy, but induces a severe scalability issue, as the number of possible impact states is very high and the optimal strategy can hardly be determined. Thus, we decided to use the third option, in which the lower-impact messages are only transmitted with a certain probability $p_{\text{server}}(\vec{m})$ on the server-side, which depends on the measurement $\vec{m}$. This leads to a lower bandwidth consumption of these messages, as only a share is transmitted, and ensures the transmission of low-impact measurements (of that high-impact state) with a low probability. To determine the probability, we divide the specific-impact $\mu(\vec{m}, t)$ of a measurement $\vec{m}$ by the base-impact of the message as shown in Equation 64. This leads to an implicit prioritization of highly relevant messages, as the transmission probability for messages with low specific-impact is reduced. Notice that per definition, the base-impact is represented by one impact-level.

$$p_{\text{server}}(\vec{m}) = \frac{\mu(\vec{m}, t)}{\mu_{0,i}} \tag{64}$$

To decide on the vehicles that receive a certain measurement $\vec{m}$, the server creates a random value $r \in [0, 1]$. The server transmits the message to a vehicle if (i) the vehicle has stated its interest in the corresponding impact-level, and (ii) the probability for the vehicle to receive this message is larger than r. As events are generally measured by multiple vehicles, this probabilistic filtering in principle reduces the update frequency of these messages to reduce the overall network load.

For this purpose, we describe the handling of the measurement-specific influence on the specific-impact in Section 5.5.1, followed by the handling of the context-specific influence on the specific-impact in Section 5.5.2.

5.5.1 *Consideration of Inaccurate Measurements and the Event State*

One influence on the specific-impact is the measurement itself, including its inaccuracy and the measured event state. As an example, the server has two measurements

$\vec{m}_1 = (0,1)^T$ and $\vec{m}_2 = (1,0)^T$ of type t, which has an impact vector $\vec{\mu}_t$ as shown in Equation 65.

$$\vec{\mu}_t = \begin{pmatrix} 1 \\ 1000 \end{pmatrix} \tag{65}$$

Thus, we can derive $\mu(\vec{m}_1, t) = 500$ and $\mu(\vec{m}_2, t) = 0.5$ based on the calculation for the specific-impact of a message. Thus, we would transmit $\vec{m}_1$ with a probability of $p_{\text{server}}(\vec{m}_1) = {}^{500}/_{1000} = 50\%$ and $\vec{m}_2$ with a probability of $p_{\text{server}}(\vec{m}_2) = {}^{0.5}/_{1000} = 0.05\%$.

This reduces the expectation for the bandwidth requirements $a_{0,i}$ of the corresponding impact state i to the corresponding probability, but we need to guarantee that the expected specific-impact of that specific measurement equals the base-impact $\mu_{0,i}$ of the corresponding impact-level $\mu_{0,i}$. For this purpose, we analyze the provided information if the measurement is received by one specific vehicle: While only one message is received, the vehicle can assume the existence of a certain number of non-transmitted measurements based on its knowledge about the message's relevance for itself and the measurement accuracy. These two factors are both considered to be known by the vehicle, either by deriving them from the actual message/measurement or by the server explicitly communicating them to the vehicles. Based on these factors, the receiving vehicle can estimate the average number $\overline{n}_{miss}$ of missed measurements with a similar specific-impact $\mu(\vec{m}, t)$, which is reciprocal to the probability $p_{\text{server}}(\vec{m})$ for $\vec{m}$ being transmitted, as shown in Equation 66.

$$\overline{n}_{miss} = \frac{1}{p_{\text{server}}(\vec{m})} \tag{66}$$

Based on that, by transmitting $\vec{m}$, the server implicitly provides information about the missed measurements. This information can be used in the measurement aggregation process by weighting the measurement accordingly. This induces some error if the probability is comparably low, as a single transmitted measurement has a high weight in this case. However, since measurements are pre-aggregated by either the vehicles locally or by the server, this effect is considered to be low. Notice that high-impact measurement (in case an event in a high-impact state) are still communicated rapidly, while communicating the resolution of this high-impact state might take slightly longer than normal. The effect of this behavior is, however, bound by the maximum lifetime of measurements, which invalidates high-impact measurements after a predefined time. These considerations change the effective specific-impact $\mu_{\text{eff}}(\vec{m}, t)$ of transmitting such a measurement $\vec{m}$ to the combined specific-impact of all these missed measurements as shown in Equation 67.

$$\mu_{\text{eff}}(\vec{m}, t) = \mu(\vec{m}, t) \cdot \overline{n}_{miss} = \frac{\mu(\vec{m}, t)}{p_{\text{server}}(\vec{m})} \tag{67}$$

When replacing $p_{\text{server}}(\vec{m})$ according to its definition in Equation 64, we obtain Equation 68.

$$\mu_{\text{eff}}(\vec{m}, t) = \frac{\mu(\vec{m}, t)}{p_{\text{server}}(\vec{m})} = \frac{\mu(\vec{m}, t) \cdot \mu_{0,i}}{\mu(\vec{m}, t)} = \mu_{0,i} \tag{68}$$

Thus, the effective specific-impact $\mu_{\text{eff}}(\vec{m}, t)$ equals the base-impact $\mu_{0,i}$ of the corresponding impact-level i, which enables the transmission of lower-accuracy measurement using this method by considering the non-transmitted measurements.

5.5.2 *Consideration of the Vehicle's Context*

The vehicle's context influences the specific-impact for a specific vehicle. While messages with a low specific-impact are considered to be less relevant for the vehicle, it is reasonable to transmit them with a low transmission frequency to reduce the effects of bad prediction. Due to this low transmission frequency, the number of available messages with low specific-impact is small, but the vehicle is still aware that the event exists. Thus, our probabilistic server-side filtering decreases the rate with which messages with low relevance are provided to the vehicle.

However, this approach reduces the bandwidth used for the transmission of messages, which might lead to a vehicle not utilizing its bandwidth fully in certain situations. To account for this issue, each vehicle transmits the share of utilized bandwidth to the server. With that, it is possible for the server to adapt its filtering if the vehicle does not fully utilize its available bandwidth. In this case, the server adapts the transmission probability such that the expected bandwidth consumption matches the available bandwidth of the vehicle (if possible). With that, the unused bandwidth of the vehicles is minimized, which further increases the performance of our approximate vehicular network.

The definition of relevance proposed in Section 4.3 can be used to influence the specific-impact, and, thus, the dissemination of messages. However, also other dissemination mechanisms might be viable and can be used to limit the dissemination of messages to, for example, a certain geographical area using the specific-impact. Area-based dissemination can be, for example, employed by using a definition of relevance according to Equation 69.

$$R(s_v, s_e) = \begin{cases} 1 \text{ if } s_v \text{ in area around } s_e \\ 0 \text{ else} \end{cases} \tag{69}$$

In this definition, a message is considered to be relevant if the vehicle is currently located in an area around the event. While this definition is very inaccurate regarding its approximation of interest as described in Section 4.3, it may already be in use for the filtering of messages in existing networks. With this redefinition of relevance, our mechanisms for approximate vehicular networks can be used with most available message dissemination mechanisms, by adapting the relevance calculation such that the dissemination behavior is reflected.

5.6 SYSTEM ANALYSIS

In this section, we describe the properties of our found solution and analyze the behavior of this solution in a controlled setting to gain insights into the behavior of our approach.

5.6.1 *Properties of the Found Solution*

In the following, we describe three properties of our solution, which are (i) optimality, (ii) stability, and (iii) coordination. Optimality and stability are important properties for the applicability of our found solution in vehicular networks, while coordination can further increase the performance of the network.

5.6.1.1 *Optimality*

In Section 5.3, we found solutions that potentially maximize the utility of each individual vehicle, given the strategies of the other vehicles. However, we did not analyze if the found solution is an optimal value, and if it is a local or global optimum, given it is an optimal value. One possibility to prove the optimality of the found solution is to prove that the expected utility is a concave function. We prove that by showing that the Hessian matrix is negative definite. In the function of the expected utility function, the derivative

$$\frac{\partial^2 \overline{u}}{\partial p_{\phi_t,i} \partial p_{\phi_t,j}} = 0$$

for $i \neq j$, as there is no dependency between any two probabilities of the same strategy. Thus, it is sufficient to show that the diagonal of the Hessian matrix smaller than 0, what we do by calculating the second derivative of the expected utility with respect to any $p_{\phi_t,l}$ as shown in Equation 70.

$$\begin{aligned}
\frac{\partial^2 \overline{u}}{\partial^2 p_{\phi_t,l}} &= -\mu_{\phi_t,l} \cdot a_{\phi_t,l} \cdot n_{\phi_t} \cdot (n_{\phi_t} - 1) \cdot (1 - p_{\phi_t,l})^{n_{\phi_t}-1} \cdot P_l(\Phi \setminus \{\phi_t\}) \\
&- \mu_{\phi_t,m} \cdot a_{\phi_t,m} \cdot \left(\frac{\partial p_{\phi_t,m}}{\partial p_{\phi_t,l}}\right)^2 n_{\phi_t} \cdot (n_{\phi_t} - 1) \cdot (1 - p_{\phi_t,m})^{n_{\phi_t}-1} \cdot P_m(\Phi \setminus \{\phi_t\})
\end{aligned} \tag{70}$$

Notice that the second derivative is always smaller than 0 in the allowed range of parameters if $p_{\phi_t,l} < 1$ and $n_{\phi_t} \geqslant 2$. That is, as every factor in the first summand is greater 0, which leads to the product to be always greater than 0. In our optimization problem, $p_{\phi_t,l} = 1$ will only occur iff the bandwidth is sufficient to receive everything, as otherwise this bandwidth utilization is comparably inefficient. If $n_{\phi_t} = 1$, we utilize a similar procedure as for the non-cooperative case, which is not captured by our considerations for cooperative vehicular networks. The same holds for the second summand, but its dependency on the other privacy-levels through $P_m(\ldots)$ might lead to the second summand being 0. As the first summand is always negative and the second summand is at most 0, we can observe that the Hessian matrix is negative definite. Thus, the expected utility is a concave function in the allowed range of the parameters. One property of a concave function is that each found optimum is always a global optimum. Thus, our found solution is a global optimum, that guarantees that there is no other strategy that achieves a higher performance for this specific setup considering the strategies of the other privacy-levels.

5.6.1.2 *Stability*

For our developed solution to work, every vehicle needs to stick to the calculated strategy. If a vehicle could improve its own utility (and potentially the utility of others), it could derivate from the developed strategy and, thus, increase its own benefit. In non-cooperative games, a state in which an actor has no incentive to change its strategy is called Nash equilibrium. A Nash equilibrium is reached if there no possibility for any actor to improve its decision while the decision of the other actors remains the same. If this holds true for every actor in the system, there is no incentive for the overall system to change, leading to a stable system. A Nash equilibrium can be reached either based on pure strategies or, as in our case, with mixed strategies. To show that our found solution is a Nash equilibrium, we optimize the strategy $\vec{q}$ of the tagged vehicle independently of the other vehicles. We assume that all other vehicles stick to the previously derived solution, while the tagged vehicle aims to improve its utility u^*, which is shown in Equation 71.

$$\overline{u}^* = \sum_{i=1}^{n_\mu} \mu_{\phi_t,i} \cdot a_{\phi_t,i} \cdot \left[1 - P_i(\Phi \setminus \{\phi_t\}) \cdot (1 - p_{\phi_t,i})^{n_{\phi_t,i}-1} \cdot (1 - q_i)\right] \tag{71}$$

The dependency of q_l and q_m remains the same as for the other vehicles, as the tagged vehicle also needs to stick to its bandwidth constraints. When calculating the derivative of the modified expected utility u^*, we derive Equation 72.

$$\begin{aligned}\frac{\partial \overline{u}^*}{\partial q_l} &= \mu_{\phi_t,l} \cdot a_{\phi_t,l} \cdot (1 - p_{\phi_t,l})^{n_{\phi_t}-1} \cdot P_l(\Phi \setminus \{\phi_t\}) \\ &\quad + \mu_{\phi_t,m} \cdot a_{\phi_t,m} \left(\frac{\partial q_m}{\partial q_l}\right) \cdot (1 - p_{\phi_t,m})^{n_{\phi_t}-1} \cdot P_m(\Phi \setminus \{\phi_t\}) \stackrel{!}{=} 0\end{aligned} \tag{72}$$

As stated previously, the dependency of q_l and q_m persists and is similar to all other vehicles. When comparing Equation 72 with the derivative of the expected utility shown in Equation 44, we can observe the similarity between these two equations. In fact, the only difference between them is the factor n_{ϕ_t}, which is only contained in the derivative of the expected utility, but not in the derivative of the modified expected utility. This relation is shown in Equation 73.

$$\frac{\partial \overline{u}^*}{\partial q_l} \cdot n_{\phi_t} = \frac{\partial \overline{u}}{\partial p_{\phi_t,l}}, \forall\, l \in \{1, \ldots, n_\mu\} \tag{73}$$

As we already showed that the derivative of the expected utility is 0, we utilize this knowledge to proof the correctness of Equation 72 as shown in Equation 74.

$$\frac{\partial \overline{u}}{\partial p_{\phi_t,l}} = 0 \Rightarrow \frac{\partial \overline{u}^*}{\partial q_l} = 0, \forall\, l \in \{1, \ldots, n_\mu\} \tag{74}$$

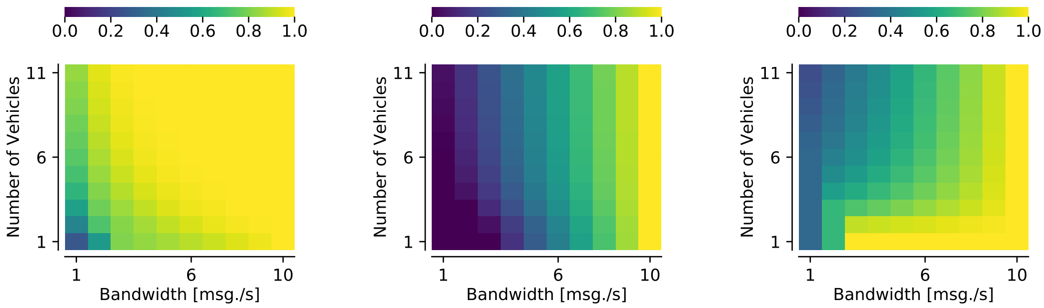
This, however, does not necessarily state that there is no other possible solution with a higher utility, but only states that the found solution is a local optimum. The proof of the non-existence of another global optimum with a higher utility again requires the analysis of the second derivative with respect to the modified expected utility. As Equation 72 does not depend on q_l anymore, the second derivative is always 0. Thus, there is no better solution available to the vehicle, i. e., there is no incentive for the vehicle to adapt its strategy. Due to the fact that the decision of the vehicle, given it utilizes its bandwidth fully, does not decrease the vehicle's performance, it is crucial that every vehicle in the network knows the derivation of the optimal strategy. Otherwise, the vehicles might deviate from the optimal solution, as a change does not decrease their performance immediately. However, if all vehicles change their strategy, the performance of the overall system will decrease.

5.6.1.3 *Coordination*

Through the introduction of multiple privacy-levels, coordination between these becomes a possibility. That is, a certain privacy-level may be responsible for the reception of a certain impact-level and all other privacy-levels ignore this impact-level, i. e., set the corresponding probability of reception to 0. The advantage of this coordination is the reduced redundancy compared to a shared responsibility, but this is at costs of reliability: If there is only one vehicle at a certain privacy-level, which is responsible for the reception of a certain impact-level, a disconnect from this specific vehicle would have similar consequences as disconnecting from a cluster-head. Thus, further questions arise, like the minimum number of vehicles at a privacy-level such that other vehicles can rely on that privacy-level.

The solution for this issue is based on the fact that the different adaptation factors $\rho_{\phi,i}$ for each privacy-level. As completely relying on another vehicles is hardly appropriate for high-impact messages, we do not only use the solution with the highest utility as described in Section 5.3, but also consider the performance of the vehicle if all other vehicles fail. Thus, we have two utility values based on the current strategy of the vehicle, one considering the cooperation with others, while the other accounts for the case of failure. With this combination, there might be better solutions available for a vehicle than the chosen one from a model perspective, which are not appropriate in a realistic setting. Thus, in a realistic setting, we cannot guarantee that there is no better strategy available for a vehicle and rely on cooperation between them.

To choose the optimal strategy for the vehicle in real-world environments, the vehicle combines these two utility values using the trust factor τ. τ is a value between 0 and 1, where 1 states full trust on the vehicles in proximity, i. e., choosing the strategy only based on the cooperative utility $\overline{u}_c$, and 0 states no trust on the vehicles in proximity,



(a) Utility.

(b) Probability of Impact Level 1.

(c) Probability of Impact Level 2.

Figure 7: Behavior of our game-theoretic approach depending on the available bandwidth.

i. e., choosing the strategy only based on the utility $\overline{u}_f$ in case of full failure. Between these values, the final utility $\overline{u}_t$ is derived from a linear combination of $\overline{u}_c$ and $\overline{u}_f$ as shown in Equation 75.

$$\overline{u}_t = \tau \cdot \overline{u}_c + (1 - \tau) \cdot \overline{u}_f \tag{75}$$

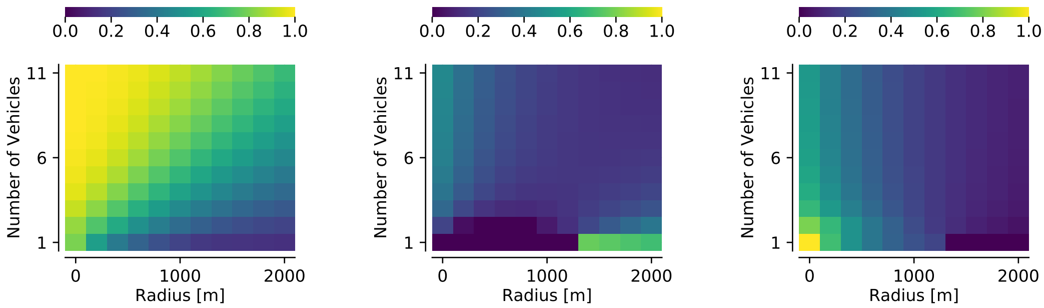
5.6.2 *Numerical Analysis*

In this section, we analyze the behavior of our game-theoretic approach in a controlled environment to gain a better understanding of the behavior of our developed approach. For this purpose, we analyze the influence of the available bandwidth per vehicle and of privacy, both depending on the size of the neighborhood.

5.6.2.1 *Influence of the Available Bandwidth per Vehicle*

A major influence to the performance of our game-theoretic approach is the available bandwidth per vehicle A. That is, a high bandwidth enables the reception of more messages, such that each individual vehicles is more self sustained and can also offer more to the vehicle in its proximity. To display the effects of the available bandwidth, we consider a scenario with two impact-levels, in which all vehicles are privacy-insensitive. The first impact-level has an expected specific-impact value $\mu_0 = 1$ and consumes bandwidth of $a_i = 7$. The second impact-level has an expected specific-impact value of $\mu_1 = 10$ and consumes bandwidth of $a_i = 3$. Thus, a bandwidth of 10 is required to receive all available messages.

In Figure 7, all figures analyze the influence of the available bandwidth A on the x-axis, while the number of vehicles in proximity is on the y-axis. When we investigate the utility in Figure 7a, we can observe that both the available bandwidth and the number of vehicles in proximity increase the utility for the tagged vehicle. For the bandwidth, a high available bandwidth increases the system performance to the maximum, while the effect of the number of vehicles in proximity strongly depends on the available



(a) Utility. (b) Probability of Impact Level 1. (c) Probability of Impact Level 2.

Figure 8: Behavior of our game-theoretic approach depending on the imprecision area.

bandwidth. That is, as a high available bandwidth enables the reliable reception of messages for a vehicle. This is not possible by increasing the number of vehicles in proximity due to the lack of explicit coordination. When investigating the strategy of the tagged vehicle depicted in Figure 7b, we can observe that the first impact-level (with lower specific-impact) is only considered if the available bandwidth increases. That is, as the specific-impact value of the second impact-level is significantly higher than the specific-impact value of the first impact-level, such that a loss of a high-impact messages can hardly be compensated by the lower impact-level. This is also observable in Figure 7c, in which the probability remains relatively constant if the available bandwidth is low. However, when the bandwidth increases, we can clearly observe that $p_{\phi_t,2}$ is decreased such that bandwidth can be used to receive messages of the first impact-level. This effect then disappears when the bandwidth reaches 10, as all messages can be received by all vehicles.

5.6.2.2 *Influence of the Radius of the Imprecision Area*

Similar to the available bandwidth, the radius of the imprecision area has a major influence to the utility of the tagged vehicle. In Figure 8, we display the geocast-based dissemination of two impact-levels. The first impact-level has a dissemination radius of $r_1 = 10000$ and an associated expected specific-impact value of $\mu_1 = 1$. The second impact-level has a dissemination radius of $r_2 = 1000$ and an associated expected specific-impact value of $\mu_2 = 4$. With that setup, the impact $\mu_{\phi,1}$ considering the privacy-level ϕ is lower than $\mu_{\phi,2}$ until $r_\phi \geqslant 1250$. The bandwidth requirements are similar for both impact-levels, and without privacy the available bandwidth is sufficient to receive either one of them completely. To analyze the influence of the imprecision area, we vary the radius r_ϕ of this area on the x-axis, and the number n_ϕ of vehicles in that privacy-level on the y-axis.

Figure 8a shows that the utility decreases with an increasing radius of the imprecision area, and increases with an increasing number of vehicles. However, it is very interesting to analyze the degree of performance loss due to the increasing size of the imprecision area depending on the number of vehicles. While the decrease for a single vehicle (without cooperation) is above 82%, the performance decrease with 11 vehicles is below 33%. Even though these vehicles all have the same privacy-level, cooperation still reduces the negative influence of privacy. Analyzing the behavior of the strategy of the vehicles is very interesting due to the change in expected specific-impact due to the change in the imprecision area. Figure 8b displays the probability to receive the impact-level with a lower base-impact. For a small imprecision area, this impact-level is only considered if the number of vehicles is really high. But starting from $r_\phi = 1250$, the vehicles start focusing on this impact-level, as the expected specific-impact per bit of this impact-level is higher due to the lower influence of privacy to this impact-level. This is most visible for the case without cooperation, in which the vehicle completely switches from receiving only messages of impact-level 2 to receiving only messages from impact-level 1. In contrast, Figure 8c displays the focus of the vehicles to impact-level 2, which is far above impact-level 1 at the beginning. However, with an increasing size of the imprecision area, an increasing amount of bandwidth is used to compensate for the additionally required bandwidth usage for the reception of these messages. Starting from $r_\phi = 1250$, the vehicles focus on the reception of impact-level 1.

Summarizing, our system dynamically adapts to the environmental conditions and only receives low-impact messages if the reception does not decrease the performance of high-impact messages drastically. In addition, our approach considers the privacy of the vehicles and its influence on the available impact-levels to adapt to the environmental conditions.

In this chapter, we developed our innovative approach to *approximate vehicular networking* based on Diverse Prioritization and Treatment (DPT), which improves the performance of the network without increasing the utilized communication resources. Our basic concept is the reduction of redundancy introduced through prioritization in the network. We employ a probabilistic mechanism derived from a utility-based game to reduce this redundancy and, thus, free bandwidth for the reception of previously not received messages. In our numerical analysis in Section 5.6, we already showed the behavior of our approach in a controlled environment. Our approach to approximate vehicular networking is additionally evaluated in Section 7.4, where we show the performance increase achieved through our concept of approximate vehicular networking and highlight the robustness of our approach against message loss and adversaries.

6 VEHICLE.KOM: PLATFORM FOR APPROXIMATE VEHICULAR NETWORKS

To evaluate the performance of our approximate vehicular network, we need to design and implement the possibility to evaluate large-scale vehicular networks in a network simulator. For this purpose, we introduce *Vehicle.KOM*, a scalable platform for the simulation of approximate vehicular networks. Our platform is based on the Simonstrator.KOM framework [150, 164] and enables the evaluation of approximate vehicular networks, which might consider the quality of data for prioritization. In Section 6.1, we present our extensions to the Simonstrator platform that are necessary to simulate large-scale vehicular networks. For this purpose, we provide an overview of our extensions to the Simonstrator.KOM platform and describe the design decisions for the platform components in Section 6.1.1, and our separation of networking and application logic in Section 6.1.2. Based on our *Vehicle.KOM* platform, we create a prototypical implementation of approximate vehicular networks based on the Publish/Subscribe (Pub/Sub) paradigm in Section 6.2. Our prototypical implementation consists of our mechanisms for data assessment in Section 6.2.1, and our impact-based data dissemination in Section 6.2.2.

6.1 OVERVIEW OF THE VEHICLE.KOM PLATFORM

In this section, we describe our VEHICLE.KOM platform and the necessary extensions of the underlying Simonstrator platform. While we utilize the existing channel models and components of the Simonstrator platform, large changes were necessary to allow for the simulation of large-scale vehicular networks. Similar to the Simonstrator platform, we provide reproducibility based on seeds, which are used to configure both the networking and the movement of the vehicles. When changing these seeds, the influence of randomness to the approaches can be investigated. In the following, we describe the design and implementation of the platform components. Next, we describe the encapsulation of networking and application logic to analyze the influence of dissemination mechanisms to the system performance.

6.1.1 *Platform Components*

Similar to the Simonstrator framework, our Vehicle.KOM is event-driven and each host (vehicle, Road Side Unit (RSU), server) is assigned a set of components defining its behavior. This behavior includes communication, message storage and processing, and decision-making as shown in Figure 9. These components are derived from the scenario described in Chapter 3. In the following, we provide a detailed overview of the components and their interaction in our *VEHICLE.KOM* platform. While RSUs can

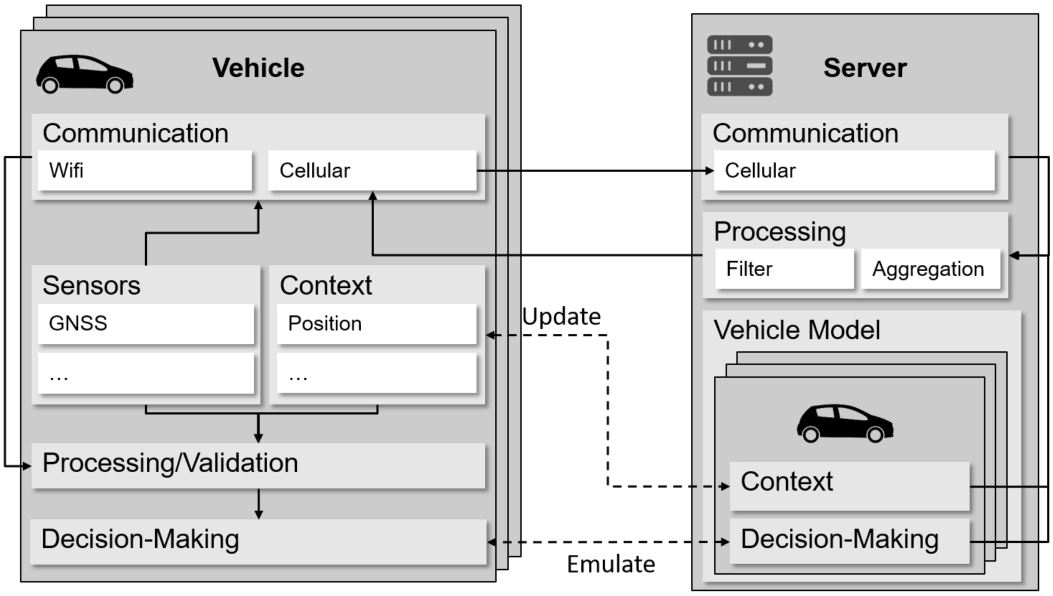


Figure 9: Components and their interaction in the Vehicle.KOM platform.

be modeled in our simulation framework, we omit them in this description due to the neglectable influence on our developed approaches.

6.1.1.1 *Road Event Modeling*

As described in Chapter 3, the measurable road events have a significant influence on the load on the vehicular network. Thus, the configurability of road events is an important aspect of our VEHICLE.KOM platform. Each road event has a location and associated road segment, a creation time, and an expected and actual lifetime. Every data entry created by a vehicle contains all these meta-information, but approaches do not access hidden simulation knowledge like the actual event lifetime.

To accurately model the network load in vehicular networks, we created events of different severity, which are fog, bumps, hazards, jams, rain, and traffic signs. While these are generally sufficient to create a diverse network load, new properties can be added easily due to the extensible design of our environment properties. This extensible design allows the creation of new road events in the simulation by adjusting the event-specific properties like lifetime and appearance probability.

Based on these properties, the events are generated using an extensible event-generator. This event generator can be extended with plugins, which are responsible for the generation of exactly one event type. As an input, a plugin requires the lifetime distribution of the event, the share of active events in the road network, and the type of the event that is generated. For the lifetime, both an exponential lifetime distribution and a Gaussian distribution are available.

When the simulation starts, there are no active events in the road network, i. e., every road is in its default state. After a configurable warmup-time, the event generator plugins generate events with a randomly selected location, and a randomly selected lifetime from its lifetime distribution until the share of active events is reached. Additionally, an event can be assigned a value parameter that may change over time to emulate changing road events, like the increasing average vehicle speed in a dissolving jam. The generated value of a road event can be either discrete or continuous. We use a Markov chain to determine the transitions between the states of the variable for both continuous and discrete values. To accurately choose values for continuous variables, we interpolate the value linearly between the two states. If an event turns active during the simulation, a new event is generated. In our generation process, we prevent the generation of the *same* event type at the same segment, but do not restrict the total number of events that are active on one segment. Thus, the event types are generated independently from each other and the share of active events is between 0 and 1.

6.1.1.2 *Vehicle Modeling*

When modeling the vehicles' behavior in a vehicular network, a realistic movement of the vehicles is pivotal due to the context-sensitivity of vehicular networks. The Simonstrator natively only supports a map-based movement targeted at human mobility [152], which does not satisfy the requirements of realistic vehicle movement. For realistic vehicle movement, we extended the Simonstrator framework by connecting it to the traffic simulator Simulation of Urban Mobility (SUMO) [110], which is considered accurately model the movement of vehicles. We connect our VEHICLE.KOM platform to SUMO using the Traffic Control Interface (TraCI) [183], which enables controlling SUMO via a network socket. While the number of hosts in the Simonstrator is considered to be stable, the number of active vehicles in SUMO may change during the simulation. We address this issue by initializing a high number of offline hosts in our VEHICLE.KOM platform, which are managed by the *VehicleMovement* class. Once a new vehicle becomes active in SUMO, the *VehicleMovement* selects one previously inactive hosts and binds it to the respective vehicle in SUMO.

This binding is managed by the *VehicleInformationComponent*, which also offers the possibility to manipulate the vehicle movement in SUMO, i. e., by breaking, accelerating, our detouring. The *VehicleInformationComponent* is accessible by every application running on a vehicle. It provides the possibility to retrieve and change the vehicle's path and speed. After the vehicle is inactive in SUMO, it is removed from the Simonstrator by turning the host's network interfaces offline.

The vehicles and their components are displayed on the left side of Figure 9. Each vehicle can sense its environment using its on-board sensors, which can be adjusted to provide data of varying quality. The data from these sensors are both processed locally and shared with other vehicles. For communication with other vehicles, two network interfaces are available, a local communication interface based on Wifi and a cellular communication interface. The communication of data via Wifi does not require the availability of a central server, while cellular communication is managed by the backend. Incoming data are merged with locally perceived data to generate the environmental model in each vehicle. Based on this environmental model, each vehicle may make decisions regarding its future path, braking, and other adaptations of driving behavior. These adaptations are executed using a bidirectional link to SUMO, which enables the manipulation of vehicle behavior. Thus, not only the network-specific metric, but also vehicle-specific metrics like driving duration and average speed can be used in our simulation.

6.1.1.3 *Backend Modeling*

The backend is responsible for coordinating the distribution of data provided by the vehicles. For this purpose, the server has processing capabilities, which can (i) preprocess data before providing it to the vehicle, which is necessary if the available resources for communication or processing are not necessary, and (ii) filter data to provide only relevant data to each vehicle. Additionally, the server can adapt the filtering to the available resources, as the different filtering mechanisms have different processing and networking demands. To determine the relevance of data for a vehicle, the server manages a *virtual vehicle* for each real vehicle driving on the streets. This *virtual vehicle* contains all the necessary data to enable the filtering process of the server. There are two possibilities to model this virtual vehicle, (i) mirror only the context of the vehicle or (ii) mirror the context and the decision-making of the vehicle. The latter enables the consideration of active vehicular applications to model the necessity of data for a vehicle more accurately. This application-aware mirroring might be computationally expensive, which might require a transition to context-only filtering if the backend resources are not sufficient.

6.1.1.4 *Interaction between Backend and Vehicles*

The vehicles need to provide data to the backend to enable the dissemination of road-related data, which are (i) road-based data sensed by the vehicles, (ii) the current context of the vehicles, and (iii) a description of the currently running applications of the vehicles. The server uses the provided road-based data to either to (i) create an environmental model and provide this model to the vehicles or (ii) directly provide the raw input to the vehicles. The current context of the vehicle and the description of the running applications are used to update the *virtual vehicle*. Depending on the current network load and available backend resources, the server can transition between different representations of the *virtual vehicle*, which changes the required data. The update frequency of the vehicle's context can also be adapted to reduce the load on the network if the amount of payload is low.

6.1.2 *Separation of Networking and Application Logic*

For an efficient analysis of the performance of different networking approaches, the separation of networking and application logic is a pivotal aspect. That is, as today's vehicular applications adapt to the network conditions very rarely, i. e., such an adaptation cannot be assumed for our scenario. Consequently, we model the application logic independently of the networking logic to analyze the influence of our networking approaches to the running vehicular applications. This analysis is pivotal to assess the performance of data-centric vehicular networks, as the achieved performance is heavily connected to the applications running on the vehicle and the vehicle's context in general.

For the separation of networking and application logic, we employed two classes, the *VehicularSender* and the *VehicularReceiver* interface. These interfaces utilize the Pub/Sub paradigm as a basis to organize the data-centric dissemination of road data. They additionally support the communication between vehicles without consideration of the underlying communication technology. Thus, these interfaces enable the exchange of the communication method without any changes to the application.

Listing 1 displays the available methods for the transmission of data to other vehicles. For this purpose, the vehicle packs the available data into the EnvironmentInformation class, which can contain any road-related and environment-related data. This container can then either be transmitted to vehicles in proximity using the local Wifi interface (sendLocally(...)), or transmitted to a central backend, which distributes the data to concerned vehicles (sendViaCellular(...)). The application can decide on the used interface to account for the relevance area of messages: If a message is relevant in a large area, long-range data transmission via the cellular interface is preferred, while a locally relevant message is only shared via Wifi. To separate the data received via cellular and Wifi, the vehicle can access the topic used via the getLocalTopic() function. The VehicularSender class is commonly very similar for the approaches presented in this thesis, as we focus on the reception of data via the cellular network. That is, as each message generated by a vehicle is disseminated to a number of vehicles, leading to a high load on the cellular network. We prevent this by adapting the reception of messages such that the network is not overloaded. In the next step, the transmission of data could further be optimized using approaches we investigated in previous works [37, 117].

```java
public interface VehicularSender extends HostComponent {
    void initialize();

    boolean sendViaCellular(EnvironmentInformation<? extends
        LocationBasedEnvironmentProperty> pInformation);

    boolean sendLocally(EnvironmentInformation<? extends
        LocationBasedEnvironmentProperty> pInformation);

    Topic getLocalTopic();
}
```

Listing 1: VehicularSender.java

The reception of data is managed by the Listing 2 interface. To manage the vehicular applications, we utilize the observer pattern: Each vehicular application is an observer of the VehicularReceiver, who manages the reception of messages. Thus, each registered application is notified if a new relevant message arrives. The advantage of this architecture is the flexibility regarding the number and type of supported vehicular applications, as each application only need to implement the CommunicationListener interface and register at the VehicularReceiver. Compared to the VehicularSender, which is commonly similar for all the approaches presented in this thesis, the VehicularReceiver is commonly adapted to implement the different communication mechanisms presented in Section 3.2.2. Thus, the application is only provided with the data received by the respective implemented VehicularReceiver, which enables the link between application performance and networking mechanism.

```java
public interface VehicularReceiver extends HostComponent {
    void initialize();

    void registerCommunicationListener(CommunicationListener pListener);

    double getAvailableBandwidth();
}
```

Listing 2: VehicularReceiver.java

6.2 PROTOTYPICAL REALIZATION OF APPROXIMATE VEHICULAR NETWORKS

In this section, we describe the necessary extensions to the Simonstrator.KOM platform to implement approximate vehicular networks. For that purpose, a quality-centric view of data is required, as well as a probabilistic and impact-centric behavior of the underlying Pub/Sub system. In the following, we describe the necessary changes to the data representation and to the implemented Pub/Sub system.

6.2.1 *Reflecting Quality in Measurements*

The main challenges to reflect the quality in measurements are the implementation of sensor inaccuracy and the tracking of inaccuracy through the aggregation of measurements. In its initial version, the Simonstrator supports the possibility to add an error to the measurement of a host, but this possibility does not allow for the adaptation of the provided measurement value. That is, a quality-centric measurement should not only be potentially erroneous but also state the possibility for a wrong value in its meta-information. In order to not only provide the possibility of erroneous data but also track the inaccuracy through aggregation and processing, we extended the data representation in the Simonstrator framework to provide this meta-information by providing the distribution of values as a probability function. Based on this function, two measurements can easily be aggregated as described in Chapter 4. Based on the quality-centric view on data, the networking can be improved, which is described in the following.

6.2.2 *Impact-based Data Dissemination*

For the impact-based data dissemination, the Bypass.KOM framework [149] needed to be adapted to support the non-atomic announcement of interest. This adaptation needs to be performed in multiple components due to the dependency of different components on each other. Thus, adaptations are necessary to the subscriptions to announce interest, the handling of messages on the server-side, and the Pub/Sub client. Notice that there are no adaptations to the notifications necessary, as the specific-impact of a message solely depends on the receiving vehicle and the message itself. As the specific-impact of message on the receiving vehicle is unknown, the server estimates this specific-impact. Thus, the performance of the network increases drastically if this estimation is accurate, and decreases otherwise.

6.2.2.1 *Subscriptions*

In the Pub/Sub paradigm, vehicles need to announce their interest to the broker to receive interesting messages. In the case of our impact-based networks, each vehicle communicates one or multiple base-impact ranges out of which it wants to receive messages. This mechanism can be used to adapt the load of the vehicle to network conditions and personal preferences, such that the vehicle receives exactly as much data as desired. For this purpose, as impact-based subscription consists of a set of base-impact ranges, in which the vehicle considers the messages to be relevant. A threshold-based approach is not applicable in this case, as vehicles might want to cooperate and coordinate their transmission. In that case, threshold-based approaches would greatly reduce the efficiency of this cooperation and decrease the performance of the network.

6.2.2.2 *Client Behavior*

At the client, the vehicular applications need to be able to create impact-based subscriptions to announce their interest in certain message types. For this purpose, we extended the existing implementation of the Bypass Pub/Sub client to support the creation of impact-based subscriptions. These changes allow for the creation of subscriptions based on the base-impact of messages, which are suitable for our impact-based vehicular applications.

6.2.2.3 *Handling of Messages at the Server*

In Bypass, the handling of messages at the broker is managed by the Subscription-Storage class, which matches of notifications to subscriptions, partially considering the context of the message and receivers. However, the current context-aware implementations of the SubscriptionStorage only support a limited set of context-aware dissemination mechanisms, which are area-based and, thus, not usable for our implementation of impact-based data dissemination. Thus, we design a new extension of the SubscriptionStorage, which considers the specific-impact of a message and matches them with the base-impact based subscriptions described previously. Before deciding on the transmission of a message, the SubscriptionStorage calculates the base-impact of the message and the specific-impact for each individual vehicle. While this process might become computationally expensive for large-scale networks, brokers with only regional responsibility and approximations for distant vehicles can be used to increase scalability. For most vehicles, there will not be an exact match between the specific-impact assigned with the message for a certain vehicle and the base-impact.

As described in Chapter 5, the broker filters messages previously to the dissemination process. The broker then matches the base-impact of the message with the base-impact ranges in which the vehicle stated interest, and transmits the message to the vehicle in case of a match. With that behavior, our designed SubscriptionStorage is capable of filtering messages based on their specific-impact and disseminate these messages to the concerned vehicles.

7 EVALUATION

Based on our VEHICLE.KOM platform, we perform an extensive evaluation of our data assessment and dissemination approaches presented in this thesis. We highlight the key properties of our developed approaches, especially regarding the robustness to errors in the network and in measurements, and the efficiency regarding bandwidth usage. In Section 7.1, we discuss our evaluation setup and the common simulation parameters of all simulation scenarios. Next, we analyze the performance of our contributions, starting the data assessment in Section 7.2 presented in Chapter 4. Based on these findings, we analyze the dissemination of data via the cellular network using our impact-aware dissemination approach in Section 7.3, which is described in both Chapter 4 and Chapter 5, but only relies on non-cooperative networks. After that, we analyze the performance of our approach to cooperation in approximate vehicular networks described in Chapter 5 and show the high performance even when compared to approaches with close-to-optimal bandwidth usage.

The goal of this evaluation is to show the adaptability of approximate vehicular networks to the data that is carried over the network, especially the prioritization of high-impact data. Additionally, we show that the explicit handling of uncertainty in these networks can greatly improve the performance of the decision made based on the data. Thus, the consideration of uncertainty and approximation is an essential aspect of future efficient vehicular networks, even though the performance of this network is stochastic. We also show the convergence towards deterministic networks for high-impact messages, which might be desired in real-world vehicular networks.

7.1 EVALUATION SETUP AND METHODOLOGY

In this section, we describe the common parameters and assumptions for the evaluation of our proposed mechanisms for approximate vehicular networks. For this purpose, we first describe the assumed communication models in Section 7.1.1, followed by a description of the underlying road networks in Section 7.1.2. Finally, we introduce the plot types used in this thesis in Section 7.1.4.

7.1.1 *Communication Models*

We use the communication models available at the Simonstrator.KOM platform, which are originally tailored for the simulation of P2P networks. For the cellular network, we rely on the implemented model for 4G communication as presented in [152], which models the latency, available bandwidth, and message drop. While this model is comparably simple, the influence of message drop on the cellular channel is not focus of this thesis, as it can be compensated for by reliable transfer protocols like the Trans-

Parameter	Value
Wifi model	802.11
Loss exponent	4
Error model	NistErrorRateModel
Cellular bandwidth (down)	$100^{\mathrm{Mbits}}/_{\mathrm{s}}$
Cellular bandwidth (up)	$10^{\mathrm{Mbits}}/_{\mathrm{s}}$
Cellular latency	$30\mathrm{ms}$

Table 1: Simulation parameters for the evaluation of data assessment.

mission Control Protocol (TCP). Additionally, we do not investigate the influence of the lack of communication infrastructure in this thesis, i. e., this simplified implementation of the cellular network is sufficient in our case. For local Vehicle to Vehicle (V2V) communication, we use a NS-3 model for the 802.11 model with distributed coordination [154], which has been included in the Simonstrator.KOM platform [149, 150]. This model simulates the noise floor caused by concurrent transmissions to determine if a sent packet is transmitted to the receiver, which is important to investigate the influence of message transmissions on the Wifi channel. While this model differs from the Decentralized Congestion Control (DCC) utilized in 802.11p, it is sufficient to accurately model the communication in our evaluation.

Table 1 provides an overview of the assumed network parameters. For the Wifi model available in the Simonstrator.KOM platform, we use a loss exponent of 4, which is considered to be suitable for urban areas [7]. For the cellular network, we consider an LTE-based network, with a maximum downlink of $100^{\mathrm{Mbits}}/_{\mathrm{s}}$, a maximum uplink of $10^{\mathrm{Mbits}}/_{\mathrm{s}}$, and a latency of $30\mathrm{ms}$.

7.1.2 *Simulation Scenarios*

In this section, we describe the properties of the two road networks we utilize to evaluate the performance of our contributions. We use different scenarios to gain clearer insights into the influence of certain parameters on the system performance: We utilize a highway scenario without detours to evaluate our contributions towards data quality assessment and aggregation. For our relevance assessment and message distribution in large-scale approximate vehicular networks, this scenario is not sufficient. Thus, we use the TAPAS Cologne scenario [81] to evaluate the performance of these approaches.

7.1.2.1 *Highway Scenario*

We utilize the highway scenario as a controlled environment, in which we can better analyze the influence of parameters like traffic density and vehicle speed. It consists of two highways, which are connected through four highway exits as shown in Figure 10. On the right side of the scenario, an event is generated at some point in the simulation, which is then measured by the vehicles in proximity and shared with the other vehicles

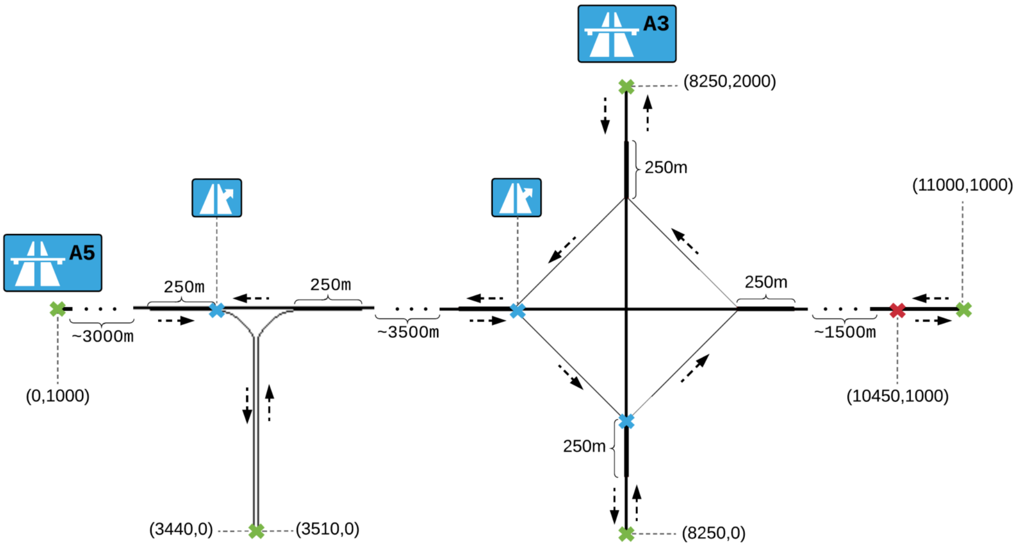


Figure 10: Highway scenario. Taken from [158].

in the scenario. With this scenario, the data assessment can be evaluated and analyzed depending on the number of available measurements.

7.1.2.2 *TAPAS Cologne Scenario*

The TAPAS Cologne Scenario is, with a simulated area of roughly $30\text{km} \times 30\text{km}$, one of the largest available scenarios for the vehicular traffic simulator Simulation of Urban Mobility (SUMO), which simulates a whole day in the city of Cologne, Germany. The mobility pattern of the population is derived from the TAPAS dataset [81], which contains information about the traveling habits of Germans. Figure 11 displays the road network of the TAPAS cologne scenario including the load on the streets. A red road is most occupied, while a blue road is barely occupied. The black roads are only part of the underlying road network but are not utilized by any vehicle. In Figure 11a, it is evident that the main road occupation is the city center of Cologne. Thus, simulating the whole scenario is generally not necessary, and we focused on network simulation to an area in the city center, which is marked with a green box in the figure and has a size of $2\text{km} \times 2\text{km}$. Figure 11b is a zoomed-in version of Figure 11a. We can observe that this area is indeed quite trafficked and contains streets of different types, like highways, rural roads, and urban roads. Thus, it is well suited to simulate the data dissemination in vehicular networks and to analyze the necessity of data at certain locations. In this scenario, a certain number of events of different types is generated randomly after the start of the simulation. These events are measured by vehicles in proximity and shared with a central server, who then forwards the data to concerned vehicles. With this scenario, the dissemination efficiency of both our impact-aware server-side approach and our game-theoretic client-side approach can be evaluated.

(a) Full scenario. (b) Network simulated.

Figure 11: TAPAS Cologne scenario with road occupation. Red roads are highly trafficked, while blue roads are barely trafficked.

7.1.3 *Threads to Validity*

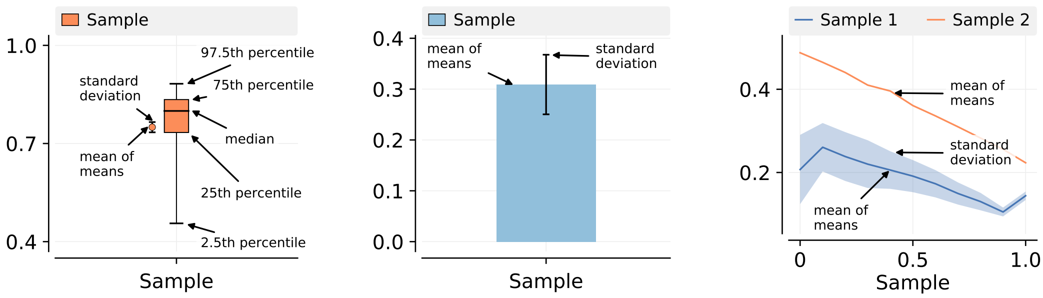
The evaluation of our work relies on several assumptions, which we discuss in the following. Additionally, we assess the influence of these assumptions on the validity of our evaluation. These assumptions can be categorized into assumptions related to the behavior of the vehicles, the behavior of the wireless communication channel, and the behavior of the environment. These factors are discussed in the following.

7.1.3.1 *Behavior of the Vehicles*

In this evaluation, we assumed that the movement of the vehicles is appropriately modeled by the traffic simulation SUMO. While this assumption is commonly used for the evaluation of vehicular networks in general, a different movement behavior of the vehicles might lead to a different behavior of the vehicular network. In this case, the number of topology changes might change, which may change the performance of the cluster-based approaches presented in Section 7.4. With a decreasing number of topology changes, the performance of the cluster-based approaches would increase, potentially outperforming our developed approach from Chapter 5. However, our developed approach would still provide highly reliable transmission of high-impact data, which is especially valuable if packets may be lost when transmitted via the local Wifi channel.

7.1.3.2 *Behavior of the Wireless Communication Channel*

The wireless communication channel has a strong influence on the performance of the developed approaches for Approximate Vehicular Networks. While a real Wifi



(a) Box Plot.

(b) Bar Plot.

(c) Line Plot.

Figure 12: Examples of the plots used in this evaluation.

and LTE channel suffer from effects like shadowing, mirroring, and interference, the utilized Wifi communication model only considers interference to keep the computational complexity appropriate. Thus, the communication channels will likely behave differently in real-world communication networks. However, when we consider the properties of our Approximate Vehicular Network, these effects are partially compensated, considering the induced high redundancy of high-impact messages. In contrast, the cluster-based approaches used for comparison are expected to decrease further in performance, as unexpected and unpredictable disconnects from the cluster-head will occur more frequently.

7.1.3.3 *Behavior of the Environment*

In our evaluation, we assumed no background traffic, which might be generated by other vehicular applications, smartphones, and other wireless communication devices. As a result, the wireless communication channel is more congested than assumed in our evaluation, which decreases the bandwidth available for our Approximate Vehicular Network evaluated in Section 7.4. To this end, the available bandwidth at the wireless communication channel might be insufficient to allow for the redundant broadcasts of high-impact messages. Additionally, the frequency of interference-caused message loss might increase, reducing the performance of our developed approaches. Thus, approaches to reduce the utilized bandwidth on the wireless communication channel might be necessary for the applicability in a real-world environment, which might reduce the performance of our developed approaches. However, the prioritization of high-impact messages in conjunction with the increased redundancy of these messages is expected to compensate for message loss on the wireless communication channel.

7.1.4 *Plot Types*

In this thesis, we rely on different types of plots, which we introduce in the following.

7.1.4.1 *Box Plot*

We utilize box plots whenever we present the aggregated results of one simulation run. In a box plot, the median and the respective percentiles refer to the performance of the active hosts in the simulation, i. e., the median represents the performance of the host, which performed better than 50% of all other hosts in the simulation. Thus, the boxes provide a good insight into the performance distribution of the hosts in the network and describes the consistency between multiple hosts. To allow statements over multiple simulation runs, we provide the mean of means as a point next to the box, which also displays the standard deviation of the mean values of all simulation runs. This provides good insights into the average performance of the hosts, and the stability of the approaches over multiple simulation runs.

7.1.4.2 *Bar Plot*

In contrast to the box plot, the bar plot is utilized to display the average performance over multiple simulation runs. For this purpose, the height of the bar refers to the mean over all simulation runs. When bar plots are used, it is often not feasible or desirable to display the performance of individual hosts, e. g., if the time until an event happens is measured. In addition, we provide the standard deviation over these simulation runs, to show the variation between different runs.

7.1.4.3 *Line Plot*

We utilize line plots to visualize a development for the change of one parameter. In these line plots, the change of the metric is often easier to observe compared to a box plot or bar plot. In the line plot, the mean performance of the approach over the simulation runs is displayed. For the main approach, we additionally display the standard deviation of the mean value over multiple simulation runs to assess the stability of our approach over multiple simulation runs.

7.2 AGGREGATION SCHEME FOR INACCURATE MEASUREMENTS

In this section, we evaluate the performance of our data accuracy assessment and data aggregation, with the goal of reducing wrong aggregates in our vehicular network. For this purpose, we compare the performance of our approach with state-of-the-art approaches. In this evaluation, we want to investigate the following two hypotheses.

1. Our approach significantly reduces the number of false decisions compared to state-of-the-art methods.
2. Our approach adapts its robustness and adaptability to the requirements given through the sensor accuracy of the measuring vehicles, the lifetime of the road event, and the environment.

First, we introduce the scenario model in Section 7.2.1 including the reference approaches, the metrics used, and the parameters. Following, we investigate the influence of different parameters: In Section 7.2.2, we analyze the influence of the sensor accuracy to the performance of the approaches. Next, we analyze the event lifetime as second information-specific parameter in Section 7.2.3. As the last influence factor, we analyze the influence of vehicular traffic in Section 7.2.4.

7.2.1 *Scenario Model*

In this section, we describe the specific properties of this evaluation and the necessary changes compared to the default setup. For the data assessment, we rely on the highway-scenario to limit the noise induced by varying traffic conditions. This scenario has a warmup period of $30\mathrm{min}$.

$2.5\mathrm{min}$ after the start of the simulation, an event with a specified duration is started, which has 15 possible states. While the event is active, the value of the underlying variable changes 10 times between these states. The vehicles at the event location generate messages using their local sensors, which have a predefined accuracy. This accuracy reflects the quality of the generated measurements and is added as meta-information to the shared message using our vector-based representation of a measurement as described in Section 4.2.1. Additionally, each message has a Time to Live (TTL) of 10% of the event lifetime to invalidate messages, as they are likely to not describe the correct value. The generated messages are then broadcasted to all vehicles in the network using the cellular network, which is assumed to have sufficient resources to deliver all available messages to all vehicles. When a vehicle receives a message, it stores it locally in its cache and creates a new aggregate based on the previously stored measurements. When a vehicle traverses the event location, the current state of the aggregate is investigated. The aggregate is considered correct if the state with the highest probability is equal to the current state of the underlying variable.

In the following, we first introduce the scenario-specific metrics and parameters and then describe the considered state-of-the-art approaches for data assessment.

7.2.1.1 *Metrics*

In this section, we utilize metrics associated with the number of correct decisions, which are the share of correct aggregates, the adaptation time, and the number of aggregate changes.

Share of Correct Aggregates: The share of correct aggregates reflects the overall performance of the approaches. A high share of correct aggregates is desirable, as the share of vehicles that need to work with false data is reduced. The share of correct aggregates is measured when the vehicles are at the event location by analyzing the current state of their aggregate. If the state with the highest probability equals the correct state, then the aggregate is considered correct, otherwise it is considered false.

Adaptation Time: The adaptation time measures the time until an approach adapts its aggregate after a state change of the underlying variable. Thus, this metric captures the

Parameter	Value
Sensor accuracy	100%, 90%, **80%**, 60%
Sensor heterogeneity	0%, 10%, **20%**
Event lifetime [min]	10, **20**, 40
Traffic density	0.05, 0.1, **0.3**, 0.9
Traffic speed [km/h]	**mixed**, 80, 100, 130, 180

Table 2: Simulation parameters for the evaluation of data assessment.

time between a state change of the underlying variable and the *first* correctly calculated aggregate. It is a metric to determine the adaptability of the approaches.

Number of Aggregate Changes: The number of aggregate changes measures the number of changes of the aggregate. This metric is measured when the aggregate has been correct, i. e., the adaptation to the new state of the underlying variable has been performed. We then track all changes of the aggregate value, which are undesired as this captures false aggregates and is caused by incorrect adaptations to false measurements. It is a metric to determine the robustness of the approaches.

7.2.1.2 *Parameters*

Besides the parameters discussed in Section 7.1, we require additional parameters to analyze the performance of our developed approaches. These parameters are centered around (i) the accuracy of the measured data values, i. e., the average sensor accuracy and the variation of the sensor accuracy, and (ii) the lifetime of measurements. The average sensor accuracy reflects the possibility of vehicles to measure accurate data, which influences the required robustness of the aggregation approach. We measure the sensor accuracy in the percentage of correct data entries delivered by the sensor. To reduce the impact of a single vehicle, we assume that every vehicle may measure every event exactly once. The variation of the sensor accuracy influences the requirement of the approaches to consider the accuracy of each individual measurement. The data lifetime, influences the required adaptability of the aggregation approaches.

In addition to data-related parameters, we also analyze the performance of our approaches in varying traffic environments. For this purpose, we vary the traffic density and the average vehicle speed. The traffic density is set using the spawn rate parameter in SUMO, which indicates the probability that a vehicle is spawned in a certain time interval. The vehicle speed is generally mixed, i. e., 10% trailers (80km/h), 30% trucks (100km/h), 50% normal cars (130km/h), and 10% sport cars (180km/h), but we also analyze the performance for the different vehicle types separately.

A summary of the parameters and their values is provided in Table 2. In the following, we analyze the influence of every parameter separately.

7.2.1.3 *Reference Approaches*

We assess the performance of our developed approaches using the reference approaches described in this section. These reference approaches are based on state-of-the-art methods for data aggregation.

Newest-Only (NEW): The *newest-only* approach always relies on the newest measurement that is received to determine the real value. This approach adapts quickly to environmental changes but lacks robustness to false measurements. Thus, this approach is heavily impacted by false measurements, which decrease its performance if the measurement accuracy is low.

Majority-based (MAJ): The *majority-based* approach always selects the measurement value that is most commonly represented in its storage. Notice that we only consider discrete variables, as continuous variables can be approximated using discrete variables. This approach is very robust to false measurements, as a high number of false measurements is required to change the result of the *majority-based* approach. However, this approach has the downside of adapting slowly to environmental changes due to the same reason. Thus, this approach is expected to perform well for static variables (without frequent changes) or if the average measurement accuracy is low.

Oracle (Oracle): The oracle approach is used as an upper baseline for our analysis. It selects the correct value out of the storage if it exists, and provides the closest value if not. However, it does not aggregate measurements, i. e., it is not considered to be an optimal aggregation strategy. The difference is only noticeable for low-quality measurements, in which the performance of the other approaches drops likewise.

Accuracy-Aware (CP): Our *accuracy-aware* approach has a similar idea as the *majority-based* approach and has been proposed by us in [118]. Similar as the *majority-based* approach, this approach generally chooses the measurement value that is most represented but also considers the accuracy of the measurements in this process. The advantage of this approach is most significant in heterogeneous environments, in which the accuracy of provided measurement can vary significantly. However, similar to the *majority-based* approach, this approach requires a high number of measurements to adapt to new values if the sensor accuracy is low and, thus, takes a long time to adapt to environmental changes.

Data-Centric (QoI): Our data-centric approach considers the average accuracy and lifetime of data to adapt the aggregation such that the number of wrong decisions is minimized and has been proposed by us in [121]. Due to the consideration of average accuracy and lifetime, the number of false aggregates can already be reduced significantly. However, this approach does not consider the potentially varying accuracy of the available measurements, which reduce the performance in highly heterogeneous environments.

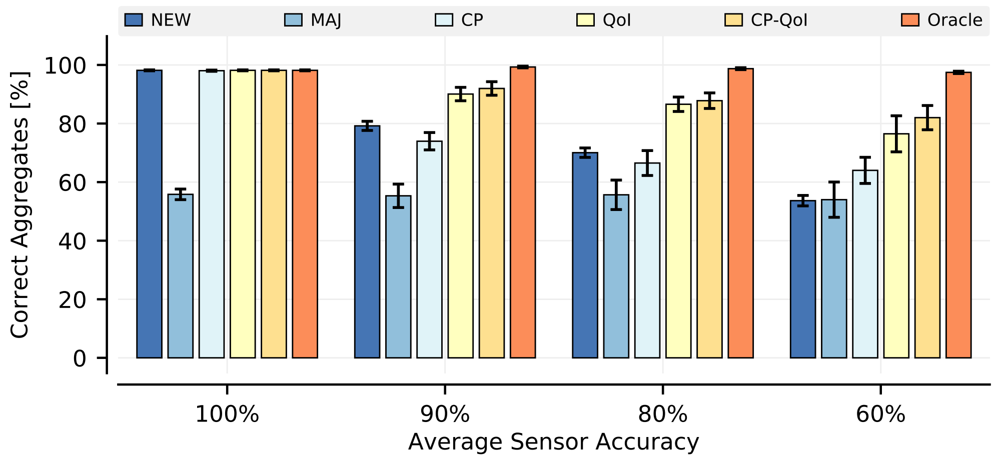


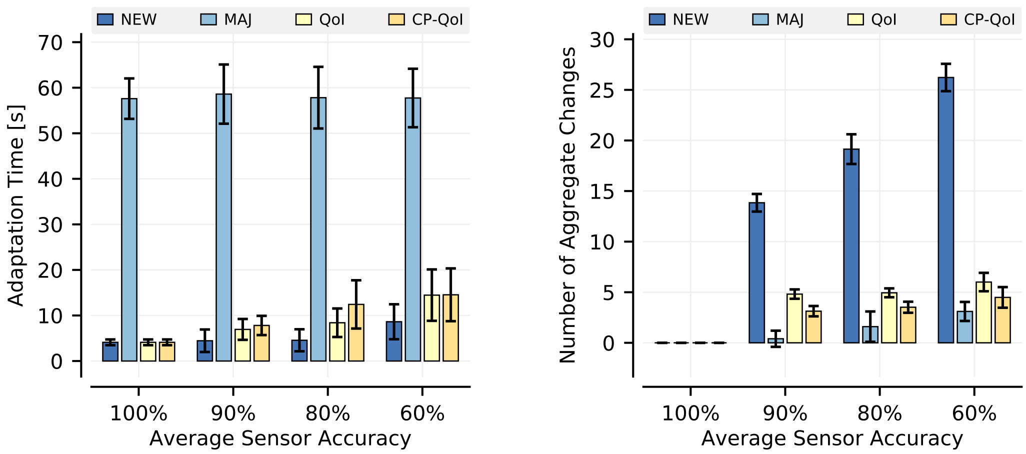
Figure 13: Influence of the average sensor accuracy on the quality of the aggregates for the different approaches.

Accuracy-Aware Data-Centric (CP-QoI): Our accuracy-aware data-centric approach combines the accuracy-aware and the QoI approach as described in Section 4.2. Thus, this approach combines the advantages of both approaches, which is assumed to further reduce false aggregates.

7.2.2 *Influence of Data Accuracy*

In this section, we analyze the influence of data accuracy on the number of correct aggregates, the robustness, and the adaptability of each approach. Notice that the total number of aggregates is similar in the simulation of each approach, as the underlying movement models are not affected by the aggregates. Thus, the share of correct aggregates corresponds to the total number of correct aggregates.

Figure 13 displays the performance of the different approaches depending on the sensor accuracy. The figure visualizes the share of correct aggregates of the overall network, as every vehicle only performs one aggregation in our scenario. For a sensor accuracy of 100%, all approaches except the *MAJ* approach perform well, as they can always rely on the newest measurement to reflect the real state of the road. Thus, an adaptation is performed immediately after an environmental change, and no false measurements need to be handled by the approaches. The *MAJ* approach performs poorly as its adaptation speed is very low, i. e., the aggregates produced by the *MAJ* approach remain in the old state for a long time. With decreasing sensor accuracy, the performance of all approaches decreases, but the level of decrease differs strongly between the different approaches. The *NEW* approach loses performance linearly with decreasing accuracy, as every wrong measurement in the system induces a false aggregate for the next vehicle passing the event. The performance of the *MAJ* approach is only marginally affected by the decreasing data quality, as it is very robust to



(a) Adaptation Time.

(b) Number of Aggregate Changes.

Figure 14: Influence of the average sensor accuracy on the adaptability and robustness of the different approaches.

false measurements. However, as the *MAJ* approach reaches a stable state very late, the overall performance still remains low. In contrast, the performance of the *CP* approach decreases, as its adaptability decreases drastically at low sensor accuracies. Our developed approaches, the *QoI* approach and the *CP-QoI* approach, also reduce in performance, but their decrease is much less severe compared to the *NEW* approach and the *CP* approach. That is, the *QoI* approach and the *CP-QoI* approach find an appropriate balance between robustness and adaptability to preserve the quality of the aggregates. Additionally, the CP-QoI approach outperforms the *QoI* approach for low sensor accuracies, as it is less prone to the heterogeneity of vehicular sensors and utilizes knowledge of the state transitions of the underlying variable to predict the future state of the event. The *Oracle* approach almost always achieves a share of correct aggregates of 100%, which is justified by the long time that the vehicles travel until they reach the event location, i. e., the correct value is almost always available in the cache. Based on these observations, we can confirm hypothesis 1 for varying sensor accuracy, i. e., our approach drastically outperforms state-of-the-art methods regarding the number of false aggregates.

In Figure 14, we further analyze the performance of the approaches regarding their adaptability to environmental changes and their robustness to false measurements. In this plot, we omit the results for the *CP* approach and the *Oracle* to improve readability. Figure 14a depicts the adaptation time to environmental changes of the different approaches, which gives insights into the adaptability of the approaches. We can clearly see that the adaptation time of the *MAJ* approach is very significant and almost independent of the sensor accuracy. That is, the *MAJ* approach fails to quickly adapt to the new state, which leads to all aggregates being false until the adaptation is performed. This aspect also explains the poor performance of the *MAJ* approach regarding its

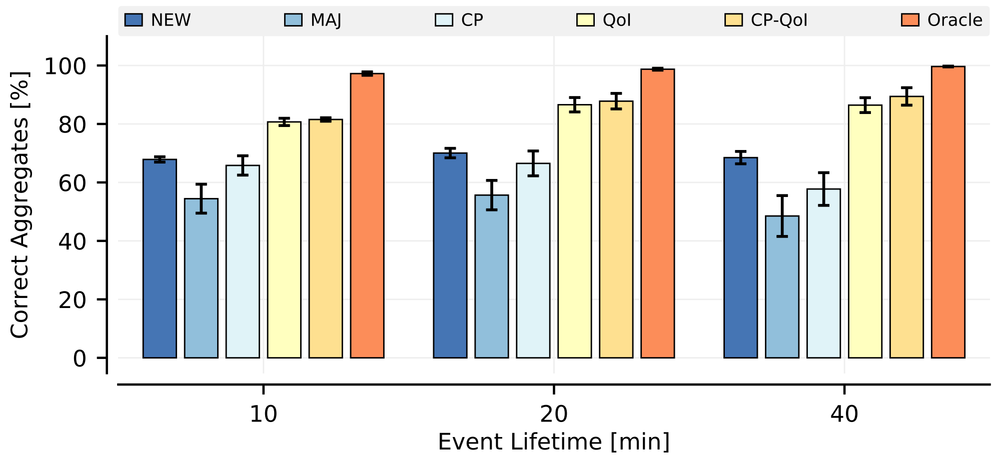
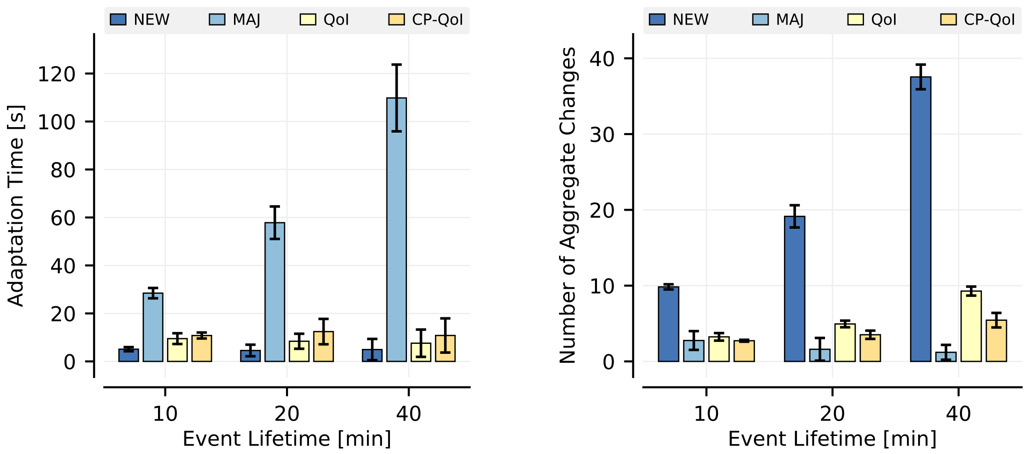


Figure 15: Influence of the event lifetime on the quality of the aggregates for the different approaches.

overall performance, even though it is considered very robust to false measurements. For the other approaches, the adaptation time increases with decreasing sensor accuracy. We can use the *NEW* approach as a reference to determine the influence of the measurement accuracy on the adaptation time. For a decreasing measurement accuracy, the adaptation time of our *QoI* approach and our *CP-QoI* approach are much higher, which is justified by the required robustness of the aggregation. As our approaches adapt to the measurement accuracy, they increase their adaptation time to gain robustness to compensate for wrong measurements. This can also be observed in Figure 14b, in which we can clearly observe that the *NEW* adapts frequently to false measurements, i. e., produces false aggregates. Compared to this, the other approaches have a much lower number of adaptations, which is due to their higher robustness to false measurements. As assumed, the *MAJ* approach is the most robust approach, as our *QoI* approach and CP-QoI approach reduce their robustness to allow for fast adaptation to environmental changes. In conclusion, Figure 14 displays the adjustment of the robustness and adaptability of our CP-QoI approach based on the sensor accuracy and, thus, confirms hypothesis 2 regarding the sensor accuracy. While the level of sensor heterogeneity has an influence on the approaches, all approaches handle the heterogeneity of sensors well due to the high number of available measurements. The corresponding figures are presented in the appendix in Section A.3.

7.2.3 *Influence of Event Lifetime*

In this section, we analyze the influence of the event lifetime on the quality of the aggregates. A long event lifetime generally benefits approaches with high robustness, as the influence of environmental changes is much lower. In contrast, a short event



(a) Adaptation Time.

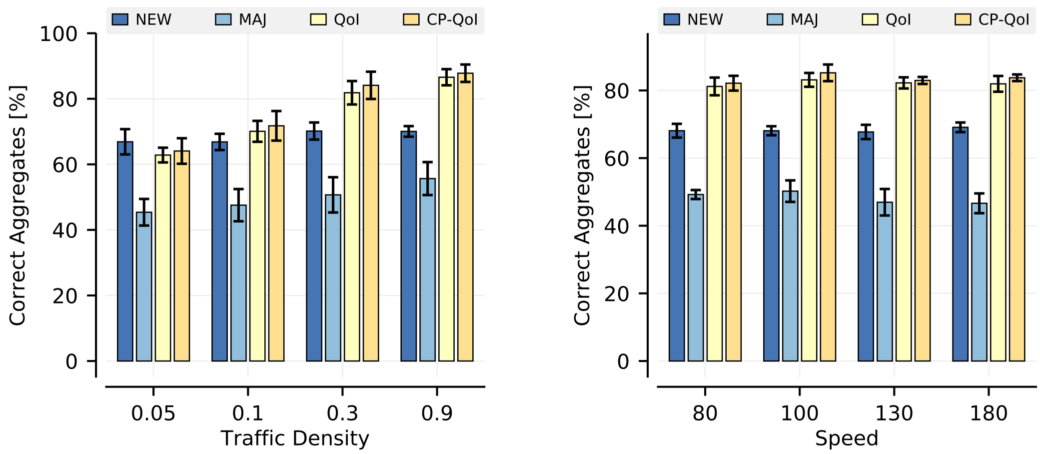
(b) Number of Aggregate Changes.

Figure 16: Influence of the event lifetime on the adaptability and robustness of the different approaches.

lifetime generally benefits approaches with high adaptability, as the influence of the environmental changes increases drastically.

Figure 15 displays the influence of the event lifetime on the share of correct aggregates. An increase in the event duration has a different influence on the different approaches. While the performance of the *NEW* approach remains unchanged as its performance mainly depends directly on the sensor accuracy, the performance of the *MAJ* approach and the *CP* approach decreases, which is unexpected as we anticipated a performance gain for these approaches. However, due to the higher number of outdated messages in the cache after an adaptation, the time until an adaptation is performed increases even further, as more messages are available in the cache. For our *QoI* approach and our *CP-QoI* approach, the performance increases with an increasing event lifetime, as the number of environmental changes is lower if the event lifetime is high, i. e., the number of certainly false aggregates (after the change of the environment) is reduced.

Figure 16 displays the adaptability and robustness depending on the event lifetime. There are two properties that change when the lifetime of an event increases: (i) the frequency of environmental changes reduces and (ii) the number of measurements of the old state increases, i. e., the difficulty of an early adaptation also increases. The time until an adaptation is performed is shown in Figure 16a. When analyzing the adaptation speed of the *MAJ* approach, we can clearly observe that the time until adaptation increases with increasing lifetime, which is justified by the aforementioned high number of measurements of the old state. Our *QoI* and CP-QoI approach remain comparably constant in terms of adaptation speed, which showcases the adaptation to the event lifetime. Thus, the influence of the high number of measurements in the old state is seemingly not required to keep the robustness of the approach high.



(a) Traffic Density.

(b) Traffic Speed.

Figure 17: Influence of the traffic conditions on the quality of the aggregates for the different approaches.

Thus, our approaches decrease the impact of old measurements to allow for a fast adaptation. Figure 16b displays the number of aggregate changes during this interval. For all approaches except for the *MAJ* approach, we can observe that the number of aggregate changes increases with increasing event lifetime, which is justified by the longer durations in which the event remains constant and robustness is required. For the *MAJ* approach, this number actually decreases, as more messages are available in the cache, which further increases the performance of the approach. In conclusion, we can observe that our CP-QoI approach has a constant adaptation time for different event lifetimes, which generally increases due to the higher number of messages in the cache. Thus, we can confirm hypothesis 2, that our CP-QoI approach adapts its robustness and adaptability to the event lifetime.

7.2.4 *Influence of Traffic Conditions*

In this section, we vary the average speed and density of vehicular traffic. This influences the number of measurements available to the vehicles: an increasing traffic density increases the number of available measurements, while an increasing vehicle speed reduces the time until the vehicle needs to determine the aggregate.

Figure 17 displays the quality of the aggregates depending on the traffic conditions. Figure 17a displays the influence of the traffic density on the quality of the aggregates for the different approaches. If the traffic density is very low, the number of available measurements is also very low. Thus, the advantage of the approaches relying on the aggregation of multiple measurements decreases, while the performance of the *NEW* approach is unaffected. This leads to the *NEW* approach being comparable or even slightly better than our developed approaches. However, with increasing traffic density

and, thus, an increasing number of available measurements, the performance of our *QoI* and *CP-QoI* approach increases drastically, while the *NEW* approach remains constant. For high traffic densities, the advantage of our approaches over the *MAJ* and *NEW* approach is most significant. While the *NEW* approach is unaffected by the traffic density, the *MAJ* approach increases in performance if the traffic density increases, as the adaptation time is similar, but its robustness to false measurement increases due to the higher number of messages in the cache. Compared to that, the speed of the vehicles seems to have only a minor influence as shown in Figure 17b. All approaches are only marginally affected by the changes in the vehicle speed. Thus, the speed of vehicles seems to have only a minor influence on the number of available measurements, as otherwise the performance of our approaches would change.
In conclusion, we can state that our approach performs better if the number of available measurements is high, as it can ignore existing measurements if required to increase the quality of the aggregates. This is a major improvement over the *MAJ* approach, which seemingly decreases in performance if the number of available measurements increases. Thus, we can again confirm hypothesis 2, that our *CP-QoI* approach adapts its robustness and adaptability to the environment.

7.3 PRIORITIZATION BASED ON THE SPECIFIC-IMPACT

In this section, we analyze the influence of relevance to the dissemination of data in a cellular-based vehicular network. For this purpose, we compare the performance of our impact-aware dissemination with state-of-the-art methods for the distribution of context-sensitive messages. In this part of the evaluation, we want to evaluate the following hypotheses:

1. Our impact-aware dissemination increases the efficiency of message distribution in vehicular networks compared to state-of-the-art approaches.
2. Our impact-aware dissemination adapts to the data impact, i. e., high-impact data is prioritized over low-impact data.

7.3.1 *Scenario Model*

In this section, we describe the specific properties of this evaluation and the necessary changes towards the default setup. For the analysis of data relevance and its impact on the message dissemination, we utilize the TAPAS Cologne scenario to simulate realistic traffic flow conditions. This is important to assess the possible improvements of our impact-aware communication, as this type of communication relies on prediction, i. e., predictable traffic flow would increase the performance of our impact-aware approach unnaturally. In order to pre-train the prediction model, we let the simulation running for 30min without interfering with it, i. e., not simulating any communication. During this time, we observe the movement of traffic flow and build our prediction model described in Section 4.3.

In this scenario, the data is not generated by the vehicles themselves, but the data is assumed to be already available at the server, as the influence of our approximate vehicular network can be analyzed much better if the randomness in the message generation is reduced. In simulations where the generation of data by the vehicles should be considered, an active monitoring strategy of incoming data at the vehicles as described in [122] can be utilized. However, these strategies cannot adapt proactively and, thus, need to predict the future bandwidth consumption, which introduces additional errors to the system. Based on the available bandwidth and the anticipated future bandwidth consumption, each vehicle may update its reception strategy considering the current state of the environment, i. e., announce what share of the available messages it can receive given its bandwidth constraints. This reception strategy is communicated to the server and utilized to decide on the transmission of data.

In the following, we first introduce the scenario-specific metrics and parameters and then describe the considered state-of-the-art approaches for cellular data dissemination in vehicular networks.

7.3.1.1 *Metrics*

In this section, we utilize metrics associated with the efficiency and the behavior of data distribution to assess the performance of the reference approaches. We use the metrics recall and precision to assess the performance of the message dissemination. In the following, we describe our definition of the aforementioned metrics.

Recall: Recall is the ratio between the number of messages that have been received and are relevant, divided by the total number of relevant messages. Thus, recall is a metric describing the quality of the communication regarding the availability of data. We consider a message to be relevant if the corresponding event is located on the vehicle's planned path and the vehicle will arrive at the event location while the event is active, which are determined using simulation knowledge.

Precision: Precision is the ratio between the number of messages that have been received and are relevant, divided by the total number of received messages. Thus, precision captures the unnecessary usage of bandwidth in a system, that is to be reduced. The definition of relevant messages is similar to the definition of relevant messages for the recall metric.

7.3.1.2 *Parameters*

To assess the influence of our impact-aware dissemination approach, we consider different parameter settings listed in Table 3. The first parameter, the event lifetime, influences the possibility of the network to propagate data to distant vehicles. That is, if the expected event lifetime is short, a dissemination of messages to a distant vehicle is not reasonable, as the event will expire by the time the vehicle arrives at the event location. The second parameter, the available bandwidth, restricts the number of messages that may be transmitted to each individual vehicle. The influence of this

Parameter	Value
Event lifetime [min]	5, 10, **30**, 60
Available bandwidth [msg./s]	0.1, **1**, 10, 100
Base-impact	1, 10, 100, **1000**

Table 3: Simulation parameters for the evaluation of impact-aware communication.

parameter is very high, as it restricts the possibility to share knowledge between the vehicles and the server. The third parameter, the base-impact, is used to determine the influence of the base-impact of a message to the performance of the approach. The performance of high-impact messages should generally be better than the performance of low-impact messages.

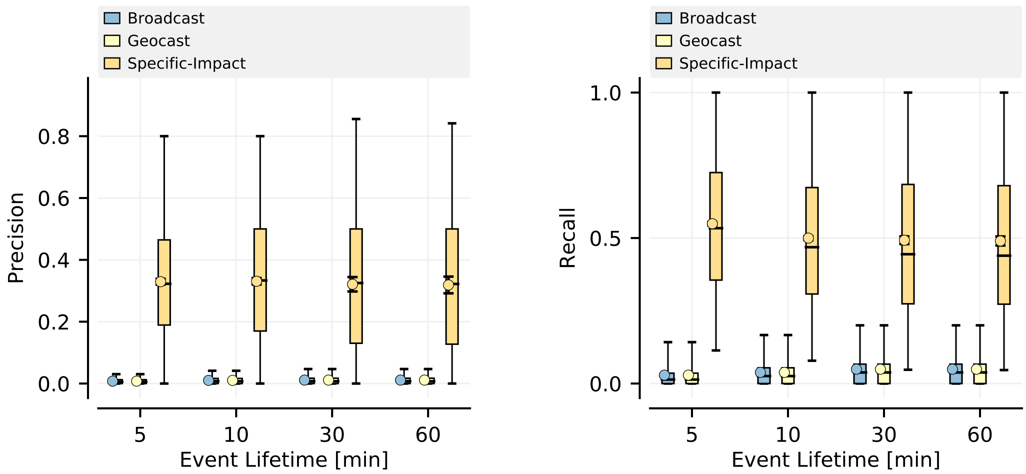
7.3.1.3 *Reference Approaches*

For reference, we use two state-of-the-art approaches for message distribution in vehicular networks, which are described in the following.

Broadcast: The broadcast approach is mainly used to estimate the amount of transferable data in the network and assess the possible savings by adding context-awareness. The broadcast approach transmits every message to every vehicle in the network, without consideration of their individual context. Thus, this approach produces the highest data traffic but also ensures that every vehicle has received every message. We expect that this approach receives every relevant message if the bandwidth is sufficient, but the share of relevant messages is the lowest, as the context of both message and vehicle are not considered in the dissemination.

Geocast: The geocast approach is a conventional approach for message dissemination in vehicular networks. Geocasting disseminates messages inside a predefined area, which can be of any shape. In this evaluation, we assume a circular dissemination area for messages with a radius of 1km. While geocasting is more efficient than broadcasting regarding bandwidth usage, it still assumes that the relevance of messages for a vehicle depends only on the distance to the contained event. The performance of the geocasting approach strongly depends on the area in which messages are distributed: If this area is large, then more vehicles receive relevant data, but the transmission of unnecessary messages also decreases. In contrast, a small area decreases the number of transmitted messages but increases the relevance of a transmitted message to the receiving vehicles.

Impact-Aware: Our impact-aware data dissemination approach has been described in Section 4.4 and considers the specific-impact of a message to increase the efficiency of bandwidth utilization. In the specific-impact, the measurement and the relevance of the measurement for a vehicle are considered. To determine the relevance of a measurement for a vehicle, the server predicts the future movement of the vehicles



(a) Precision.

(b) Recall.

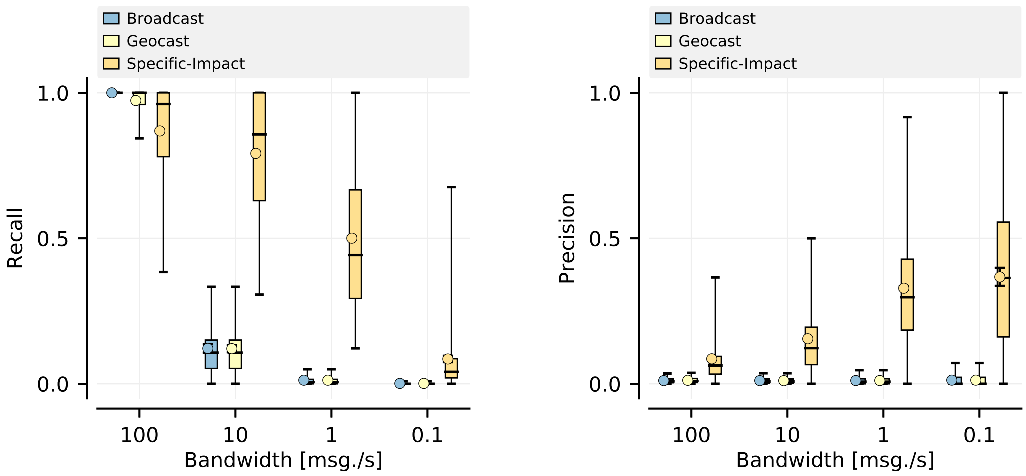
Figure 18: Influence of the event lifetime on the performance of message dissemination for the different approaches.

and the future state of the event. While our impact-aware dissemination approach significantly reduces the utilized bandwidth, its performance depends greatly on the performance of the utilized prediction mechanism, leading to a performance drop if the prediction is poor. In general, we expect this approach to be much more efficient than the geocast and broadcast approach.

7.3.2 *Influence of Event Lifetime*

The degree of consideration of context varies drastically between the reference approaches. While the broadcast dissemination does not consider context during the dissemination process, the geocast approach introduces some level of context-awareness to limit the dissemination area of data. However, even the geocast approach simplifies the assessment of context, although an exact assessment of context is pivotal for the data dissemination. Our impact-aware approach assesses the probability of the vehicle encountering the contained road event and can drastically increase the usability of data for each vehicle.

Figure 18 displays the performance of the impact-aware approach. In Figure 18a, we can see the relevance of the transmitted message for the receiving vehicle, i. e., the precision score. It is evident that our impact-aware dissemination drastically increases the average relevance of received messages compared to the broadcast and geocast approach because it models the relevance of data much more accurately. Thus, it does not only rely on the distance to the event but also considers the topology of the road network and the lifetime of the event. As described in Section 5.5.2, relevance reduces the probability of a message being transmitted. Thus, messages of low relevance are transmitted with a lower probability, i. e., the average relevance of a transmitted message



(a) Recall.

(b) Precision.

Figure 19: Influence of the available bandwidth on the performance of message dissemination for the different approaches.

increases. In contrast, the broadcast approach transmits all messages if the bandwidth is sufficient, which decreases the share of relevant messages compared to the number of overall transmitted messages. The geocast approach considers a message relevant if the vehicle is within a certain distance of the event. However, this does not reflect the actual relevance of a message, which is shown by the very low performance of the geocast approach. As our impact-aware approach explicitly considers the lifetime of the event, it can keep the precision high even if the event lifetime decreases. The share of received relevant messages compared to all relevant messages by an individual vehicle is displayed in Figure 18b. It is evident that the performance of our impact-aware approach is much higher compared to both static approaches in a limited bandwidth setting, as it uses the available bandwidth much more efficiently. Additionally, the performance of our impact-aware approach increases with decreasing event lifetime, which is justified by the more accurate predictability of short paths. That is, for short event lifetimes all paths with long travel times can be directly discarded from the prediction, as the event will most probably be disappeared by the time the vehicle arrives at the event location. For all other approaches, the performance increases with increasing event lifetime, as they utilize static attributes which are independent of the event lifetime. Thus, these approaches benefit from longer paths, which increase the number of vehicles in the network that potentially benefit from that message.

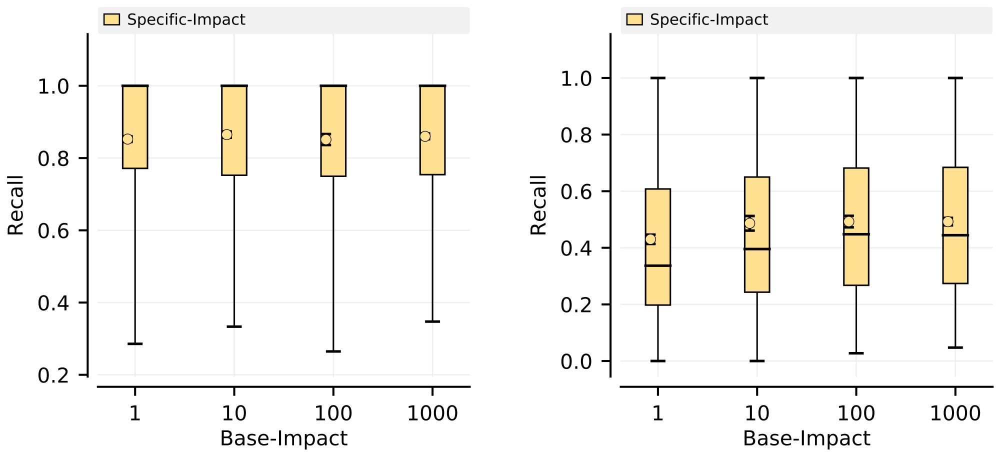
When we analyze the influence of the available bandwidth to the different approaches in Figure 19, the influence of the less efficient bandwidth utilization for the geocast and broadcast approach becomes even more evident. In Figure 19a, the recall of the different dissemination approaches depending on the available communication bandwidth is displayed. The geocast and the broadcast approach both perform well if the available bandwidth is high, but degrade drastically in performance if the bandwidth is limited. In this case, messages are dropped randomly at some point, as the available communication bandwidth is not sufficient to handle the load. In contrast, our impact-aware approach performs slightly worse than these approaches if the bandwidth is high, which is justified by a lack of accuracy in the relevance calculation. To keep the runtime of our impact-aware approach reasonable, we used a lower threshold to cancel the computation if the expected gain in accuracy of the result is low. While this approach is working well in general, it decreases the performance in high bandwidth scenarios as some relevance values of the transmitted messages are below that threshold. Thus, the relevance of a message for certain vehicles is estimated to be 0, which leads to these vehicles never receiving this message. The server can only increase the transmission probability if the relevance is non-zero, as it multiplies the ratio between the available resources and the used resources with the original transmission probability. However, our impact-aware approach is designed for networks with restricted available bandwidth, this behavior should generally not affect the system performance. We can see the improvement of our impact-aware approach when the available bandwidth decreases. The performance of our approach remains much higher than the performance of the geocast and broadcast approach due to the more accurate evaluation of relevance. Since messages of high relevance are prioritized, this leads to a higher share of relevant messages received by the vehicles. This can also be observed in Figure 19b, which shows the share of relevant messages out of all messages received by a vehicle. It is evident that the precision of the geocast and broadcast approach is independent of the bandwidth, while the precision of our impact-aware approach increases due to the prioritization of highly relevant messages. This increase of our impact-aware approach is an important prospect of the efficiency of our proposed approach, as it displays the usability of path prediction for message dissemination in vehicular networks.

These observations let us confirm hypothesis 1, i. e., our approach utilizes bandwidth much more efficiently compared to state-of-the-art approaches. This is mainly visible in the plots showing the precision metric in different environments. In the same time, our impact-aware approach achieves a higher recall compared to state-of-the-art approaches if the bandwidth is limited due to the more efficient bandwidth utilization.

7.3.3 *Influence of Data Impact*

Data impact should generally influence the data dissemination if the available bandwidth is limited. This means that high-impact data should be prioritized if the communication bandwidth is insufficient to transmit all available data to the vehicles. Thus, we analyze the dissemination of high-impact and low-impact messages in networks with different bandwidths per vehicle.

Figure 20 displays the recall of each message type depending on the base-impact of the message. As expected, there is no observable influence of the message impact on the traffic load in an unconstraint network as shown in Figure 20a. Therefore, all messages can be transmitted by the server and the vehicles do not need to limit the reception of messages themselves as described in Chapter 5. Thus, the recall is only limited by the

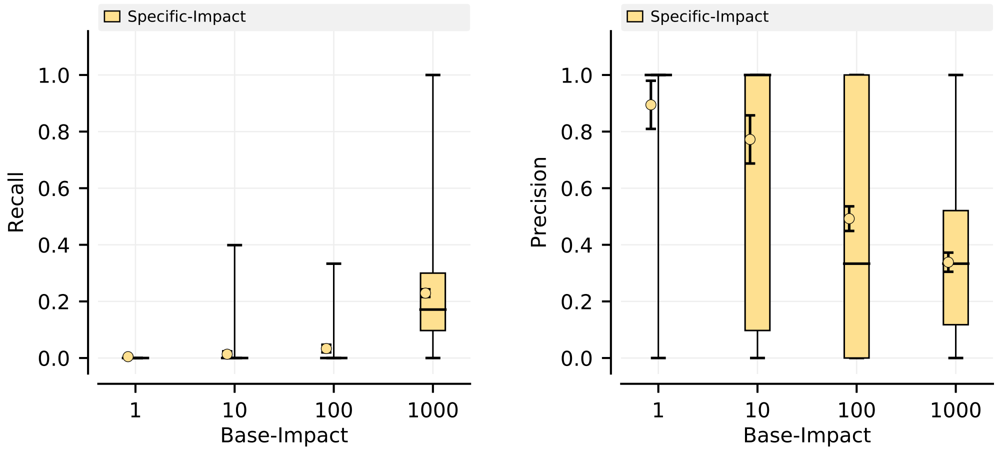


(a) Unconstrained bandwidth. (b) Constrained bandwidth.

Figure 20: Influence of the base-impact of messages on the performance of message dissemination for our impact-aware approach.

imprecision of the relevance assessment as discussed previously. As the impact does not influence the server-side transmission probability, this leads to an equal recall for all impact-levels. However, if we limit the available bandwidth and force the vehicles to limit the reception of low-impact messages as shown in Figure 20b, the recall of low-impact messages drops. Thus, the vehicle prioritizes high-impact messages over the low-impact messages, which enables the server to also transmit high-impact messages with lower relevance values. This also increases the recall, as the number of received relevant messages increases, but the change in the recall is comparably small. This is justified by the fact that a message with a low relevance value is generally less relevant, i. e., the chance of belonging to the set of relevant messages is much lower than for messages with a high relevance value.

The effect of the message impact becomes even more significant if we reduce the available bandwidth further as shown in Figure 21. Through the reduction of available bandwidth, the vehicle-side filtering of data becomes even more significant, leading to a significant reduction in recall for messages with low base-impact as shown in Figure 21a. Thus, our impact-aware approach correctly prioritizes messages with high base-impact as expected, receiving a high share of this data. When analyzing the precision of the message dissemination in Figure 21b, we observe that the precision of messages with high base-impact is much lower than the precision of messages with low base-impact. This is justified by the fact that the vehicles filter messages with low base-impact already on the client-side, leading to a significant reduction of these low-impact messages that are considered for transmission. If messages with a high low base-impact shall be transmitted to a vehicle, both the server and the vehicle need to decide on a transmission, which is much more unlikely compared to high-impact messages, for which only the server needs to decide on a transmission. This behavior



(a) Traffic load.

(b) Precision.

Figure 21: Influence of the base-impact of messages on the performance of message dissemination for our impact-aware approach with very low bandwidth.

is justified by the limited bandwidth of the vehicle, which is insufficient to receive all messages. Based on these findings, we can confirm hypothesis 2, that our impact-aware approach appropriately prioritizes high-impact data, while low-impact data is dropped if the available communication bandwidth is insufficient.

7.4 COOPERATION IN APPROXIMATE VEHICULAR NETWORKS

In this section, we evaluate the performance of our developed approximate vehicular networks considering the contributions regarding impact-aware data dissemination and efficient offloading using our game-theoretic approach as described in Chapter 5. We analyze the performance of our developed game-theoretic approach in comparison with state-of-the-art approaches using a geocast approach for the dissemination of messages via the cellular network. That is, as the server-side performs deterministically for the geocast approach, which enables us to provide a clearer view to the improvements achieved by our game-theoretic approach. For this purpose, we first determine the influence of approximate vehicular networks under varying conditions. At first, we determine the trust factor τ, which is a parameter for the remaining evaluation runs. Next, we analyze the influence of the available bandwidth and of location privacy. Finally, analyze the robustness of our game-theoretic approach to misbehaving vehicles.

In this evaluation, we evaluate the following hypotheses:

1. Our game-theoretic approach manages to transmit high-impact messages much more reliable than realistic clustering approaches and modulates the redundancy

of messages such that bandwidth can be utilized for the reception of previously not received messages.

2. Our game-theoretic approach mitigates the negative effects of location privacy and manages to maintain a constant performance if the share of privacy in the network is low.

3. Our game-theoretic approach is very robust to message loss in the network, and compensates for misbehaving vehicles in the network.

7.4.1 *Scenario Model*

In this section, we describe the specific properties of this evaluation and the necessary changes towards the default setup. For the evaluation of our approximate vehicular networks, we rely on the TAPAS Cologne scenario to guarantee realistic neighborhoods of the vehicles. This scenario has a warmup period of $30\mathtt{min}$.

The scenario used in this section is very similar to the previous scenario in Section 7.3, but differs in the dissemination strategy, the consideration of privacy, and the cooperation between vehicles. For the dissemination strategy, we utilize a range-based dissemination strategy to reduce the probabilistic influence of the filtering based on the specific-impact of the messages. Regarding the consideration of privacy, the vehicles only provide a representation of their location to the server, which may be inaccurate for privacy-sensitive vehicles. The server considers this imprecision-area in the dissemination of messages, by providing every message that might be relevant for a vehicle at any possible location in this area. This also increases the transmission range of events, which we solve by generating events around the simulated area, i. e., events at locations outside of the area in which vehicle traffic is simulated. This is necessary to (i) analyze the influence of privacy appropriately and (ii) keep the simulation runtime at an acceptable level. As the third difference to the previous scenario, a vehicle broadcasts every message, received via the cellular network, in its one-hop environment if not stated differently. This is required to enable the cooperative reception of messages, which is analyzed in this section.

In the following, we first introduced the scenario-specific metrics and parameters and then describe the considered state-of-the-art approaches for hybrid communication in vehicular networks.

7.4.1.1 *Metrics*

For the evaluation of our approach to approximate vehicular networking, we utilize metrics capturing achieved communication quality and the consumed bandwidth.

Relative Utility: The relative utility measures the achieved utility compared to the maximum achievable utility for each vehicle. For the maximum achievable utility, it is assumed that all relevant messages had been received. Thus, a relative utility of 1 states that all sent messages have been received, while a relative utility of 0 states that not a single message sent to the vehicle has been received.

Parameter	Value
Trust factor	0%, 25%, **50%**, 75%, 100%
Assigned bandwidth [msg./s]	0.01, 0.1, **1**, 10
Radius of imprecision area (vector) [km]	(0, 1, 10)
Distribution of privacy (vector)	(50%, 30%, 20%)
Share of misbehaving vehicles	**0%**, 25%, 50%, 75%, 100%
Base-impact (vector)	(1, 10, 100, 1000)
Relative event frequency (vector)	(90%, 9%, 0.9%, 0.1%)
Dissemination radius [km] (vector)	(10, 1, 100, 100)

Table 4: Simulation parameters for the evaluation of hybrid vehicular networks.

Used Bandwidth: The used bandwidth captures the bandwidth usage of the vehicles and is always relative to the available bandwidth. Thus, a used bandwidth of 1 states that the vehicles have used all available bandwidth without exceeding it. If the used bandwidth is above 1, it means that the vehicles exceed the available bandwidth. Similarly, a used bandwidth below 1 states that the available bandwidth has not been used fully.

7.4.1.2 *Parameters*

In Table 4, a list of parameters is presented, which are expected to influence the performance of the different approaches. The bold values are the values used as default if not stated differently. In the beginning, we determine the trust factor τ, which is utilized to increase the robustness of our game-theoretic approach at the cost of the maximum achievable performance. Thus, we first determine the trust factor, which maximizes the performance of our game-theoretic approach in our simulation scenario. After that, we investigate the influence of bandwidth and the influence of privacy to the system. Our last parameter is the share of malicious vehicles, that do not share any data with the vehicles in their proximity.

In addition, we need to generate load for the network. For this purpose, we introduced four different event types, with an exponentially distributed impact. We assume that the event appearance is high for messages with a low base-impact, and high for messages with a high base-impact. Besides, the generated messages are disseminated in a circular area around the event, which is chosen such that the influence of privacy to the event types differs.

7.4.1.3 *Reference Approaches*

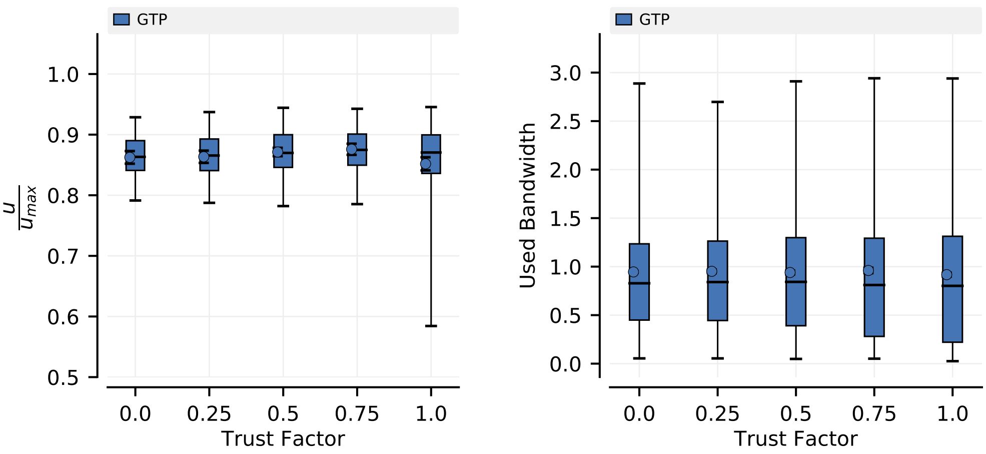
In this section, we describe the reference approaches used to evaluate the performance of our approach to approximate vehicular networking. For better comparability, we implemented our impact-based prioritization for all of the approaches. Thus, all approaches generally prioritize high-impact messages but vary in the coordination of the transmission between vehicles.

Non-Cooperative Approach (NC): The non-cooperative approach does not exchange any messages with other vehicles, i. e., relies only on its cellular network interface to receive messages about its environment. This approach is used as a baseline to evaluate the performance gain through performance in various settings. It is expected to perform worse than the cooperative approaches in most of the cases, as the cooperative approach can share bandwidth to coordinate the transmission of messages.

Cluster-based Approach (CL): In the cluster-based approach, vehicles form clusters with vehicles in their proximity to coordinate the transmission of data [130, 196, 200]. For this purpose, they elect a leader, so-called cluster-head, which is responsible for the reception of messages provided by the backend and shares them locally. All other vehicles disable their network interfaces and fully rely on the cluster-head. As a result, the transmission is very efficient, but the reliance on a single other vehicle might cause message loss in case of a disconnect or misbehavior of the cluster-head. Due to the complexity of fault detection in wireless networks, the cluster-members wait for the cluster-head to timeout in case of a connection loss until they reorganize the cluster. This limits the maximum performance of this approach but is a necessary assumption for the real-world applicability of cluster-based approaches. In conclusion, this approach is expected to perform well if the number of disconnects and misbehaving vehicles in the network is low.

Global-Knowledge Cluster-based Approach (GK): Similar to the cluster-based approach, the global-knowledge cluster-based approach forms clusters to increase the efficiency of the message transmission. In contrast to the cluster-based approach, this approach can immediately detect disconnects and initiate a reclustering. Thus, the time of disconnect is minimized, which improves the performance of the cluster-based approach. Additionally, every message transmitted via Wifi is immediately delivered to every vehicle in range without considering message loss. These assumptions are not realistic, but this approach is only used as a reference to analyze the maximum capabilities of the network. We expect this approach to perform well in settings without misbehaving vehicles.

Game-Theoretic Approach (GTP): The last of the reference approaches is our game-theoretic approach as described in Chapter 5. This approach adapts to the environmental conditions by adapting its robustness to the impact of data. That is, the vehicles optimize their communication compared to the *NC* approach by reducing the redundancy of data transmission. The level of reduction depends on the impact of the data that could be missed and the impact of the data that could be additionally received based on the freed bandwidth. We expect this approach to perform well in most settings, but perform slightly worse than the global-knowledge cluster-based approach due to the additional redundancy of our approach.



(a) Relative utility. (b) Utilized bandwidth.

Figure 22: Influence of the trust factor τ on the performance of our *GTP* approach.

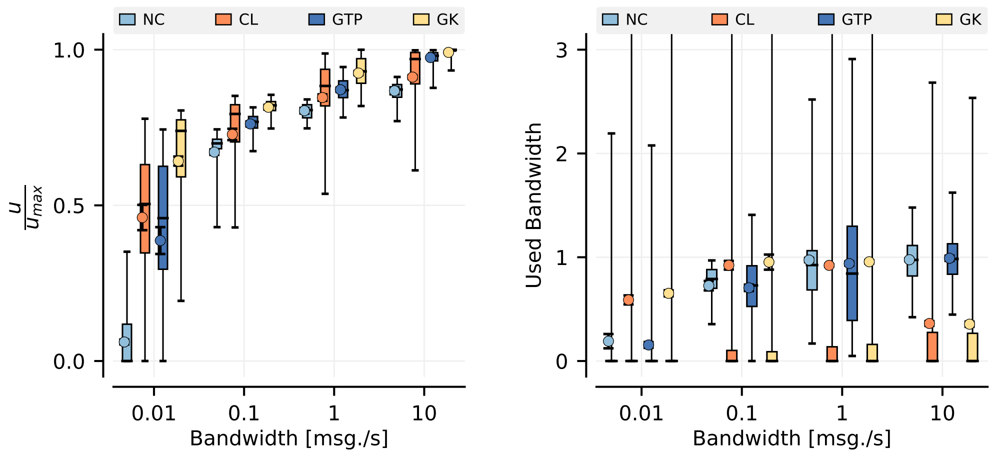
7.4.2 *Determining the Trust Factor* τ

In order to analyze the performance of our *GTP* approach, we need to analyze the influence of the trust factor τ. This trust factor τ can be between 0 and 1, which influences the selected strategy of a specific vehicle. That is, the chosen strategy might not be optimal regarding our developed game. However, every strategy that may be selected is developed using the methods proposed in Chapter 5, i. e., even a trust factor $\tau = 0$ is not assuming non-cooperative networks but selects the strategy for a cooperative network with the least negative influence if other vehicles do not cooperate. While no cooperation is one possibility for $\tau \neq 1$, frequent topology changes and the induced disconnects also might require a lower τ.

We aim at selecting the τ that leads to the highest average performance for a vehicle. Figure 22 displays the performance of our *GTP* approach for different trust factors τ. We can observe in Figure 22a that the influence of the trust factor on the performance of the results is small. The only major change is the standard deviation of the achieved utility, which is the smallest for between $\tau = 0.25$ and $\tau = 0.75$. The utilized bandwidth is almost independent of the trust factors as shown in Figure 22b. Thus, we choose a trust factor of 0.5 for the remainder of this evaluation to increase the predictability and expressiveness regarding the performance of the obtained results.

7.4.3 *Influence of the Available Communication Bandwidth*

Figure 23 displays the performance of the approaches depending on the available bandwidth. In general, all approaches increase in utility with an increasing amount of available bandwidth as shown in Figure 23a. A higher bandwidth leads to a higher

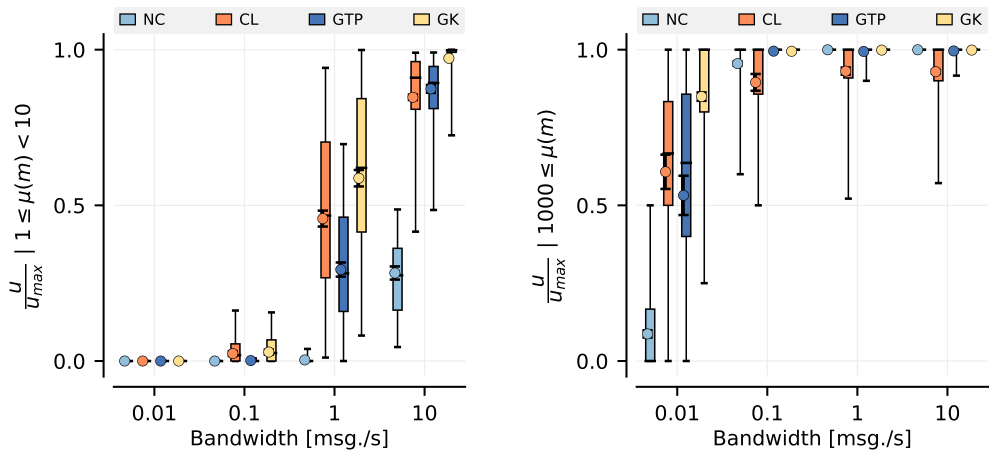


(a) Relative utility.

(b) Utilized bandwidth.

Figure 23: Influence of the available bandwidth on the performance of the approaches.

share of messages that can be received by each individual vehicle, which consequently increases the total impact of received messages and thus the utility. We can observe that our *GTP* approach always performs better than the approach without cooperation, which shows the advantages of cooperation between vehicles in a vehicular network. However, compared to the cluster-based approaches, our approach performs worse if the available bandwidth is 0.01, as our approach induces redundancy to ensure the reception of high-impact messages even after topology changes. However, this redundancy reduces the available bandwidth for the reception of other messages. In contrast, the cluster-based approaches utilize the available bandwidth generally more efficiently as the messages are not received redundantly. For a low bandwidth of 0.01, the performance decrease due to redundancy is higher than the performance gain by the robustness to topology changes. This changes when we increase the available bandwidth to 0.1 and 1, for which our approach outperforms the *CL* approach. For these bandwidths, the negative influence of a limited redundancy on the utility decreases, while the negative influence of topology changes remains constant. We perform worse compared to the *GK* approach, which is justified by the optimal coordination and bandwidth utilization of this approach. At a bandwidth of 10, the performance increase of our *GTP* approach over the realistic cluster-based approach becomes even more significant, as the utility increase of the *CL* approach is small. This is justified by the difficulty of failure-detection for the *CL* approach, which limits its maximum utility. Thus, the robustness of our approach gained by redundancy improves the performance significantly in this setting. Additionally, the difference to the global-knowledge-based approach becomes very small, although our approach is less efficient due to its induced redundancy. The reason for this is shown in the bandwidth utilization in Figure 23b. While the cluster-based approaches do not fully utilize the available bandwidth at a bandwidth of 10, our *GTP* approach and the *NC* approach



(a) Base-impact per bit $\mu_B(m) = 1$.

(b) Base-impact per bit $\mu_B(m) = 1000$.

Figure 24: Influence of the available bandwidth on the utility of messages with different base-impact for the different approaches.

still utilize the bandwidth fully. The cluster-based approach cannot fully utilize the bandwidth due to the selection of only a single cluster-head. If this cluster-head aggregates enough bandwidth to receive all available messages, bandwidth might be left unused. In contrast, our *GTP* approach always selects the number of receiving vehicle based on its neighborhood, the message importance, and the available bandwidth. Thus, it ensures the full utilization of bandwidth and utilizes it to increase the robustness of the system. Besides, we can observe that all approaches do not exceed the predefined bandwidth requirements on average, while the distribution of bandwidth utilization varies greatly between vehicles. Due to the explicit coordination in cluster-based approaches, these approaches select few vehicles with very high bandwidth usage, which are responsible for the reception of all messages. Compared to that, our *GTP* approaches distribute the load much more evenly in the network, as every vehicle is considered to be similar. It is evident that there is still some fluctuation in bandwidth between our vehicles, but this fluctuation is much less significant compared to the cluster-based approaches. Thus, our *GTP* approach distributes the load in the network much more equally compared to cluster-based approaches.

When investigating the utility score only considering certain impact-levels as shown in Figure 24, we can observe a clear dependency between the achieved utility and the message impact. That is, as high-impact messages are generally prioritized by all approaches, the low-impact messages generally achieve a worse communication quality than the high-impact messages. Figure 24a displays the relative utility for the low-impact messages. If the assigned bandwidth is 0.01, the bandwidth is not sufficient to transmit these messages, which leads to a relative utility of almost 0. With increasing bandwidth, the possibility of receiving these messages increases likewise. For a bandwidth of 0.1, we can observe the efficiency of the cluster-based approaches, which can receive low-impact messages in certain situations, while our *GTP* approach does not receive them due to the increased redundancy for high-impact messages. For an assigned bandwidth of 1, our *GTP* approach performs worse than both of the cluster-based approaches due to a similar reason. However, when the assigned bandwidth is 10, our approach outperforms the *CL* approach slightly, as the bandwidth is sufficient to perform a reliable transmission even for low-impact messages. For the *CL* approach, the missing compensation for topology changes reduces the achieved utility. Additionally, we can observe that our *GTP* approach has a much higher relative utility than the *NC* approach for the low-impact messages, which is caused by lowering the probability to receive high-impact messages. In Figure 24b, the relative utility for high-impact messages is shown. In contrast to the low-impact messages, the relative utility is always very high due to the high importance of these messages. The cooperative approaches achieve a relative utility of almost 1 if the bandwidth is at least 0.1, while the *NC* approach requires at least a bandwidth of 1 to receive the majority of high-impact messages. Thus, we can clearly observe the benefit of cooperation in the low-bandwidth case, as only the cooperative approaches can receive the majority of high-impact messages. When the bandwidth is at least 0.1, we can see that our *GTP* approach achieves a much higher relative utility compared to the *CL* approach. This is justified by the poor performance of the *CL* approach , which fails to receive 100% of the messages independent of the available bandwidth due to the lack of robustness to topology changes. In contrast, our approach adapts its robustness to the impact of the messages. For high-impact messages, our *GTP* approach is very robust, which is observable due to the non-observable difference between the *GK* approach and our *GTP* approach. Thus, we can see that our approach adapts its robustness to the impact of messages, which is a pivotal property in the dissemination of impact-aware data in a vehicular network. Thus, our *GTP* approach receives high-impact messages much more reliably compared to the *CL* approach, which lets us confirm hypothesis 1.

7.4.4 *Influence of Location Privacy*

Figure 25 displays the influence of location privacy to the system. For this purpose, we varied the share of privacy-sensitive vehicles, which have an imprecision area with a radius of 10km. We can clearly observe that a high share of location privacy decreases the performance of all analyzed approaches. However, we can observe a small performance gain of our *GTP* approach while the share of privacy increases to 20%, as the additional possibility of coordination between multiple privacy-levels is utilized. Additionally, the slope of the relative utility of our *GTP* is much lower at low levels of location privacy than for the *NC* approach and the cluster-based approaches. Thus, our *GTP* approach can maintain its performance until roughly 30% of vehicles are privacy-sensitive, after which its performance also starts to decrease. Interestingly, the performance gain of the *GK* approach is lower if the share of privacy is high. That is, as the bandwidth gained through its more efficient coordination of the transmission cannot be utilized, as it cannot receive the context-sensitive messages efficiently. For the same reason, the performance of the *CL* approach decreases even below the *NC* ap-

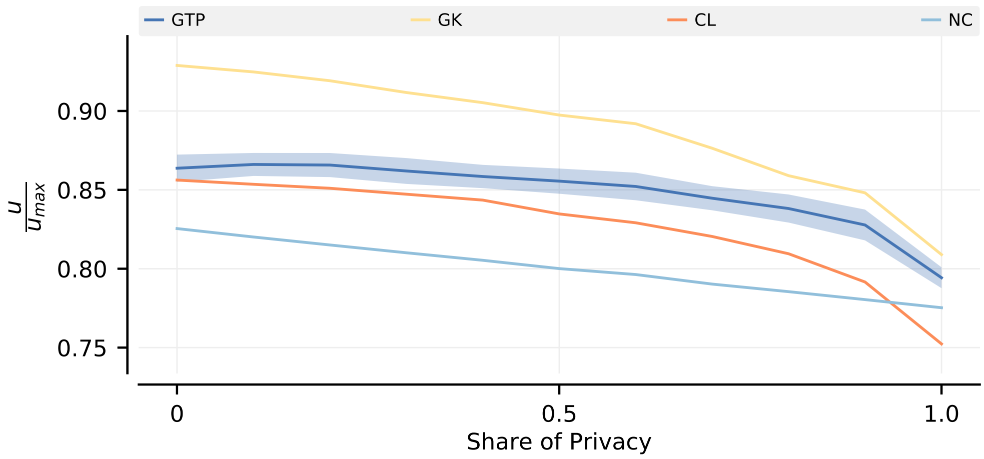
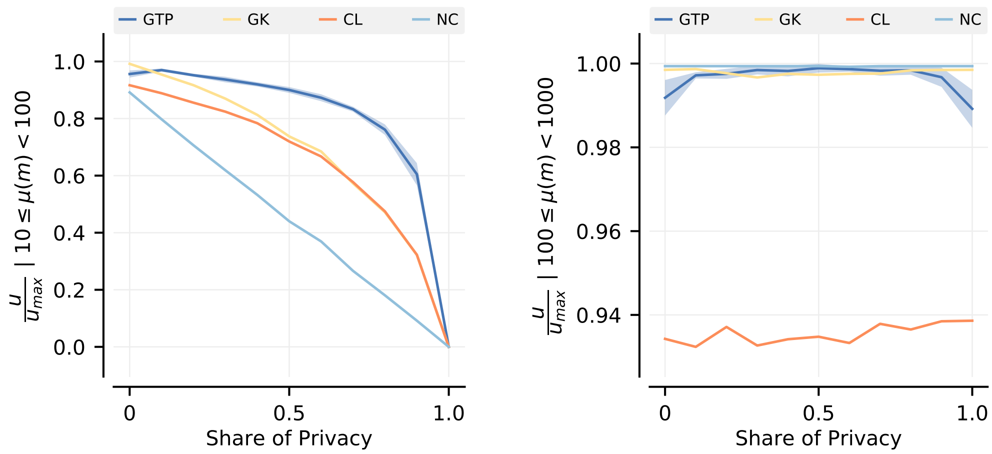


Figure 25: Influence of the location privacy on the utility for the different approaches.

proach. These observations are confirmed in Figure 26. Figure 26a displays the relative utility for highly context-sensitive messages of low impact, while Figure 26b displays the relative utility for barely context-sensitive messages of high impact. The figures for the impact-levels with a base-impact of 1 and 1000 are available in Section A.4. In Figure 26a, the performance of all approaches is relatively similar for a privacy share of 0%. However, the behavior changes drastically when the share of privacy increases. Our *GTP* approach can maintain a high level of utility until a privacy-level of 70%, after which it decreases drastically and drops to 0% like the other approaches. This constant performance is possible due to the additional coordination that can be achieved between two privacy-levels. With an increasing privacy-level, the utility of all approaches except our *GTP* approach decreases almost linearly with the privacy-level, which is due to the higher bandwidth consumption if the privacy-level increases. For our *GTP* approach, however, the utility first decreases, then increases again, and decreases finally to 0. The reason for this decrease to 0 is that the messages of impact 10 are highly context-sensitive, i. e., an efficient reception of these messages is impossible for a privacy-sensitive vehicle. For high-impact messages with low context-sensitive, the behavior is different as shown in Figure 26b. For the cluster-based approaches, the utility is independent of the share of privacy, as these approaches utilize the available bandwidth very efficiently and can compensate for the increased bandwidth consumption through privacy, which is 21% of the available bandwidth. However, we can observe the aforementioned instability of the *CL* approach, which can only achieve a utility below 94%. The performance of the *NC* approach is similar to the *GK* approach, as the *NC* approach cannot efficiently receive lower-impact messages and, thus, is forced to receive these high-impact messages with high reliability. For our *GTP* approach, the behavior of the utility is very interesting, as it is lower than the *NC* approach and the *GK* approach for a share of privacy of 0%, but then increases to roughly the same level, and decreases again at a share of privacy of 100%. At first sight, this behavior is unexpected, as the addition of privacy leads to both an increase and then a



(a) Base-impact per bit $\mu_B(m) = 10$.

(b) Base-impact per bit $\mu_B(m) = 100$.

Figure 26: Influence of location privacy on the utility of messages with different base-impact for the different approaches.

decrease in utility. This behavior can be explained again with the implicit coordination that can be achieved if vehicles of multiple privacy-levels are close to each other. In this case, the privacy-sensitive and the privacy-insensitive vehicles can coordinate the transmission of data, which increases the available bandwidth for both privacy-levels through cooperation. Thus, the performance of our *GTP* approach increases if this potential is available, i. e., if the share of privacy in the system is neither 0% nor 100%. This analysis lets us confirm hypothesis 2, i. e., our *GTP* approach mitigates the negative effects of location privacy and is able to maintain a constant performance if the share of privacy is low.

7.4.5 *Influence of Misbehaving Vehicles*

In this section, we analyze the robustness of the approaches to misbehaving vehicles. As already shown in Chapter 5, it is not reasonable to utilize a different strategy to increase the benefit of a single vehicle. Similarly, it is not feasible for any vehicle to forge messages, as the messages disseminated in a vehicular network are generally considered to be cryptographically protected. However, a misbehaving vehicle might not share received messages with vehicles in its proximity, which does not influence the utility of the misbehaving vehicle, but decreases the performance of its surroundings. We assume that the vehicles in proximity of the misbehaving vehicle (both behaving and misbehaving) are not aware of the misbehaving vehicle's unwillingness to cooperate, and the misbehaving vehicle behaves as a cooperating vehicle in role-finding and coordination processes.

Figure 27 displays the influence of misbehaving vehicles on the behaving vehicles in the network. Notice that this plots ends at an share of misbehaving vehicles of 75%,

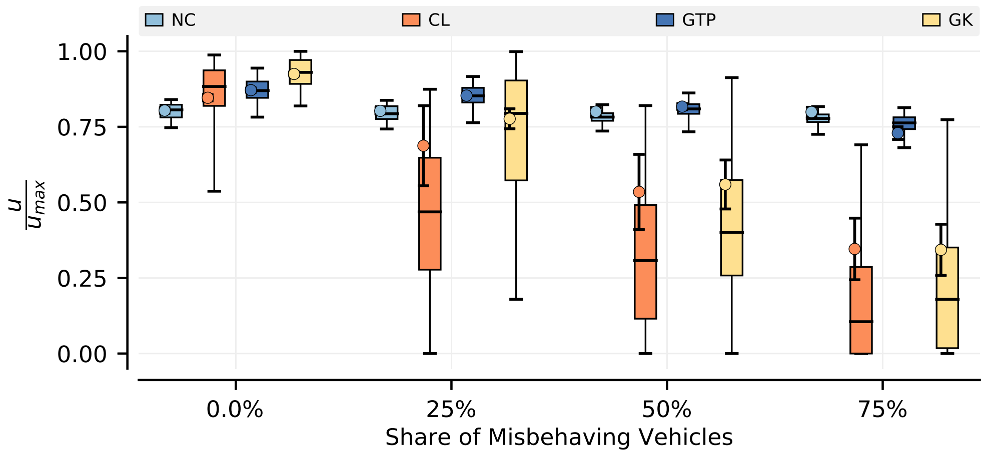


Figure 27: Influence of the misbehaving vehicles on the utility of behaving vehicles for the different approaches.

as there are no behaving vehicles if the share of misbehaving vehicles is 100%. It is evident that all cooperative approaches are negatively influenced by an increasing share of misbehaving vehicles, while the *NC* approach remains unaffected. For the *NC* approach, there is no communication between vehicles, such that an misbehaving vehicle does not behave any different to the other vehicles. The only noticeable change to the *NC* approach is a decrease in the variation, as the number of behaving vehicles decreases with an increasing share of misbehaving vehicles. For the cluster-based approaches, a big performance drop can be observed, which is almost linear with the share of misbehaving vehicles. That is expected, as the vehicles in the cluster-based approaches generally rely on exactly one other vehicle to deliver messages to them. If this vehicle is misbehaving, they do not receive any messages, leading to a utility of 0 while they rely on that vehicle. Thus, the performance of the cluster-based approaches drops below the *NC* approach already if 25% of vehicles are misbehaving. While 25% of vehicles is a large share of vehicles when considering the whole network, the local performance of a vehicle will always drop if an misbehaving vehicle is elected as cluster-head. Compared to that, our *GTP* approach never relies only on a single vehicle when receiving data, but ensures that high-impact messages are transmitted redundantly. Thus, the performance stays relatively stable even for a high share of misbehaving vehicles, while only decreasing below the *NC* approach if this share is 75%. But even at a share of misbehaving vehicles of 75%, the performance decrease of our *NC* approach is only marginal. In total, our *GTP* approach decreases its performance by only 16.4%, which shows the comparably high robustness of our approach to this type of misbehavior. While such a high share of misbehaving vehicles is unlikely due to the highly restricted access to the hardware in vehicles, it also shows the robustness of our *GTP* approach to message drop or any form of transmission errors, which would have a similar effect as this type of misbehavior, which lets us confirm hypothesis 3.

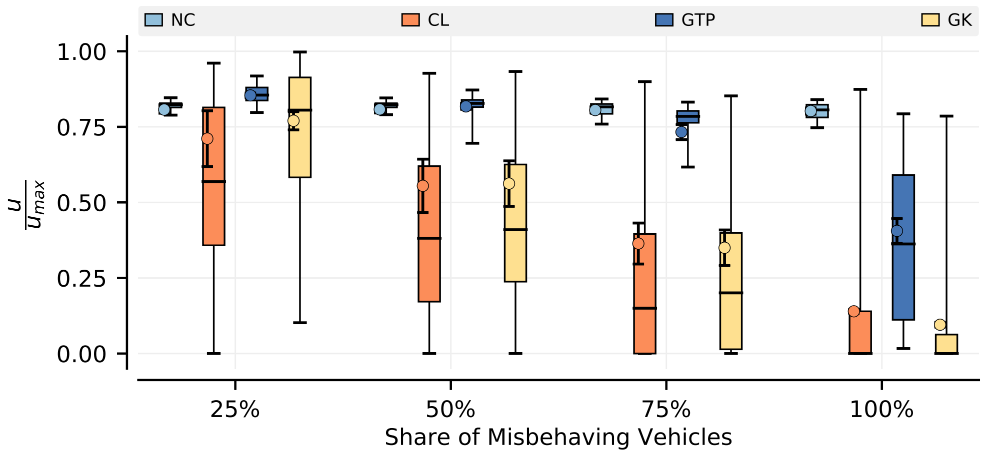


Figure 28: Influence of the misbehaving vehicles on the utility of misbehaving vehicles for the different approaches.

As an addition, we review our claim that this type of non-cooperation does not increase the performance of a misbehaving vehicle. For this purpose, Figure 28 displays the performance of only the misbehaving vehicles in the network. Notice that a share of misbehaving vehicles of 0% is excluded from this graph, as there are no misbehaving vehicles in the network in this case. When comparing the performance of the misbehaving vehicles with the corresponding performance of behaving vehicles, we can see that their performance is only slightly higher. However, when comparing the results to the case without misbehavior, we can clearly observe that the performance of misbehaving vehicles does not increase, but only decreases slightly less compared to behaving vehicles. That is, as the misbehaving vehicles also rely on other misbehaving vehicles for the reception of messages in some situations, which also decreases their overall performance. Thus, there is no gain for misbehaving vehicles from refusing to cooperate. At a share of misbehaving vehicles of 100%, we can observe that the performance of all approaches except the *NC* approach decreases drastically, but our *GTP* approach is able to maintain a higher utility than the cluster-based approaches.

Summarizing, we provide an insight into the performance of our approach to approximate vehicular networking based on probabilistic mechanisms. Based these mechanisms, vehicles can coordinate the reception of messages and increase their overall utility significantly. Compared to cluster-based approaches, which rely on only a single other vehicle, our game-theoretic approach adapts the number of recipients of the transmission based on the impact for each individual vehicle, such that the transmission of high-impact messages is very robust. This robustness is an important factor to compensate for the frequent topology changes, message drops, and misbehavior. Thus, our concept to approximate vehicular networking adapts its behavior to the transmitted data and networking conditions, such that the overall performance of the network participants is maximized.

8 SUMMARY, CONCLUSIONS, AND OUTLOOK

To conclude our work, we summarize the content of the previous chapters and state our main contributions in the following. We then draw conclusions based on our obtained results. Finally, we discuss potential future work regarding data assessment and approximate vehicular networks.

8.1 SUMMARY OF THE THESIS

In Chapter 1, we described the challenges for future vehicular networks, focusing on the data dissemination over large distances to support future autonomous vehicles. We motivated the necessity for approximate vehicular networks and the advantages of prioritization based on the quality of the disseminated data. In Chapter 2, we analyzed existing mechanisms for data dissemination in large-scale vehicular networks and analyzed existing concepts for data quality assessment in these networks. As a possible combination, we analyzed the concept of approximate networks in general and the possibilities of quality-aware vehicular networks. For that purpose, we provide a detailed insight into the necessary components and mechanisms of large-scale vehicular networks in Chapter 3. Based on our analysis of the state-of-the-art and our scenario, we present and discuss the following contributions of this thesis.

8.1.1 *Contributions*

Chapter 4 contains our first two contributions, which focus on the determination of data impact considering different influence factors. As the *first contribution*, we proposed our accuracy-aware aggregation scheme for measurements produced by multiple vehicles with different accuracy in Section 4.2. This is a common issue for vehicular networks, as the quality of measurements provided by a certain vehicle depends on a multitude of factors, like manufacturer and vehicle type. Additionally, the lifetime of an event influences the weight that old measurements should have on the current aggregate. Thus, our innovative aggregation scheme considers the accuracy of sensors on a measurement basis and the properties of the underlying event to reduce the number of false aggregates in the network. For this purpose, we designed a weighting function, that constitutes the influence of measurements depending on their age. To determine the behavior of this function, we developed and solved an optimization problem aiming at reducing false aggregates in the system.

In Section 4.3 and Section 4.4, we describe our second contribution, which is the determination of data relevance and its impact on a receiving vehicle. An accurate estimation of data relevance and impact is an important aspect of future vehicular networks due to the increase in data availability in these networks. To cope with this issue, we develop a prediction-based approach for the determination of data relevance, which predicts the future context of the vehicle and the future state of the event. That is, we utilize the expected lifetime of the event in comparison with the available routes of the vehicle to estimate the probability that the vehicle will encounter that event while it is active. We then combine these two contributions to determine the impact of a piece of data on a vehicle, considering its accuracy, its relevance, and the importance given by its data type. This impact is an important metric to prioritize and filter data in large-scale vehicular networks.

As our *third contribution* in Chapter 5, we propose our concept of approximate vehicular networks using probabilistic filtering of messages based on their impact. We investigate both cellular networks and hybrid networks, i. e., networks in which the vehicles may cooperate locally via the Wifi interface. As both these networks rely on the transmission of data by a server, the server generally requires an accurate representation of the vehicle's context at the server, which compromises the privacy of the passengers. Our innovative concept of approximate vehicular networks naturally includes the privacy of passengers in the transmission process, as vehicles can choose to provide only an inaccurate representation of their context. We then develop an approach to coordinate the transmission of data in a network, in which vehicles with different privacy constraints are present. For this purpose, we model the transmission of data as a utility-based game, in which the vehicles utilize a mixed strategy considering the strategies of vehicles in proximity. We then develop a concept of determining the optimal strategy for each vehicle in this network and analyze the properties of our found solution. Furthermore, we designed and developed our VEHICLE.KOM platform described in Chapter 6, which is used as the basis for our extensive evaluation in Chapter 7. For this purpose, we modeled the environment and the vehicles according to our findings from Chapter 3, and connected the traffic simulator Simulation of Urban Mobility (SUMO) to utilize realistic vehicle movement in our simulation.

8.1.2 *Conclusions*

In our extensive evaluation, we provided insights into the behavior of our approaches under varying environmental conditions and show that our approaches improve the performance of future vehicular networks.

In Section 7.2, we showed that our innovative aggregation scheme drastically improves the data quality when aggregating measurements of heterogeneous sensor sources. Our approach can compensate low sensor accuracies by increasing its robustness dynamically and, thus, increasing the influence of old measurements, such that the aggregation relies on more data. Additionally, our approach adapts its behavior to the lifetime of the event, by increasing its adaptability to changes if the expected event lifetime is low. Thus, our aggregation scheme constitutes a significant advance towards an increased data quality in future vehicular network, in which heterogeneous sensor sources will dominate.

We then evaluate the performance of our relevance-based dissemination approach in cellular-only vehicular networks in Section 7.3. Our relevance-based dissemination approach is a pivotal contribution to reduce the bandwidth usage in future vehicular networks. In the evaluation, we show that our approach achieves a significantly higher precision compared to geocast-based approaches, i. e., our approach transmits a higher share of relevant data to the receiving vehicles. Our relevance-based approach performs slightly worse regarding the total number of relevant messages received compared to the geocast-based approaches if the available bandwidth is very high, but utilizes the bandwidth much more efficiently. This enables our relevance-based approach to maintain its performance if the bandwidth decreases, while the geocast-based approach drastically drops in performance. We also showed that our relevance-based approach considers the impact of a message in the dissemination, leading to a higher share of high-impact messages being transmitted if the bandwidth is limited. This is an important aspect of future vehicular networks to reduce the influence of insufficient bandwidth. With that, our relevance-based approach constitutes a significant contribution towards efficient dissemination of messages in future vehicular networks.

In Section 7.4, we then evaluate the performance of our innovative approach for approximate vehicular networks based on probabilistic mechanisms. Our approach naturally incorporates privacy in the system and enables the vehicles to cooperate without explicit coordination. We investigated the performance of our approach compared with approaches for explicit coordination and showed that our approach outperforms these approaches in a realistic setting. That is, as our approach adapts its robustness to the impact of messages transferred through the network, which makes the reception of high-impact messages very probable. Approaches with explicit coordination lack this robustness, which leads to a stronger influence of topology changes and message drop to these approaches. We then investigated the influence of privacy on our developed approach and observed that our approach compensates for low levels of privacy and maintains the network performance. To additionally emphasize the robustness of our approach, we investigated its performance in networks with adversaries, which do not cooperate with other vehicles. Our approach can handle this type of adversaries very well, which again proves the robustness of our developed approach. Thus, our concept for approximate vehicular networks based on probabilistic mechanisms is a significant contribution, which enables cooperation in future vehicular networks by adapting the coordination to the impact of the shared messages.

8.2 OUTLOOK

Our concept for approximate networks with probabilistic behavior builds the foundation for further research. While we proposed the usage of this paradigm and concepts for data-centric data dissemination in vehicular networks, other research field can greatly benefit from the insights gained in this thesis. Especially when data is shared that imposes the user's privacy, our approach can be used as a concept to incorporate different privacy demands into the system. Additionally, our concept for the implicit cooperation between nodes in decentralized networks provides very interesting challenges when applied to other (potentially less mobile) networks, in which, e. g., a reliance on a single other node is not desired. Additionally, the impact-based prioritization and filtering offer new research challenges in other areas, which should be further investigated. As an example, the impact of data in disaster networks needs to be determined differently from our impact considerations in this thesis and is arguably more human-centric. In addition to that, incentive mechanisms for cooperation and sharing of data have not been considered in this thesis, which are considered to be interesting especially in Smart Cities and in the Internet of Things. For that purpose, also pricing mechanisms for data need to be developed, which can base our considerations of the event impact. As an example, high-quality data can be sold for a higher price compared to low-quality data. With these considerations, intermediate nodes could use our aggregation scheme to aggregate low-quality and low-price data, and sell the obtained high-quality and high-price data. This provides network nodes with spare resources the possibility to utilizes these resources to benefit the network and obtain a benefit by themselves.

Our mechanisms for the assessment of data quality and for the consideration of this quality in the dissemination, provide the foundation of further research in the direction of approximate networks. With our VEHICLE.KOM platform, we additionally provide the possibility for other researchers to develop new concepts for approximate vehicular networks and test them in a simulative environment.

ACKNOWLEDGMENTS

This thesis was funded by the LOEWE initiative (Hessen, Germany) within the NICER project, by German Federal Ministry of Education and Research (BMBF) within the SMARTER project, and by the German Research Foundation (DFG) as part of the project B1 within the Collaborative Research Center (CRC) 1053 - MAKI.

[1] *User Equipment (UE) radio transmission and reception (FDD)*. Technical specification. Sophia Antipolis, FR: 3GPP, 1999.

[2] *User Equipment (UE) radio transmission and reception (TDD)*. Technical specification. Sophia Antipolis, FR: 3GPP, 1999.

[3] *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN)*. Technical specification. Sophia Antipolis, FR: 3GPP, 2010.

[4] *Technical Specifications and Technical Reports for a UTRAN-based 3GPP system*. Technical specification. Sophia Antipolis, FR: 3GPP, 2011.

[5] *Dedicated Short Range Communications (DSRC) Service*. Standard. Washington DC, USA: Federal Communications Commission, Mar. 1999.

[6] Omid Abedi, Mahmood Fathy, and Jamshid Taghiloo. "Enhancing AODV routing protocol using mobility parameters in VANET." In: *Proc. of International Conference on Computer Systems and Applications (AICCSA)*. Mar. 2008, pp. 229–235. DOI: `10.1109/AICCSA.2008.4493539`.

[7] V. S. Abhayawardhana, I. J. Wassell, D. Crosby, M. P. Sellars, and M.G. Brown. "Comparison of empirical propagation path loss models for fixed wireless access systems." In: *Proc. of Vehicular Technology Conference (VTC)*. Vol. 1. IEEE. 2005, pp. 73–77.

[8] Kazi Masudul Alam, Mukesh Saini, and Abdulmotaleb El Saddik. "Toward social internet of vehicles: Concept, architecture, and applications." In: *IEEE access* 3 (2015), pp. 343–357.

[9] Waleed Alasmary and Weihua Zhuang. "Mobility impact in IEEE 802.11p infrastructureless vehicular networks." In: *Ad Hoc Networks* 10.2 (2012). Recent Advances in Analysis and Deployment of IEEE 802.11e and IEEE 802.11p Protocol Families, pp. 222–230. ISSN: 1570-8705. DOI: `https://doi.org/10.1016/j.adhoc.2010.06.006`.

[10] Salim Allal and Saadi Boudjit. "Geocast routing protocols for vanets: Survey and guidelines." In: *Proc. of International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*. IEEE. 2012, pp. 323–328.

[11] Sabri Allani, Taoufik Yeferny, and Richard Chbeir. "A scalable data dissemination protocol based on vehicles trajectories analysis." In: *Ad Hoc Networks* 71 (2018), pp. 31–44. ISSN: 1570-8705. DOI: `https://doi.org/10.1016/j.adhoc.2017.12.003`.

[12] Bastian Alt, Markus Weckesser, Christian Becker, Matthias Hollick, Sounak Kar, Anja Klein, Robin Klose, Roland Kluge, Heinz Koeppl, Boris Koldehofe, Wasiur R. Khudabukhsh, Manisha Luthra, Mahdi Mousavi, Max Mühlhäuser, Martin Pfannemüller, Amr Rizk, Andy Schürr, and Ralf Steinmetz. "Transitions: A Protocol-Independent View of the Future Internet." In: *Proceedings of the IEEE* 107.4 (Apr. 2019), pp. 835–846. DOI: `10.1109/JPROC.2019.2895964`.

[13] Marica Amadeo, Claudia Campolo, and Antonella Molinaro. "CRoWN: Content-centric networking in vehicular ad hoc networks." In: *IEEE Communications Letters* 16.9 (2012), pp. 1380–1383.

[14] Marica Amadeo, Claudia Campolo, and Antonella Molinaro. "Enhancing IEEE 802.11 p/WAVE to provide infotainment applications in VANETs." In: *Ad Hoc Networks* 10.2 (2012), pp. 253–269.

[15] Marica Amadeo, Claudia Campolo, and Antonella Molinaro. "Information-centric networking for connected vehicles: a survey and future perspectives." In: *IEEE Communications Magazine* 54.2 (2016), pp. 98–104.

[16] Giuseppe Araniti, Claudia Campolo, Massimo Condoluci, Antonio Iera, and Antonella Molinaro. "LTE for vehicular networking: a survey." In: *IEEE communications magazine* 51.5 (2013), pp. 148–157.

[17] Mohammad Hadi Arbabi and Michele Weigle. "Using vehicular networks to collect common traffic data." In: *Proc. of International Workshop on Vehicular Internetworking (VANET)*. ACM. 2009, pp. 117–118.

[18] Arash Asadi, Qing Wang, and Vincenzo Mancuso. "A survey on device-to-device communication in cellular networks." In: *IEEE Communications Surveys & Tutorials* 16.4 (2014), pp. 1801–1819.

[19] Alessia Autolitano, Claudia Campolo, Antonella Molinaro, Riccardo M. Scopigno, and Andrea Vesco. "An insight into Decentralized Congestion Control techniques for VANETs from ETSI TS 102 687 V1.1.1." In: *Proc. of IFIP Wireless Days (WD)*. Nov. 2013, pp. 1–6. DOI: `10.1109/WD.2013.6686471`.

[20] Fan Bai, Narayanan Sadagopan, Bhaskar Krishnamachari, and Ahmed Helmy. "Modeling path duration distributions in MANETs and their impact on reactive routing protocols." In: *IEEE Journal on Selected Areas in Communications* 22.7 (2004), pp. 1357–1373.

[21] Mohamed Bakhouya, Jaafar Gaber, and Pascal Lorenz. "An adaptive approach for information dissemination in Vehicular Ad hoc Networks." In: *Journal of Network and Computer Applications* 34.6 (2011). Control and Optimization over Wireless Networks, pp. 1971–1978. ISSN: 1084-8045. DOI: `https://doi.org/10.1016/j.jnca.2011.06.010`.

[22] Aruna Balasubramanian, Ratul Mahajan, and Arun Venkataramani. "Augmenting mobile 3G using WiFi." In: *Proc. of International Conference on Mobile Systems, Applications, and Services (MobiSys)*. ACM. 2010, pp. 209–222.

[23] Xuan Bao, Uichin Lee, Ivica Rimac, and Romit R. Choudhury. "DataSpotting: offloading cellular traffic via managed device-to-device data transfer at data spots." In: *ACM SIGMOBILE Mobile Computing and Communications Review* 14.3 (2010), pp. 37–39.

[24] Javier Barrachina, Piedad Garrido, Manuel Fogue, Francisco J. Martinez, Juan-Carlos Cano, Carlos T. Calafate, and Pietro Manzoni. "VEACON: A Vehicular Accident Ontology designed to improve safety on the roads." In: *Journal of Network and Computer Applications* 35.6 (2012), pp. 1891–1900. ISSN: 1084-8045. DOI: `https://doi.org/10.1016/j.jnca.2012.07.013`.

[25] Hrishav Bakul Barua and Kartick Chandra Mondal. "Approximate Computing: A Survey of Recent Trends—Bringing Greenness to Computing and Communication." In: *Journal of The Institution of Engineers (India): Series B* (June 2019). ISSN: 2250-2114. DOI: `10.1007/s40031-019-00418-8`.

[26] Prithwish Basu, Naved Khan, and Thomas DC Little. "A mobility based metric for clustering in mobile ad hoc networks." In: *Proc. of International Conference on Distributed Computing Systems Workshops (ICDCS)*. IEEE. 2001, pp. 413–418.

[27] Alessandro Bazzi, Barbara M. Masini, Alberto Zanella, and Gianni Pasolini. "IEEE 802.11 p for cellular offloading in vehicular sensor networks." In: *Computer Communications* 60 (2015), pp. 97–108.

[28] Alessandro Bazzi, Barbara M. Masini, Alberto Zanella, and Ilaria Thibault. "On the Performance of IEEE 802.11p and LTE-V2V for the Cooperative Awareness of Connected Vehicles." In: *IEEE Transactions on Vehicular Technology* 66.11 (Nov. 2017), pp. 10419–10432. DOI: `10.1109/TVT.2017.2750803`.

[29] Richard Bellman. "On a routing problem." In: *Quarterly of applied mathematics* 16.1 (1958), pp. 87–90.

[30] Abderrahim Benslimane, Tarik Taleb, and Rajarajan Sivaraj. "Dynamic clustering-based adaptive mobile gateway management in integrated VANET3G heterogeneous wireless networks." In: *IEEE Journal on Selected Areas in Communications* 29.3 (2011), pp. 559–570.

[31] David Bernstein, Nino Vidovic, and Sohrab Modi. "A Cloud PAAS for High Scale, Function, and Velocity Mobile Applications - With Reference Application as the Fully Connected Car." In: *Proc. of International Conference on Systems and Networks Communications (ICSNC)*. Aug. 2010, pp. 117–123. DOI: `10.1109/ICSNC.2010.24`.

[32] Filipe Betzel, Karen Khatamifard, Harini Suresh, David J. Lilja, John Sartori, and Ulya Karpuzcu. "Approximate Communication: Techniques for Reducing Communication Bottlenecks in Large-Scale Parallel Systems." In: *ACM Comput. Surv.* 51.1 (Jan. 2018), 1:1–1:32. ISSN: 0360-0300. DOI: `10.1145/3145812`.

[33] Daniel Bischoff and Tobias Meuser. "Geo-based Backend Dissemination for Safety-relevant V2X Applications." In: *Proc. of Automotive meets Electronics (AmE)*. VDE Verlag, Mar. 2019, pp. 1–6. ISBN: 978-3-8007-4877-8.

[34] Salim Bitam and Abdelhamid Mellouk. "ITS-cloud: Cloud computing for Intelligent transportation system." In: *Proc. of Global Communications Conference (GLOBECOM)*. Dec. 2012, pp. 2054–2059. DOI: `10.1109/GLOCOM.2012.6503418`.

[35] Jakob Breu, Achim Brakemeier, and Michael Menth. "Analysis of cooperative awareness message rates in vanets." In: *Proc. of International Conference on ITS Telecommunications (ITST)*. IEEE. 2013, pp. 8–13.

[36] Jakob Breu, Achim Brakemeier, and Michael Menth. "A quantitative study of Cooperative Awareness Messages in production VANETs." In: *EURASIP Journal on Wireless Communications and Networking* 2014.1 (2014), p. 98.

[37] Daniel Burgstahler, Tobias Meuser, Ulrich Lampe, Doreen Böhnstedt, and Ralf Steinmetz. "ProbSense.KOM: A Probabilistic Sensing Approach for Gathering Vehicular Sensed Data." In: *Proc. of International Conference on Mobile Services (MS)*. Washington, DC, USA: IEEE Computer Society, June 2016, pp. 9–16. ISBN: 978-1-5090-2625-8.

[38] Carlos T. Calafate, Giancarlo Fortino, Sascha Fritsch, Janio Monteiro, Juan-Carlos Cano, and Pietro Manzoni. "An efficient and robust content delivery solution for IEEE 802.11 p vehicular environments." In: *Journal of Network and Computer Applications* 35.2 (2012), pp. 753–762.

[39] Murat Caliskan, Andreas Barthels, Bjorn Scheuermann, and Martin Mauve. "Predicting Parking Lot Occupancy in Vehicular Ad Hoc Networks." In: *Proc. of Vehicular Technology Conference (VTC-Spring)*. Apr. 2007, pp. 277–281. DOI: `10.1109/VETECS.2007.69`.

[40] Cristina Cano and Douglas J. Leith. "Coexistence of WiFi and LTE in unlicensed bands: A proportional fair allocation scheme." In: *Proc. of International Conference on Communication Workshop (ICCW)*. IEEE. 2015, pp. 2288–2293.

[41] Bin B Chen and Mun C. Chan. "MobTorrent: A framework for mobile Internet access from vehicles." In: *Proc. of International Conference on Computer Communications (INFOCOM)*. IEEE. 2009, pp. 1404–1412.

[42] Tsu-Wei Chen and Mario Gerla. "Global state routing: A new routing scheme for ad-hoc wireless networks." In: *Proc. of International Conference on Communications (ICC)*. Vol. 1. IEEE. 1998, pp. 171–175.

[43] Xianbo Chen, Hazem H. Refai, and Xiaomin Ma. "On the enhancements to IEEE 802.11 MAC and their suitability for safety-critical applications in VANET." In: *Wireless Communications & Mobile Computing* 10.9 (2010), pp. 1253–1269.

[44] Nan Cheng, Ning Lu, Ning Zhang, Xuemin S. Shen, and Jon W. Mark. "Vehicular WiFi offloading: Challenges and solutions." In: *Vehicular Communications* 1.1 (2014), pp. 13–21.

[45] Sejin Chun, Sangjin Shin, Seungmin Seo, Sungkwang Eom, Jooik Jung, and Kyong-Ho Lee. "A Pub/Sub-Based Fog Computing Architecture for Internet-of-Vehicles." In: *Proc. of International Conference on Cloud Computing Technology and Science (CloudCom)*. Dec. 2016, pp. 90–93. DOI: `10.1109/CloudCom.2016.0029`.

[46] IMAGinE consortium. *IMAGinE*. 2017. URL: https://www.imagine-online.de/home/.

[47] Craig Cooper, Daniel Franklin, Montserrat Ros, Farzad Safaei, and Mehran Abolhasan. "A comparative survey of VANET clustering techniques." In: *IEEE Communications Surveys & Tutorials* 19.1 (2016), pp. 657–681.

[48] L. G. Cuthbert. "Data Compression and Error Control Techniques and Applications." In: *IEE Proceedings F - Communications, Radar and Signal Processing* 134.1 (Feb. 1987), pp. 118–119. DOI: 10.1049/ip-f-1.1987.0020.

[49] Pedro M. d'Orey and Mate Boban. "Empirical evaluation of cooperative awareness in vehicular communications." In: *Proc. of Vehicular Technology Conference (VTC Spring)*. IEEE. 2014, pp. 1–5.

[50] Kashif Dar, Mohamed Bakhouya, Jaafar Gaber, Maxime Wack, and Pascal Lorenz. "Wireless communication technologies for ITS applications [Topics in Automotive Networking]." In: *IEEE Communications Magazine* 48.5 (2010), pp. 156–162.

[51] Thierry Delot, Nicolas Cenerario, and Sergio Ilarri. "Estimating the relevance of information in inter-vehicle ad hoc networks." In: *Proc. of International Conference on Mobile Data Management Workshops (MDMW)*. IEEE. 2008, pp. 151–158.

[52] Pralhad Deshpande, Xiaoxiao Hou, and Samir R. Das. "Performance comparison of 3G and metro-scale WiFi for vehicular network access." In: *Proc. of Conference on Internet Measurement (IMC)*. ACM. 2010, pp. 301–307.

[53] Ben Ding, Zehua Chen, Yan Wang, and Hui Yu. "An improved AODV routing protocol for VANETs." In: *Proc. of International Conference on Wireless Communications and Signal Processing (WCSP)*. Nov. 2011, pp. 1–5. DOI: 10.1109/WCSP.2011.6096736.

[54] Yong Ding and Li Xiao. "SADV: Static-Node-Assisted Adaptive Data Dissemination in Vehicular Networks." In: *IEEE Transactions on Vehicular Technology* 59.5 (June 2010), pp. 2445–2455. DOI: 10.1109/TVT.2010.2045234.

[55] Klaus Doppler, Mika P. Rinne, Pekka Janis, Cassio B. Ribeiro, and Klaus Hugl. "Device-to-device communications; functional prospects for LTE-advanced networks." In: *Proc. of International Conference on Communications (ICC) Workshops*. IEEE. 2009, pp. 1–6.

[56] Wassim Drira and Fethi Filali. "A Pub/Sub extension to NDN for efficient data collection and dissemination in V2X networks." In: *Proc. of International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. June 2014, pp. 1–7. DOI: 10.1109/WoWMoM.2014.6918991.

[57] Amit Dua, Neeraj Kumar, and Seema Bawa. "A systematic review on routing protocols for vehicular ad hoc networks." In: *Vehicular Communications* 1.1 (2014), pp. 33–52.

[58] Matt Duckham and Lars Kulik. “Location privacy and location-aware computing.” In: *Dynamic and Mobile GIS*. CRC press, 2006, pp. 63–80.

[59] David Eckhoff, Nikoletta Sofra, and Reinhard German. “A performance study of cooperative awareness in ETSI ITS G5 and IEEE WAVE.” In: *Proc. of Annual Conference on Wireless On-demand Network Systems and Services (WONS)*. IEEE. 2013, pp. 196–200.

[60] Stephan Eichler. “Performance Evaluation of the IEEE 802.11p WAVE Communication Standard.” In: *Proc. of Vehicular Technology Conference (VTC)*. Sept. 2007, pp. 2199–2203. DOI: 10.1109/VETECF.2007.461.

[61] *Intelligent Transport Systems (ITS); Access layer specification for Intelligent Transport Systems operating in the 5 GHz frequency band*. Standard. Sophia Antipolis, FR: European Telecommunications Standards Institute, Mar. 2012.

[62] *Intelligent Transport Systems (ITS); Framework for Public Mobile Network*. Technical Report. Sophia Antipolis, FR: European Telecommunications Standards Institute, Feb. 2012.

[63] *Intelligent Transport System (ITS); Vehicular communications; GeoNetworking; Part 4: Geographical addressing and forwarding for point-to-point and point-to-multipoint communications; Subpart 1: Media independent functionalities*. Technical Specification. Sophia Antipolis, FR: European Telecommunications Standards Institute, May 2014.

[64] *Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service*. Technical Specification. Sophia Antipolis, FR: European Telecommunications Standards Institute, Mar. 2011.

[65] *Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 3: Specifications of Decentralized Environmental Notification Basic Service*. Technical Specification. Sophia Antipolis, FR: European Telecommunications Standards Institute, Sept. 2010.

[66] Patrick T. Eugster, Pascal A. Felber, Rachid Guerraoui, and Anne-Marie Kermarrec. “The many faces of publish/subscribe.” In: *ACM computing surveys (CSUR)* 35.2 (2003), pp. 114–131.

[67] Andreas Festag. “Standards for vehicular communicationfrom IEEE 802.11 p to 5G.” In: *e & i Elektrotechnik und Informationstechnik* 132.7 (2015), pp. 409–416.

[68] Marco Fiore and Jérôme Härri. “The networking shape of vehicular mobility.” In: *Proc. of International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*. ACM. 2008, pp. 261–272.

[69] Gábor Fodor, Erik Dahlman, Gunnar Mildh, Stefan Parkvall, Norbert Reider, György Miklós, and Zoltán Turányi. “Design aspects of network assisted device-to-device communications.” In: *IEEE Communications Magazine* 50.3 (2012), pp. 170–177.

[70] Lester R. Ford. *Network flow theory*. Tech. rep. Rand Corp Santa Monica Ca, 1956.

[71] Hui Gao, Chi Harold Liu, Jian Tang, Dejun Yang, Pan Hui, and Wendong Wang. "Online Quality-Aware Incentive Mechanism for Mobile Crowd Sensing with Extra Bonus." In: *IEEE Transactions on Mobile Computing* (2018), pp. 1–1. DOI: 10.1109/TMC.2018.2877459.

[72] Jose J. Garcia-Luna-Aceves and Marcelo Spohn. "Source-tree routing in wireless networks." In: *Proc. of International Conference on Network Protocols (ICNP)*. IEEE. 1999, pp. 273–282.

[73] Mario Gerla, Eun-Kyu Lee, Giovanni Pau, and Uichin Lee. "Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds." In: *Proc. of World Forum on Internet of Things (WF-IoT)*. IEEE. 2014, pp. 241–246.

[74] Amitava Ghosh, Rapeepat Ratasuk, Bishwarup Mondal, Nitin Mangalvedhe, and Tim Thomas. "LTE-advanced: next-generation wireless broadband technology." In: *IEEE wireless communications* 17.3 (2010), pp. 10–22.

[75] Marco Giordani, Andrea Zanella, Takamasa Higuchi, Onur Altintas, and Michele Zorzi. "Emerging Trends in Vehicular Communication Networks." In: *Emerging Wireless Communication and Network Technologies: Principle, Paradigm and Performance*. Ed. by Karm Veer Arya, Robin Singh Bhadoria, and Narendra S. Chaudhari. Singapore: Springer Singapore, 2018, pp. 37–57. ISBN: 978-981-13-0395-1. DOI: 10.1007/978-981-13-0396-8_3.

[76] Hendrik-Jörn Günther, Raphael Riebl, Lars Wolf, and Christian Facchi. "Collective perception and decentralized congestion control in vehicular ad-hoc networks." In: *2016 IEEE Vehicular Networking Conference (VNC)*. Dec. 2016, pp. 1–8. DOI: 10.1109/VNC.2016.7835931.

[77] Hendrik-Jörn Günther, Raphael Riebl, Lars Wolf, and Christian Facchi. "The Effect of Decentralized Congestion Control on Collective Perception in Dense Traffic Scenarios." In: *Computer Communications* 122 (2018), pp. 76–83. ISSN: 0140-3664. DOI: https://doi.org/10.1016/j.comcom.2018.03.009.

[78] Hannes Hartenstein and Kenneth P. Laberteaux. "A tutorial survey on vehicular ad hoc networks." In: *IEEE Communications Magazine* 46.6 (June 2008), pp. 164–171. DOI: 10.1109/MCOM.2008.4539481.

[79] Heecheol Song and Hwang Soo Lee. "A survey on how to solve a decentralized congestion control problem for periodic beacon broadcast in vehicular safety communications." In: *Proc. of International Conference on Advanced Communications Technology (ICACT)*. Jan. 2013, pp. 649–654.

[80] Wendi R. Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan. "Energy-efficient communication protocol for wireless microsensor networks." In: *Proc. of Annual Hawaii international Conference on System Sciences (HICSS)*. IEEE. 2000, 10–pp.

[81] Georg Hertkorn. "Mikroskopische Modellierung von zeitabhängiger Verkehrsnachfrage und von Verkehrsflußmustern." PhD dissertation. University of Cologne, 2004.

[82] Guido R. Hiertz, Dee Denteneer, Lothar Stibor, Yunpeng Zang, Xavier P. Costa, and Bernhard Walke. "The IEEE 802.11 universe." In: *IEEE Communications Magazine* 48.1 (Jan. 2010), pp. 62–70. DOI: `10.1109/MCOM.2010.5394032`.

[83] Xiaoxiao Hou, Pralhad Deshpande, and Samir R. Das. "Moving bits from 3G to metro-scale WiFi for vehicular network access: An integrated transport layer solution." In: *Proc. of International Conference on Network Protocols (ICNP)*. IEEE. 2011, pp. 353–362.

[84] Christoph Ide, Bjoern Dusza, Markus Putzke, and Christian Wietfeld. "Channel sensitive transmission scheme for V2I-based Floating Car Data collection via LTE." In: *Proc. of International Conference on Communications (ICC)*. IEEE. 2012, pp. 7151–7156.

[85] *Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions*. Standard. Geneva, CH: International Organization for Standardization, Dec. 1994.

[86] Mohammad R. Jabbarpour, Ali Jalooli, Erfan Shaghaghi, Rafidah M. Noor, Leon Rothkrantz, Rashid H. Khokhar, and Nor B. Anuar. "Ant-based vehicle congestion avoidance system using vehicular networks." In: *Engineering Applications of Artificial Intelligence* 36 (2014), pp. 303–319. ISSN: 0952-1976. DOI: `https://doi.org/10.1016/j.engappai.2014.08.001`.

[87] Florian Jomrich, Daniel Bischoff, Steffen Knapp, Tobias Meuser, Björn Richerzhagen, and Ralf Steinmetz. "Lane Accurate Detection of Map Changes based on Low Cost Smartphone Data." In: SCITEPRESS, May 2019, pp. 126–137. ISBN: 978-989-758-374-2.

[88] Florian Jomrich, Josef Schmid, Steffen Knapp, Alfred Höß, Ralf Steinmetz, and Björn Schuller. "Analysing communication requirements for crowd sourced backend generation of HD Maps used in automated driving." In: *Proc. of Vehicular Networking Conference (VNC)*. Dec. 2018, pp. 1–8. DOI: `10.1109/VNC.2018.8628335`.

[89] Florian Jomrich, Aakash Sharma, Tobias Rückelt, Daniel Burgstahler, and Doreen Böhnstedt. "Dynamic Map Update Protocol for Highly Automated Driving Vehicles." In: *Proc. of International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS)*. SCITEPRESS, Apr. 2017, pp. 68–78. ISBN: 978-989-758-242-4.

[90] Luo Junhai, Ye Danxia, Xue Liu, and Fan Mingyu. "A survey of multicast routing protocols for mobile ad-hoc networks." In: *IEEE communications surveys & tutorials* 11.1 (2009), pp. 78–91.

[91] Mohammed Saeed Al-Kahtani. "Survey on security attacks in Vehicular Ad hoc Networks (VANETs)." In: *Proc. of International Conference on Signal Processing and Communication Systems*. IEEE. 2012, pp. 1–9.

[92] Seiya Kato, Matti Hiltunen, Kaustubh Joshi, and Richard Schlichting. "Enabling vehicular safety applications over LTE networks." In: *Proc. of International Conference on Connected Vehicles and Expo (ICCVE)*. Dec. 2013, pp. 747–752. DOI: 10.1109/ICCVE.2013.6799889.

[93] Brett Kaufman and Behnaam Aazhang. "Cellular networks with an overlaid device to device network." In: *Proc. of Asilomar Conference on Signals, Systems and Computers (ACSSC)*. IEEE. 2008, pp. 1537–1541.

[94] Uzma Khan, Shikha Agrawal, and Sanjay Silakari. "Detection of malicious nodes (DMN) in vehicular ad-hoc networks." In: *Procedia computer science* 46 (2015), pp. 965–972.

[95] Evangelia Kokolaki, Merkouris Karaliopoulos, and Ioannis Stavrakakis. "Leveraging Information in Parking Assistance Systems." In: *IEEE Transactions on Vehicular Technology* 62.9 (Nov. 2013), pp. 4309–4317. DOI: 10.1109/TVT.2013.2269015.

[96] Charalampos Konstantopoulos, Damianos Gavalas, and Grammati Pantziou. "Clustering in mobile ad hoc networks through neighborhood stability-based mobility prediction." In: *Computer Networks* 52.9 (2008), pp. 1797–1824.

[97] Sjaanie Koppel, Judith Charlton, Brian Fildes, and Michael Fitzharris. "How important is vehicle safety in the new vehicle purchase process?" In: *Accident Analysis & Prevention* 40.3 (2008), pp. 994–1004. ISSN: 0001-4575. DOI: https://doi.org/10.1016/j.aap.2007.11.006.

[98] Taek Jin Kwon and Mario Gerla. "Efficient Flooding with Passive Clustering (PC) in Ad Hoc Networks." In: *SIGCOMM Computer Communication Review* 32.1 (Jan. 2002), pp. 44–56. ISSN: 0146-4833. DOI: 10.1145/510726.510730.

[99] Lei Lei, Zhangdui Zhong, Chuang Lin, and Xuemin Shen. "Operator controlled device-to-device communications in LTE-advanced networks." In: *IEEE Wireless Communications* 19.3 (2012), pp. 96–104.

[100] Ilias Leontiadis, Paolo Costa, and Cecilia Mascolo. "Persistent content-based information dissemination in hybrid vehicular networks." In: *Proc. of International Conference on Pervasive Computing and Communications (PerCom)*. Mar. 2009, pp. 1–10. DOI: 10.1109/PERCOM.2009.4912758.

[101] Ilias Leontiadis and Cecilia Mascolo. "GeOpps: Geographical Opportunistic Routing for Vehicular Networks." In: *Proc. of International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. June 2007, pp. 1–6. DOI: 10.1109/WOWMOM.2007.4351688.

[102] Ilias Leontiadis and Cecilia Mascolo. "Opportunistic Spatio-temporal Dissemination System for Vehicular Networks." In: *Proc. of International MobiSys Workshop on Mobile Opportunistic Networking*. MobiOpp '07. San Juan, Puerto Rico: ACM, 2007, pp. 39–46. ISBN: 978-1-59593-688-2. DOI: 10.1145/1247694.1247702.

[103] Fan Li and Yu Wang. "Routing in vehicular ad hoc networks: A survey." In: *IEEE Vehicular Technology Magazine* 2.2 (June 2007), pp. 12–22. DOI: 10.1109/MVT.2007.912927.

[104] Meng Li, Pengbo Si, and Yanhua Zhang. "Delay-tolerant data traffic to software-defined vehicular networks with mobile edge computing in smart city." In: *IEEE Transactions on Vehicular Technology* 67.10 (2018), pp. 9073–9086.

[105] Patrick Lieser, Nils Richerzhagen, Simon Luser, Björn Richerzhagen, and Ralf Steinmetz. "Understanding the Impact of Message Prioritization in Post-Disaster Ad Hoc Networks." In: *Proc. of International Conference on Networked Systems (NetSys)*. Mar. 2019.

[106] Stephanie Lindsey and Cauligi S. Raghavendra. "PEGASIS: Power-efficient gathering in sensor information systems." In: *Proc. of Aerospace Conference*. Vol. 3. IEEE. 2002, pp. 3–3.

[107] Xiao Liu, Yuxin Liu, Neal N. Xiong, Ning Zhang, Anfeng Liu, Hailan Shen, and Changqin Huang. "Construction of Large-Scale Low-Cost Delivery Infrastructure Using Vehicular Networks." In: *IEEE Access* 6 (2018), pp. 21482–21497. DOI: `10.1109/ACCESS.2018.2825250`.

[108] Shou-Chih Lo, Yi-Jen Lin, and Jhih-Siao Gao. "A multi-head clustering algorithm in vehicular ad hoc networks." In: *International Journal of Computer Theory and Engineering* 5.2 (2013), p. 242.

[109] Christian Lochert, Hannes Hartenstein, Jing Tian, Holger Fussler, D. Hermann, and Martin Mauve. "A routing strategy for vehicular ad hoc networks in city environments." In: *Proc. of Intelligent Vehicles Symposium (IV)*. June 2003, pp. 156–161. DOI: `10.1109/IVS.2003.1212901`.

[110] Pablo A. Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. "Microscopic Traffic Simulation using SUMO." In: *Proc. of International Conference on Intelligent Transportation Systems (ITSS)*. IEEE, 2018.

[111] Manisha Luthra, Boris Koldehofe, and Ralf Steinmetz. "Transitions for Increased Flexibility in Fog Computing: A Case Study on Complex Event Processing." In: *Informatik Spektrum* 42.4 (2019), pp. 244–255.

[112] Manisha Luthra, Boris Koldehofe, Pascal Weisenburger, Guido Salvaneschi, and Raheel Arif. "TCEP: Adapting to dynamic user environments by enabling transitions between operator placement mechanisms." In: *Proc. of International Conference on Distributed and Event-based Systems (DEBS)*. ACM. 2018, pp. 136–147.

[113] Guilherme Maia, Andre LL Aquino, Aline Viana, Azzedine Boukerche, and Antonio AF Loureiro. "HyDi: a hybrid data dissemination protocol for highway scenarios in vehicular ad hoc networks." In: *Proc. of International Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications (DIVANet)*. ACM. 2012, pp. 115–122.

[114] Francesco Malandrino, Claudio Casetti, Carla-Fabiana Chiasserini, and Marco Fiore. "Offloading cellular networks through ITS content download." In: *Proc. of Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*. IEEE. 2012, pp. 263–271.

[115] K. Massow, B. Kwella, N. Pfeifer, F. Häusler, J. Pontow, I. Radusch, J. Hipp, F. Dölitzscher, and M. Haueis. "Deriving HD maps for highly automated driving from vehicular probe data." In: *Proc. of International Conference on Intelligent Transportation Systems (ITSC)*. Nov. 2016, pp. 1745–1752. DOI: 10.1109/ITSC.2016.7795794.

[116] Chetan B. Math, Hong Li, Sonia H. de Groot, and Ignas Niemegeers. "Fair decentralized data-rate congestion control for V2V communications." In: *Proc. of International Conference on Telecommunications (ICT)*. May 2017, pp. 1–7. DOI: 10.1109/ICT.2017.7998229.

[117] Tobias Meuser, Daniel Burgstahler, Tobias Rueckelt, Doreen Boehnstedt, and Ralf Steinmetz. "Hybrid-ProbSense.KOM: Probabilistic Sensing with Hybrid Communication for Gathering Vehicular Sensed Data." In: *Proc. of Automotive meets Electronics (AmE)*. Vol. 87. Mar. 2017, pp. 1–6. ISBN: 978-3-8007-4369-8.

[118] Tobias Meuser, Patrick Lieser, The An Binh Nguyen, Doreen Böhnstedt, and Ralf Steinmetz. "Adaptive Information Aggregation for Application-specific Demands." In: *Proc. of International Balkan Conference on Communications and Networking (BalkanCom)*. May 2017, pp. 1–6.

[119] Tobias Meuser, Björn Richerzhagen, Ioannis Stavrakakis, The An Binh Nguyen, and Ralf Steinmetz. "Relevance-Aware Information Dissemination in Vehicular Networks." In: *Proc. of International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, June 2018, pp. 588–590. ISBN: 978-1-5386-4725-7. DOI: 10.1109/WoWMoM.2018.8449751.

[120] Tobias Meuser, Ioannis Stavrakakis, and Antonio Fernández Anta. "Dynamic Vehicle Path-Planning in the Presence of Traffic Events." In: *Proc. of Conference on Local Computer Networks (LCN)*. IEEE, Oct. 2019, pp. 1–4.

[121] Tobias Meuser, Martin Wende, Patrick Lieser, Björn Richerzhagen, and Ralf Steinmetz. "Adaptive Decision Making Based on Temporal Information Dynamics." In: *Proc. of International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS)*. SCITEPRESS, Mar. 2018, pp. 91–102. ISBN: 978-989-758-293-6.

[122] Tobias Meuser, Daniel Bischoff, Björn Richerzhagen, and Ralf Steinmetz. "Cooperative Offloading in Context-Aware Networks: A Game-Theoretic Approach." In: *Proc. of International Conference on Distributed and Event-based Systems (DEBS)*. ACM, June 2019, pp. 55–66. ISBN: 978-1-4503-6794-3. DOI: 10.1145/3328905.3338535.

[123] Tobias Meuser, Daniel Bischoff, and Ralf Steinmetz. "Location Privacy in Heterogeneous Vehicular Networks." In: *Proc. of International Conference on Distributed and Event-based Systems (DEBS)*. ACM, 2019, pp. 260–261. ISBN: 978-1-4503-6794-3. DOI: 10.1145/3328905.3332515.

[124] Tobias Meuser, Patrick Lieser, The An Binh Nguyen, Doreen Böhnstedt, and Ralf Steinmetz. "Lifetime-aware information aggregation under uncertainty for advanced driver assistance systems." In: *Ad Hoc Networks* 82 (2019), pp. 56–67. ISSN: 1570-8705. DOI: https://doi.org/10.1016/j.adhoc.2018.07.018.

[125] Tobias Meuser, Oluwasegun Taiwo Ojo, Daniel Bischoff, Antonio Fernández Anta, Ioannis Stavrakakis, and Ralf Steinmetz. *Hide Me: Enabling Location Privacy in Heterogeneous Vehicular Networks*. 2020. arXiv: 2001.07170 [cs.CR].

[126] Zeeshan Hameed Mir and Fethi Filali. "LTE and IEEE 802.11 p for vehicular networking: a performance evaluation." In: *EURASIP Journal on Wireless Communications and Networking* 2014.1 (2014), p. 89.

[127] Sasa Misailovic, Stelios Sidiroglou, and Martin C. Rinard. "Dancing with Uncertainty." In: *Proc. of Workshop on Relaxing Synchronization for Multicore and Manycore Scalability*. RACES '12. Tucson, Arizona, USA: ACM, 2012, pp. 51–60. ISBN: 978-1-4503-1632-3. DOI: 10.1145/2414729.2414738.

[128] Tulika Mishra, Deepak Garg, and Manoj Madhav Gore. "A publish/subscribe communication infrastructure for VANET applications." In: *Proc. of Workshops of International Conference on Advanced Information Networking and Applications (AINA)*. IEEE. 2011, pp. 442–446.

[129] Shima Mohseni, Rosilah Hassan, Ahmed Patel, and Rozilawati Razali. "Comparative review study of reactive and proactive routing protocols in MANETs." In: *Proc. of International Conference on Digital Ecosystems and Technologies (DEST)*. IEEE. 2010, pp. 304–309.

[130] Mildred M. C. Morales, Choong Seon Hong, and Young-Cheol Bang. "An adaptable mobility-aware clustering algorithm in vehicular networks." In: *Proc. of Asia-Pacific Network Operations and Management Symposium*. IEEE. 2011, pp. 1–6.

[131] Johannes Müller, Tobias Meuser, Ralf Steinmetz, and Michael Buchholz. "A Trust Management and Misbehaviour Detection Mechanism for Multi-Agent Systems and its Application to Intelligent Transportation Systems." In: *Proc. of International Conference on Control and Automation (ICCA)*. IEEE, June 2019, pp. 1–7.

[132] Vinod Namboodiri and Lixin Gao. "Prediction-Based Routing for Vehicular Ad Hoc Networks." In: *IEEE Transactions on Vehicular Technology* 56.4 (June 2007), pp. 2332–2345. DOI: 10.1109/TVT.2007.897656.

[133] Gustavo Niemeyer. *Geohash*. 2008. URL: http://geohash.org/.

[134] Zhaolong Ning, Xiping Hu, Zhikui Chen, MengChu Zhou, Bin Hu, Jun Cheng, and Mohammad S. Obaidat. "A Cooperative Quality-Aware Service Access System for Social Internet of Vehicles." In: *IEEE Internet of Things Journal* 5.4 (Aug. 2018), pp. 2506–2517. DOI: 10.1109/JIOT.2017.2764259.

[135] Dusit Niyato, Ekram Hossain, and Ping Wang. "Optimal channel access management with QoS support for cognitive vehicular networks." In: *IEEE Transactions on Mobile Computing* 10.4 (2010), pp. 573–591.

[136] Konstantinos Oikonomou, Dimitrios Kogias, and Ioannis Stavrakakis. "Probabilistic flooding for efficient information dissemination in random graph topologies." In: *Computer Networks* 54.10 (2010), pp. 1615–1629. ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2010.01.007.

[137] Konstantinos Oikonomou and Ioannis Stavrakakis. "Performance Analysis of Probabilistic Flooding Using Random Graphs." In: *Proc. of International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. June 2007, pp. 1–6. DOI: 10.1109/WOWMOM.2007.4351694.

[138] Sergio Ortiz, Carlos T. Calafate, Juan-Carlos Cano, Pietro Manzoni, and Chai K. Toh. "A UAV-Based Content Delivery Architecture for Rural Areas and Future Smart Cities." In: *IEEE Internet Computing* 23.1 (Jan. 2019), pp. 29–36. DOI: 10.1109/MIC.2018.2884277.

[139] Christoforos Panos, Christos Xenakis, and Ioannis Stavrakakis. "A novel Intrusion Detection System for MANETs." In: *Proc. of International Conference on Security and Cryptography (SECRYPT)*. July 2010, pp. 1–10.

[140] Dan Peng, Fan Wu, and Guihai Chen. "Pay As How Well You Do: A Quality Based Incentive Mechanism for Crowdsensing." In: *Proc. of International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*. MobiHoc '15. Hangzhou, China: ACM, 2015, pp. 177–186. ISBN: 978-1-4503-3489-1. DOI: 10.1145/2746285.2746306.

[141] Tao Peng, Qianxi Lu, Haiming Wang, Shaoyi Xu, and Wenbo Wang. "Interference avoidance mechanisms in the hybrid cellular and device-to-device systems." In: *Proc. of International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE. 2009, pp. 617–621.

[142] Charles E. Perkins and Pravin Bhagwat. *Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers*. New York, NY, USA, Oct. 1994. DOI: 10.1145/190809.190336.

[143] "Personen + Unternehmen." In: *ATZelektronik* 12.4 (Aug. 2017), pp. 6–7. ISSN: 2192-8878. DOI: 10.1007/s35658-017-0055-9.

[144] Mai-Anh Phan, Rene Rembarz, and Sabine Sories. "A capacity analysis for the transmission of event and cooperative awareness messages in LTE networks." In: *Proc. of ITS World Congress*. 2011.

[145] Nuno K. Pratas and Petar Popovski. "Low-rate machine-type communication via wireless device-to-device (D2D) links." In: *arXiv preprint arXiv:1305.6783* (2013).

[146] Junaid Qadir, Arjuna Sathiaseelan, Liang Wang, and Jon Crowcroft. "Taming Limits with Approximate Networking." In: *Proc. of Workshop on Computing Within Limits*. LIMITS 2016. Irvine, California: ACM, 2016, 9:1–9:8. ISBN: 978-1-4503-4260-5. DOI: 10.1145/2926676.2926678.

[147] Filippo Rebecchi, Marcelo D. de Amorim, Vania Conan, Andrea Passarella, Raffaele Bruno, and Marco Conti. "Data Offloading Techniques in Cellular Networks: A Survey." In: *IEEE Communications Surveys Tutorials* 17.2 (Apr. 2015), pp. 580–603. DOI: 10.1109/COMST.2014.2369742.

[148] Guillaume Rémy, Sidi-Mohammed Senouci, François Jan, and Yvon Gourhant. "LTE4V2XCollection, dissemination and multi-hop forwarding." In: *Proc. of International Conference on Communications (ICC)*. IEEE. 2012, pp. 120–125.

[149] Björn Richerzhagen, Dominik Stingl, Ronny Hans, Christian Gross, and Ralf Steinmetz. "Bypassing the cloud: Peer-assisted event dissemination for augmented reality games." In: *Proc. of International Conference on Peer-to-Peer Computing (P2P)*. Sept. 2014, pp. 1–10. DOI: 10.1109/P2P.2014.6934296.

[150] Björn Richerzhagen, Dominik Stingl, Julius Rückert, and Ralf Steinmetz. "Simonstrator: Simulation and Prototyping Platform for Distributed Mobile Applications." In: *Proc. of International Conference on Simulation Tools and Techniques (SIMUTools)*. SIMUTools '15. Athens, Greece: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2015, pp. 99–108. ISBN: 978-1-63190-079-2. DOI: 10.4108/eai.24-8-2015.2261064.

[151] Nils Richerzhagen, Björn Richerzhagen, Boris Koldehofe, and Ralf Steinmetz. "Towards Transitions between Role Assignment Schemes: Enabling Adaptive Offloading in Challenged Networks." In: *Proc. of Workshop on Challenged Networks (CHANTS)*. ACM, Oct. 2017, pp. 1–6. ISBN: 978-1-4503-5144-7/17/10.

[152] Nils Richerzhagen, Björn Richerzhagen, Dominik Stingl, and Ralf Steinmetz. "The human factor: A simulation environment for networked mobile social applications." In: *Proc. of International Conference on Networked Systems (NetSys)*. Mar. 2017, pp. 1–8. DOI: 10.1109/NetSys.2017.7903955.

[153] Nils Richerzhagen, Björn Richerzhagen, Michael Walter, Dominik Stingl, and Ralf Steinmetz. "Buddies, not Enemies: Fairness and Performance in Cellular Offloading." In: *Proc. of International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, June 2016, pp. 1–9. ISBN: 978-1-5090-2185-7/16.

[154] George F. Riley and Thomas R. Henderson. "The ns-3 network simulator." In: *Modeling and tools for network simulation*. Springer, 2010, pp. 15–34.

[155] Tobias Rückelt, Daniel Burgstahler, Frank Englert, Christian Gottron, Sebastian Zöller, and Ralf Steinmetz. "A concept for vehicle internet connectivity for non-safety applications." In: *Proc. of Local Computer Networks Workshops (LCN)*. IEEE. 2014, pp. 718–721.

[156] Tobias Rückelt, Ioannis Stavrakakis, Tobias Meuser, Imane Horiya Brahmi, Doreen Böhnstedt, and Ralf Steinmetz. "Data transmission plan adaptation complementing strategic time-network selection for connected vehicles." In: *Ad Hoc Networks* 82 (2019), pp. 146–154. ISSN: 1570-8705. DOI: `https://doi.org/10.1016/j.adhoc.2018.08.006`.

[157] César A. Santiváñez, Ram Ramanathan, and Ioannis Stavrakakis. "Making Link-state Routing Scale for Ad Hoc Networks." In: *Proc. of International Symposium on Mobile Ad Hoc Networking &Amp; Computing*. MobiHoc '01. Long Beach, CA, USA: ACM, 2001, pp. 22–32. ISBN: 1-58113-428-2. DOI: `10.1145/501417.501420`.

[158] Sophie Schönherr. "Hybrid Communication in Vehicular Networks: Combining the Advantages of Cellular and V2V Communication." Master Thesis. TU Darmstadt, 2017.

[159] Ramon S. Schwartz, Rafael R. R. Barbosa, Nirvana Meratnia, Geert Heijenk, and Hans Scholten. "A simple and robust dissemination protocol for VANETs." In: *Proc. of European Wireless Conference (EW)*. IEEE. 2010, pp. 214–222.

[160] Sayandeep Sen, Syed Gilani, Shreesha Srinath, Stephen Schmitt, and Suman Banerjee. "Design and Implementation of an "Approximate" Communication System for Wireless Media Applications." In: *Proc. of SIGCOMM*. SIGCOMM '10. New Delhi, India: ACM, 2010, pp. 15–26. ISBN: 978-1-4503-0201-2. DOI: `10.1145/1851182.1851187`.

[161] Mihail L. Sichitiu and Maria Kihl. "Inter-vehicle communication systems: a survey." In: *IEEE Communications Surveys & Tutorials* 10.2 (2008), pp. 88–105.

[162] Jabari Stegall and Alexander Wyglinski. "Secure Distributed Anonymous Data Collection for Vehicular Ad-Hoc Networks." In: *Proc. of Vehicular Technology Conference (VTC-Fall)*. Aug. 2018, pp. 1–5. DOI: `10.1109/VTCFall.2018.8690568`.

[163] Ralf Steinmetz and Klaus Wehrle. *Peer-to-Peer Systems and Applications*. Vol. 3485. Springer, 2005.

[164] Dominik Stingl, Christian Gross, Julius Rückert, Leonhard Nobach, Aleksandra Kovacevic, and Ralf Steinmetz. "PeerfactSim.KOM: A simulation framework for Peer-to-Peer systems." In: *Proc. of International Conference on High Performance Computing Simulation (HPCS)*. July 2011, pp. 577–584. DOI: `10.1109/HPCSim.2011.5999877`.

[165] Sundar Subramanian, Marc Werner, Shihuan Liu, Jubin Jose, Radu Lupoaie, and Xinzhou Wu. "Congestion Control for Vehicular Safety: Synchronous and Asynchronous MAC Algorithms." In: *Proc. of International Workshop on Vehicular Internetworking, Systems, and Applications (VANET)*. VANET '12. Low Wood Bay, Lake District, UK: ACM, 2012, pp. 63–72. ISBN: 978-1-4503-1317-9. DOI: `10.1145/2307888.2307900`.

[166] Saif Al-Sultan, Moath M. Al-Doori, Ali H. Al-Bayatti, and Hussien Zedan. "A comprehensive survey on vehicular Ad Hoc network." In: *Journal of Network and Computer Applications* 37 (2014), pp. 380–392. ISSN: 1084-8045. DOI: `https://doi.org/10.1016/j.jnca.2013.02.036`.

[167] Saif Al-Sultan, Moath M. Al-Doori, Ali H. Al-Bayatti, and Hussien Zedan. "A comprehensive survey on vehicular ad hoc network." In: *Journal of network and computer applications* 37 (2014), pp. 380–392.

[168] Tarik Tabet, Farouk Belghoul, Dawei Zhang, and Huarui Liang. *Wi-Fi signaling by cellular devices for coexistence in unlicensed frequency bands*. US Patent 9,549,080. Jan. 2017.

[169] Mohsen Nader Tehrani, Murat Uysal, and Halim Yanikomeroglu. "Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions." In: *IEEE Communications Magazine* 52.5 (2014), pp. 86–92.

[170] Ozan K. Tonguz, Nawaporn Wisitpongphan, and Fan Bai. "DV-CAST: A distributed vehicular broadcast protocol for vehicular ad hoc networks." In: *IEEE Wireless Communications* 17.2 (2010), pp. 47–57.

[171] Jamal Toutouh, José Garca-Nieto, and Enrique Alba. "Intelligent OLSR routing protocol optimization for VANETs." In: *IEEE transactions on vehicular technology* 61.4 (2012), pp. 1884–1894.

[172] Hamed Vahdat-Nejad, Azam Ramazani, Tahereh Mohammadi, and Wathiq Mansoor. "A survey on context-aware vehicular network applications." In: *Vehicular Communications* 3 (2016), pp. 43–57. ISSN: 2214-2096. DOI: `https://doi.org/10.1016/j.vehcom.2016.01.002`.

[173] Andrea Vesco, Riccardo Scopigno, Claudio Casetti, and Carla-Fabiana Chiasserini. "Investigating the effectiveness of decentralized congestion control in vehicular networks." In: *2013 IEEE Globecom Workshops*. Dec. 2013, pp. 1314–1319. DOI: `10.1109/GLOCOMW.2013.6825176`.

[174] Alexey Vinel. "3GPP LTE Versus IEEE 802.11p/WAVE: Which Technology is Able to Support Cooperative Vehicular Safety Applications?" In: *IEEE Wireless Communications Letters* 1.2 (Apr. 2012), pp. 125–128. DOI: `10.1109/WCL.2012.022012.120073`.

[175] Omar A. Wahab, Hadi Otrok, and Azzam Mourad. "VANET QoS-OLSR: QoS-based clustering protocol for vehicular ad hoc networks." In: *Computer Communications* 36.13 (2013), pp. 1422–1435.

[176] Hongjian Wang, Yanmin Zhu, and Qian Zhang. "Compressive sensing based monitoring with vehicular networks." In: *Proc. of International Conference on Computer Communications (INFOCOM)*. Apr. 2013, pp. 2823–2831. DOI: `10.1109/INFCOM.2013.6567092`.

[177] Lucas Wang, Alexander Afanasyev, Romain Kuntz, Rama Vuyyuru, Ryuji Wakikawa, and Lixia Zhang. "Rapid Traffic Information Dissemination Using Named Data." In: *Proc. of Workshop on Emerging Name-Oriented Mobile Networking Design - Architecture, Algorithms, and Applications*. NoM '12. Hilton Head, South Carolina, USA: ACM, 2012, pp. 7–12. ISBN: 978-1-4503-1291-2. DOI: `10.1145/2248361.2248365`.

[178] Lucas Wang, Ryuji Wakikawa, Romain Kuntz, Rama Vuyyuru, and Lixia Zhang. "Data naming in vehicle-to-vehicle communications." In: *Proc. of International Conference on Computer Communications (INFOCOM) Workshops*. IEEE. 2012, pp. 328–333.

[179] Yi Wang, Akram Ahmed, Bhaskar Krishnamachari, and Konstantinos Psounis. "IEEE 802.11p performance evaluation and protocol enhancement." In: *Proc. of International Conference on Vehicular Electronics and Safety (ICVES)*. Sept. 2008, pp. 317–322. DOI: 10.1109/ICVES.2008.4640898.

[180] Yu Wang and Fan Li. "Vehicular Ad Hoc Networks." In: *Guide to Wireless Ad Hoc Networks*. Ed. by Sudip Misra, Isaac Woungang, and Subhas Chandra Misra. London: Springer London, 2009, pp. 503–525. ISBN: 978-1-84800-328-6. DOI: 10.1007/978-1-84800-328-6_20.

[181] Waymo. *Building maps for a self-driving car*. 2016. URL: https://medium.com/waymo/building-maps-for-a-self-driving-car-723b4d9cd3f4.

[182] Jan W. Wedel, Björn Schünemann, and Ilja Radusch. "V2X-Based Traffic Congestion Recognition and Avoidance." In: *Proc. of International Symposium on Pervasive Systems, Algorithms, and Networks (I-SPAN)*. Dec. 2009, pp. 637–641. DOI: 10.1109/I-SPAN.2009.71.

[183] Axel Wegener, Micha Piórkowski, Maxim Raya, Horst Hellbrück, Stefan Fischer, and Jean-Pierre Hubaux. "TraCI: An Interface for Coupling Road Traffic and Network Simulators." In: *Proc. of Communications and Networking Simulation Symposium (CNS)*. CNS '08. Ottawa, Canada: ACM, 2008, pp. 155–163. ISBN: 1-56555-318-7. DOI: 10.1145/1400713.1400740.

[184] Hao Wu, Richard Fujimoto, Randall Guensler, and Michael Hunter. "MDDV: a mobility-centric data dissemination algorithm for vehicular networks." In: *Proc. of International Workshop on Vehicular Ad Hoc Networks (VANET)*. ACM. 2004, pp. 47–56.

[185] Yue Wu, Weisi Guo, Hu Yuan, Long Li, Siyi Wang, Xiaoli Chu, and Jie Zhang. "Device-to-device meets LTE-unlicensed." In: *IEEE Communications Magazine* 54.5 (2016), pp. 154–159.

[186] Christos Xenakis, Christoforos Panos, and Ioannis Stavrakakis. "A comparative evaluation of intrusion detection architectures for mobile ad hoc networks." In: *Computers & Security* 30.1 (2011), pp. 63–80. ISSN: 0167-4048. DOI: https://doi.org/10.1016/j.cose.2010.10.008.

[187] Yong Xiang, Zheng Liu, Ruilin Liu, Weizhen Sun, and Wei Wang. "GeoSVR: A map-based stateless VANET routing." In: *Ad Hoc Networks* 11.7 (2013), pp. 2125–2135.

[188] Shouzhi Xu, Pengfei Guo, Bo Xu, and Huan Zhou. "QoS evaluation of VANET routing protocols." In: *Journal of Networks* 8.1 (2013), p. 132.

[189] Fangchun Yang, Shangguang Wang, Jinglin Li, Zhihan Liu, and Qibo Sun. "An overview of internet of vehicles." In: *China communications* 11.10 (2014), pp. 1–15.

[190] Shuo Yang, Fan Wu, Shaojie Tang, Xiaofeng Gao, Bo Yang, and Guihai Chen. "On Designing Data Quality-Aware Truth Estimation and Surplus Sharing Method for Mobile Crowdsensing." In: *IEEE Journal on Selected Areas in Communications* 35.4 (Apr. 2017), pp. 832–847. DOI: 10.1109/JSAC.2017.2676898.

[191] Chia-Hao Yu, Klaus Doppler, Cassio B. Ribeiro, and Olav Tirkkonen. "Resource sharing optimization for device-to-device communication underlaying cellular networks." In: *IEEE Transactions on Wireless communications* 10.8 (2011), pp. 2752–2763.

[192] Rong Yu, Yan Zhang, Stein Gjessing, Wenlong Xia, and Kun Yang. "Toward cloud-based vehicular networks with efficient resource management." In: *arXiv preprint arXiv:1308.6208* (2013).

[193] Yu-Ting Yu, Mario Gerla, and MY Sanadidi. "Scalable VANET content routing using hierarchical bloom filters." In: *Wireless Communications and Mobile Computing* 15.6 (2015), pp. 1001–1014.

[194] Xiaoxiao Yu, H. Zhao, Lin Zhang, S. Wu, Bhaskar Krishnamachari, and Victor O. K. Li. "Cooperative Sensing and Compression in Vehicular Sensor Networks for Urban Monitoring." In: *Proc. of International Conference on Communications (ICC)*. May 2010, pp. 1–5. DOI: 10.1109/ICC.2010.5502562.

[195] Ran Zhang, Miao Wang, Lin X. Cai, Zhongming Zheng, Xuemin Shen, and Liang-Liang Xie. "LTE-unlicensed: the future of spectrum aggregation for cellular networks." In: *IEEE Wireless Communications* 22.3 (2015), pp. 150–159.

[196] Yan Zhang, Jim Mee Ng, and Chor Ping Low. "A distributed group mobility adaptive clustering algorithm for mobile ad hoc networks." In: *Computer Communications* 32.1 (2009), pp. 189–202.

[197] Yang Zhang, Jing Zhao, and Guohong Cao. "On scheduling vehicle-roadside data access." In: *Proc. of International Workshop on Vehicular Ad Hoc Networks (VANET)*. ACM. 2007, pp. 9–18.

[198] Cong Zhao, Shusen Yang, Ping Yan, Qing Yang, Xinyu Yang, and Julie McCann. "Data Quality Guarantee for Credible Caching Device Selection in Mobile Crowdsensing Systems." In: *IEEE Wireless Communications* 25.3 (June 2018), pp. 58–64. DOI: 10.1109/MWC.2018.1700299.

[199] Jing Zhao and Guohong Cao. "VADD: Vehicle-assisted data delivery in vehicular ad hoc networks." In: *IEEE transactions on vehicular technology* 57.3 (2008), pp. 1910–1922.

[200] Ghayet el mouna Zhioua, Houda Labiod, Tabbane Nabil, and Sami Tabbane. "An efficient QoS based gateway selection algorithm for VANET to LTE advanced hybrid cellular network." In: *Proc. of International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*. ACM. 2012, pp. 353–356.

[201] Ghayet el mouna Zhioua, Nabil Tabbane, Houda Labiod, and Sami Tabbane. "A Fuzzy Multi-Metric QoS-Balancing Gateway Selection Algorithm in a Clustered VANET to LTE Advanced Hybrid Cellular Network." In: *IEEE Transactions on Vehicular Technology* 64.2 (Feb. 2015), pp. 804–817. DOI: 10.1109/TVT.2014.2323693.

[202] Bin Zhou, Honglin Hu, Sheng-Qiang Huang, and Hsiao-Hwa Chen. "Intracluster device-to-device relay algorithm with optimal resource utilization." In: *IEEE transactions on vehicular technology* 62.5 (2013), pp. 2315–2326.

[203] L. Zhou, G. Cui, H. Liu, Z. Wu, and D. Luo. "NPPB: A broadcast scheme in dense VANETs." In: *Information Technology Journal* 9.2 (2010), pp. 247–256.

[204] Julian Zobel, Patrick Lieser, Bastian Drescher, Bernd Freisleben, and Ralf Steinmetz. "Optimizing Inter-Cluster Flights of Post-Disaster Communication Support UAVs." In: *Proc. of Conference on Local Computer Networks (LCN)*. IEEE. Oct. 2019.

All web pages cited in this work have been checked in October 2019. However, due to the dynamic nature of the World Wide Web, their long-term availability cannot be guaranteed.

A APPENDIX

A.1 Transformation to obtain the weight function for the adaptive aggregation scheme

In the following, we provide the derivation of the weight function using Equation 15 from Section 4.2.3 and the two side conditions shown in Equation 77 and Equation 78. For better readability, we copied Equation 15 to Equation 76.

$$f_w(t) = a \cdot e^{bt} + d \tag{76}$$

$$f_w(0) = a + d = 1 \Rightarrow a = 1 - d \tag{77}$$

$$f_w(T) = a \cdot e^{bT} + d = 0 \tag{78}$$

When we insert $a = 1 - d$ from Equation 77 into Equation 76, we obtain Equation 79.

$$f_w(t) = (1 - d) \cdot e^{bt} + d \tag{79}$$

This also changes the second side condition as shown in Equation 78, which leads to Equation 80.

$$f_w(T) = (1 - d) \cdot e^{bT} + d = 0 \Rightarrow d = -\left(\frac{e^{bT}}{1 - e^{bT}}\right) \tag{80}$$

When we now insert Equation 80 into Equation 79, we obtain Equation 81.

$$\begin{aligned} f_w(t) &= \left(1 - \left[-\left(\frac{e^{bT}}{1 - e^{bT}}\right)\right]\right) \cdot e^{bt} - \left(\frac{e^{bT}}{1 - e^{bT}}\right) \\ &= e^{bt} + \frac{e^{bt} \cdot e^{bT}}{1 - e^{bT}} - \frac{e^{bT}}{1 - e^{bT}} = \frac{e^{bt} - e^{bt} \cdot e^{bT}}{1 - e^{bT}} + \frac{e^{bt} \cdot e^{bT}}{1 - e^{bT}} - \frac{e^{bT}}{1 - e^{bT}} \end{aligned} \tag{81}$$

From Equation 81, Equation 82 can be derived directly, which completes the derivation of the modified weight function.

$$f_w(t) = \frac{e^{bt} - e^{bT}}{1 - e^{bT}} \tag{82}$$

A.2 TRANSFORMATIONS TO OBTAIN A SOLUTION FOR THE APPROXIMATE VEHICULAR NETWORK

We provide the necessary transformations to obtain Equation 62 based on Equation 61 from Section 5.3. For better readability, we copied Equation 61 to Equation 83.

$$1 - p_{\phi_t,l} = \left[1 - \left(\frac{A}{a_{\phi_t,m}} - \sum_{i=1 \mid i \neq m \wedge \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \left[\Lambda_i \left(\frac{p_{\phi_t,l} - 1}{\Lambda_l} \right) + 1 \right]}{a_{\phi_t,m}} \right) \right] \cdot \Lambda_l \tag{83}$$

This equation can be transformed by resolving the brackets step by step as shown in Equation 84.

$$\begin{aligned}
&1 - p_{\phi_t,l} \\
&= \Lambda_l - \Lambda_l \cdot \frac{A}{a_{\phi_t,m}} + \Lambda_l \cdot \sum_{i=1 \mid i \neq m \wedge \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \left[\Lambda_i \left(\frac{p_{\phi_t,l} - 1}{\Lambda_l} \right) + 1 \right]}{a_{\phi_t,m}} \\
&= \Lambda_l - \Lambda_l \cdot \frac{A}{a_{\phi_t,m}} \\
&+ \sum_{i=1 \mid i \neq m \wedge \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \Lambda_i \cdot p_{\phi_t,l} - a_{\phi_t,i} \cdot \Lambda_i + a_{\phi_t,i} \cdot \Lambda_l}{a_{\phi_t,m}} \\
&= \Lambda_l - \Lambda_l \cdot \frac{A}{a_{\phi_t,m}} + \sum_{i=1 \mid i \neq m \wedge \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \Lambda_i \cdot p_{\phi_t,l}}{a_{\phi_t,m}} \\
&\quad - \sum_{i=1 \mid i \neq m \wedge \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \Lambda_i}{a_{\phi_t,m}} + \sum_{i=1 \mid i \neq m \wedge \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \Lambda_l}{a_{\phi_t,m}}
\end{aligned} \tag{84}$$

Additionally, we move all parts of the equation that contain $p_{\phi_t,l}$ to the left side.

$$\begin{aligned}
&\left(\sum_{i=1 \mid i \neq m \wedge \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \Lambda_i}{a_{\phi_t,m}} + 1 \right) \cdot p_{\phi_t,l} = \Lambda_l \cdot \frac{A}{a_{\phi_t,m}} \\
&\quad + \sum_{i=1 \mid i \neq m \wedge \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \Lambda_i}{a_{\phi_t,m}} - \sum_{i=1 \mid \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \Lambda_l}{a_{\phi_t,m}} + 1
\end{aligned} \tag{85}$$

An important aspect of this simplification is the fact that $\Lambda_m = 1$, which is justified by the fact that m is similar for all privacy-levels and that Λ_m captures the relation from m to itself. When we now replace some of 1 with Λ_m, we can drop the requirement that $i \neq m$ in the sums.

$$\sum_{i=1 \mid \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \Lambda_i}{a_{\phi_t,m}} \cdot p_{\phi_t,l} = \Lambda_l \cdot \frac{A}{a_{\phi_t,m}} + \sum_{i=1 \mid \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \Lambda_i}{a_{\phi_t,m}} - \sum_{i=1 \mid \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \Lambda_l}{a_{\phi_t,m}} \tag{86}$$

We then divide by the sum to separate $p_{\phi_t,l}$.

$$p_{\phi_t,l} = \frac{\Lambda_l \cdot \frac{A}{a_{\phi_t,m}} + \sum_{i=1 \mid \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \Lambda_i}{a_{\phi_t,m}} - \sum_{i=1 \mid \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \Lambda_l}{a_{\phi_t,m}}}{\sum_{i=1 \mid \phi_t \notin \Phi^-(i)}^{n_\mu} \frac{a_{\phi_t,i} \cdot \Lambda_i}{a_{\phi_t,m}}} \tag{87}$$

With canceling $a_{\phi_t,m}$ and other minor transformations, this then leads to

$$p_{\phi_t,l} = \frac{\left(A - \sum_{i=1 \mid \phi_t \notin \Phi^-(i)}^{n_\mu} a_{\phi_t,i}\right) \cdot \Lambda_l}{\sum_{i=1 \mid \phi_t \notin \Phi^-(i)}^{n_\mu} a_{\phi_t,i} \cdot \Lambda_i} + 1 \tag{88}$$

This equation is similar to Equation 62 from Section 5.3.

A.3 ADDITIONAL INSIGHTS ON THE INFLUENCE OF SENSOR HETEROGENEITY

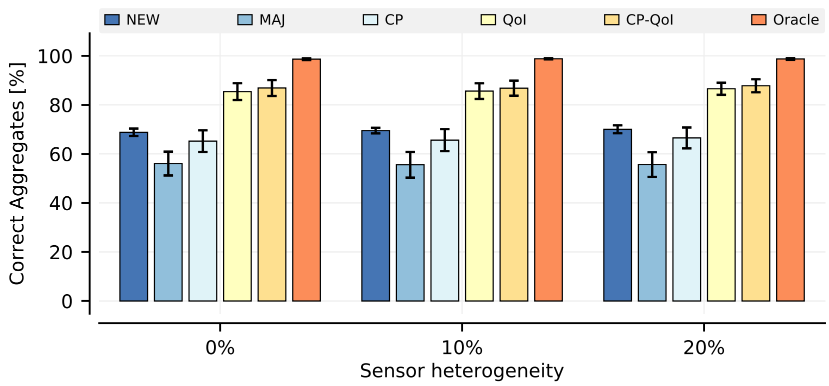
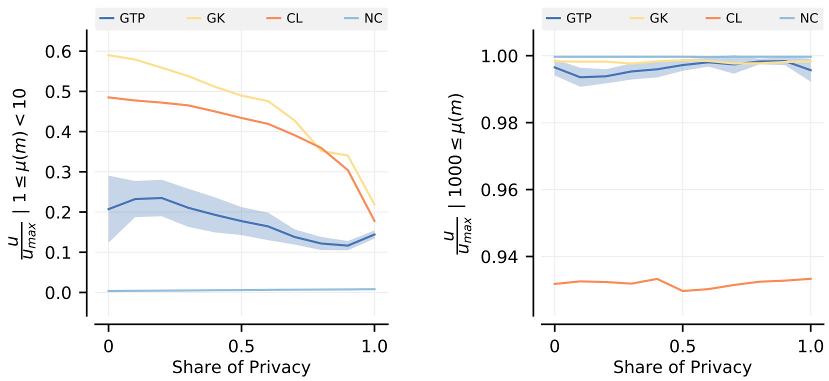


Figure 29: Influence of the sensor heterogeneity on the quality of the aggregates for the different approaches.

In addition to the average sensor accuracy, the heterogeneity of sensors might influence the performance of the approaches. The overall performance depending on the sensor heterogeneity is shown in Figure 29. We can observe that the influence of sensor heterogeneity is only marginal regarding the share of correct aggregates. That is, as a high number of aggregates partially compensates for the heterogeneity of sensors, which is the reason for the marginal influence on the performance of the approaches. There is a minor performance increase for the CP approach and the CP-QoI approach, but this increase is hardly visible in the figure. Thus, we can observe that the heterogeneity of sensors is well compensated for by the presented approaches.

A.4 ADDITIONAL INSIGHTS INTO THE INFLUENCE OF LOCATION PRIVACY



(a) Base-impact per bit $\mu_B(m) = 1$. (b) Base-impact per bit $\mu_B(m) = 1000$.

Figure 30: Influence of location privacy on the utility of messages with different base-impact.

Figure 30 describes the utility of the impact-levels with the lowest and the highest base-impact, depending on the share of privacy. In Figure 30a, the performance of the privacy-level with the lowest base-impact is displayed. While the *NC* approach cannot receive messages from this impact-level, the cooperative approaches can all receive messages of this impact-level. The cluster-based approaches are very efficient in the utilization of bandwidth and, thus, can often receive data associated with this impact-level. With an increasing privacy-level, the relative utility decreases constantly, as the messages associated with this impact-level are context-sensitive (see Section 7.4.1). Interestingly, the relative utility of our *GTP* approach has a completely different behavior, which is caused by two factors: (i) the possibilities of implicit coordination between privacy-levels allows for a more efficient bandwidth utilization (explaining the first increase in utility at a low share of privacy), and (ii) this impact-level is more efficient than the impact-level with a base-impact of 10 if received by a privacy-sensitive vehicle (explaining the second increase in utility at a high share of privacy). In contrast, the relative utility for the impact-level with the highest base-impact, shown in Figure 30b is only marginally affected by the share of privacy. This is caused by the low context-sensitivity of these messages and the high base-impact, which leads to a prioritization of these messages independent of the privacy-level.

A.5 LIST OF ACRONYMS

CAM	Cooperative Awareness Message
DCC	Decentralized Congestion Control
DCF	Distributed Coordination Function
DEN	Decentralized Environment Notification
DENM	Decentralized Environment Notification Message
DPT	Diverse Prioritization And Treatment
DSRC	Dedicated Short-Range Communication
ETSI	European Telecommunications Standards Institute
FDD	Frequency Division Duplex
HSPA	High Speed Packet Access
IoV	Internet Of Vehicles
LSB	Least Significant Bit
LTE	Long Term Evolution
LTE-A	Long Term Evolution-Advanced
LTE-D2D	LTE-Device-to-Device
M2M	Machine-to-Machine
MANET	Mobile Ad-hoc Network
MPTCP	Multipath TCP
MSB	Most Significant Bit
P2P	Peer-to-Peer
Pub/Sub	Publish/Subscribe
QoS	Quality Of Service
QUIC	Quick UDP Internet Connections
RSU	Road Side Unit
SUMO	Simulation Of Urban Mobility
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TMC	Traffic Message Channel
TraCI	Traffic Control Interface
TTL	Time To Live
UEP	Unequal Error Protection
UMTS	Universal Mobile Telecommunications System
V2V	Vehicle To Vehicle
VANET	Vehicular Ad-hoc Network

A.6 SUPERVISED STUDENT THESES

[1] Jan-Thomas Becker. “Positioning with Sensor Inaccuracy: A Robust Walking Navigation for Last Mile Logistics.” Master Thesis. TU Darmstadt, 2019.

[2] Yahia Brini. “Utilizing Machine Learning for Efficient Information Dissemination in Vehicular Networks.” Bachelor Thesis. TU Darmstadt, 2018.

[3] Tobias Brudermüller. “Interpreting Feature Spaces of Neural Networks for Vehicular Perception.” Bachelor Thesis. TU Darmstadt, 2018.

[4] Björn Büschke. “Using Crowd-sensed Data for Predictive Information-Centric Vehicle Feature Enhancement.” Master Thesis. TU Darmstadt, 2018.

[5] Sven Dotzauer-Klier. “Scalability Analysis of Context-Aware Publish/Subscribe via MQTT.” Bachelor Thesis. TU Darmstadt, 2019.

[6] Lukas Heidt. “Heterogeneous Communication in Vehicular Networks: Influence Factors on the Performance of 802.11p and LTE-A.” Student Research Project. TU Darmstadt, 2018.

[7] Mathias Hornjak. “Analyse und Einsatz von IoT-Technologien im Parkraummanagement.” Master Thesis. TU Darmstadt, 2019.

[8] Deepak Jayaram. “Cooperative Caching in Vehicular Networks: Distributed Cache Invalidation using Information Freshness.” Master Thesis. TU Darmstadt, 2018.

[9] Li Ming. “Information Quality in Mobile Ad-Hoc Networks: Development of a Model for Information Relevancy Evaluation.” Master Thesis. TU Darmstadt, 2017.

[10] Martin Möbus. “Self-Learning Location-based Information to Support Autonomous Vehicle Drive-Planning in Urban Environments.” Master Thesis. TU Darmstadt, 2017.

[11] Sophie Schönherr. “Hybrid Communication in Vehicular Networks: Combining the Advantages of Cellular and V2V Communication.” Master Thesis. TU Darmstadt, 2017.

[12] Torben Unzicker. “Assessing Information Quality for Efficient Information Dissemination in Vehicular Networks.” Master Thesis. TU Darmstadt, 2019.

[13] Marcel Verst. “Design of a Disaster Mobility Model based on real-world Movement Traces.” Master Thesis. TU Darmstadt, 2018.

[14] Tim Walter. “Modell basiertes Eco-Routing: Entwicklung eines Modells zur Bestimmung des Energiebedarfs von Fahrzeugrouten.” Bachelor Thesis. TU Darmstadt, 2017.

[15] Martin Wende. “Data Validity Model: Evaluation Of Information Freshness In Mobile Ad-Hoc Networks.” Bachelor Thesis. TU Darmstadt, 2017.

B AUTHOR’S PUBLICATIONS

MAIN PUBLICATIONS

[1] Flor Álvarez, Lars Almon, Patrick Lieser, Tobias Meuser, Yannick Dylla, Björn Richerzhagen, Matthias Hollick, and Ralf Steinmetz. “Conducting a Large-scale Field Test of a Smartphone-based Communication Network for Emergency Response.” In: *Proc. of Workshop on Challenged Networks (CHANTS)*. ACM, Nov. 2018, pp. 3–10. DOI: `10.1145/3264844.3264845`.

[2] Tobias Meuser, Daniel Bischoff, Ralf Steinmetz, and Björn Richerzhagen. “Simulation Platform for Connected Heterogeneous Vehicles.” In: *Proc. of International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS)*. SCITEPRESS, May 2019, pp. 412–419. ISBN: 978-989-758-374-2.

[3] Tobias Meuser, Daniel Burgstahler, Tobias Rueckelt, Doreen Boehnstedt, and Ralf Steinmetz. “Hybrid-ProbSense.KOM: Probabilistic Sensing with Hybrid Communication for Gathering Vehicular Sensed Data.” In: *Proc. of Automotive meets Electronics (AmE)*. Vol. 87. Mar. 2017, pp. 1–6. ISBN: 978-3-8007-4369-8.

[4] Tobias Meuser, Patrick Lieser, The An Binh Nguyen, Doreen Böhnstedt, and Ralf Steinmetz. “Adaptive Information Aggregation for Application-specific Demands.” In: *Proc. of International Balkan Conference on Communications and Networking (BalkanCom)*. May 2017, pp. 1–6.

[5] Tobias Meuser, Björn Richerzhagen, Ioannis Stavrakakis, The An Binh Nguyen, and Ralf Steinmetz. “Relevance-Aware Information Dissemination in Vehicular Networks.” In: *Proc. of International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, June 2018, pp. 588–590. ISBN: 978-1-5386-4725-7. DOI: `10.1109/WoWMoM.2018.8449751`.

[6] Tobias Meuser, Ioannis Stavrakakis, and Antonio Fernández Anta. “Dynamic Vehicle Path-Planning in the Presence of Traffic Events.” In: *Proc. of Conference on Local Computer Networks (LCN)*. IEEE, Oct. 2019, pp. 1–4.

[7] Tobias Meuser, Martin Wende, Patrick Lieser, Björn Richerzhagen, and Ralf Steinmetz. “Adaptive Decision Making Based on Temporal Information Dynamics.” In: *Proc. of International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS)*. SCITEPRESS, Mar. 2018, pp. 91–102. ISBN: 978-989-758-293-6.

[8] Tobias Meuser, Martin Wende, Patrick Lieser, Björn Richerzhagen, and Ralf Steinmetz. “Distributed Quality of Information-Aware Decision-Making in Vehicular Networks.” In: *Smart Cities, Green Technologies and Intelligent Transport Systems*. Ed. by Brian Donnellan, Cornel Klein, Markus Helfert, and Oleg

Gusikhin. Cham: Springer International Publishing, 2019, pp. 304–327. ISBN: 978-3-030-26633-2.

[9] Tobias Meuser, Daniel Bischoff, Björn Richerzhagen, and Ralf Steinmetz. "Cooperative Offloading in Context-Aware Networks: A Game-Theoretic Approach." In: *Proc. of International Conference on Distributed and Event-based Systems (DEBS)*. ACM, June 2019, pp. 55–66. ISBN: 978-1-4503-6794-3. DOI: 10.1145/3328905.3338535.

[10] Tobias Meuser, Daniel Bischoff, and Ralf Steinmetz. "Location Privacy in Heterogeneous Vehicular Networks." In: *Proc. of International Conference on Distributed and Event-based Systems (DEBS)*. ACM, 2019, pp. 260–261. ISBN: 978-1-4503-6794-3. DOI: 10.1145/3328905.3332515.

[11] Tobias Meuser, Patrick Lieser, The An Binh Nguyen, Doreen Böhnstedt, and Ralf Steinmetz. "Lifetime-aware information aggregation under uncertainty for advanced driver assistance systems." In: *Ad Hoc Networks* 82 (2019), pp. 56–67. ISSN: 1570-8705. DOI: https://doi.org/10.1016/j.adhoc.2018.07.018.

CO-AUTHORED PUBLICATIONS

[12] Lars Almon, Flor Alvarez, Patrick Lieser, Tobias Meuser, and Fabian Schaller. "Ad-Hoc-Kommunikation - Gesellschaftlich wünschenswert, rechtlich ungeregelt." In: *Die Fortentwicklung des Datenschutzes - Zwischen Systemgestaltung und Selbstregulierung*. Ed. by Prof. Marit Hansen Prof. Dr. Alexander Roßnagel Dr. Michael Friedewald. Springer Fachmedien Wiesbaden, Oct. 2018, pp. 77–98. ISBN: 978-3-658-23726-4.

[13] Lars Baumgärtner, Jonas Höchst, and Tobias Meuser. "B-DTN7: Browser-based Disruption-tolerant Networking via Bundle Protocol 7." In: *Proc. of International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*. IEEE, Dec. 2019, pp. 1–8.

[14] Daniel Bischoff, Harald Berninger, Steffen Knapp, Tobias Meuser, Björn Richerzhagen, Lars Häring, and Andreas Czylwik. "Safety-relevant V2X beaconing in realistic and scalable heterogeneous radio propagation fading channels." In: *Proc. of International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS)*. SCITEPRESS, May 2019, pp. 404–411. ISBN: 978-989-758-374-2.

[15] Daniel Bischoff and Tobias Meuser. "Geo-based Backend Dissemination for Safety-relevant V2X Applications." In: *Proc. of Automotive meets Electronics (AmE)*. VDE Verlag, Mar. 2019, pp. 1–6. ISBN: 978-3-8007-4877-8.

[16] Daniel Burgstahler, Tobias Meuser, Ulrich Lampe, Doreen Böhnstedt, and Ralf Steinmetz. "ProbSense.KOM: A Probabilistic Sensing Approach for Gathering Vehicular Sensed Data." In: *Proc. of International Conference on Mobile Services (MS)*. Washington, DC, USA: IEEE Computer Society, June 2016, pp. 9–16. ISBN: 978-1-5090-2625-8.

[17] Daniel Burgstahler, Martin Möbus, Tobias Meuser, Doreen Böhnstedt, and Ralf Steinmetz. "A Categorization Scheme for Information Demands of Future Connected ADAS." In: *Proc. of Automotive meets Electronics (AmE)*. Vol. 87. Berlin: VDE Verlag, Mar. 2017, pp. 32–37. ISBN: 978-3-8007-4369-8.

[18] Florian Jomrich, Daniel Bischoff, Steffen Knapp, Tobias Meuser, Björn Richerzhagen, and Ralf Steinmetz. "Lane Accurate Detection of Map Changes based on Low Cost Smartphone Data." In: SCITEPRESS, May 2019, pp. 126–137. ISBN: 978-989-758-374-2.

[19] Florian Jomrich, Florian Fischer, Steffen Knapp, Tobias Meuser, Björn Richerzhagen, and Ralf Steinmetz. "Enhanced Cellular Bandwidth Prediction for Highly Automated Driving." In: *Smart Cities, Green Technologies and Intelligent Transport Systems*. Ed. by Brian Donnellan, Cornel Klein, Markus Helfert, and Oleg Gusikhin. Cham: Springer International Publishing, 2019, pp. 328–350. ISBN: 978-3-030-26633-2.

[20] Florian Jomrich, Markus Grau, Tobias Meuser, The An Binh Nguyen, Doreen Böhnstedt, and Ralf Steinmetz. "ICCOMQS - Intelligent measuring framework to ensure reliable communication for highly automated vehicles." In: *Proc. of Vehicular Networking Conference (VNC)*. IEEE, Nov. 2017, pp. 311–318. ISBN: 978-1-5386-0985-9.

[21] Florian Jomrich, Alexander Herzberger, Tobias Meuser, Björn Richerzhagen, Ralf Steinmetz, and Cornelius Wille. "Cellular Bandwidth Prediction for Highly Automated Driving - Evaluation of Machine Learning Approaches based on Real-World Data." In: *Proc. of International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS)*. Ed. by Oleg Gusikhin Markus Helfert. 4. SCITEPRESS, Mar. 2018, pp. 121–131. ISBN: 978-989-758-293-6.

[22] Patrick Lieser, Nils Richerzhagen, Tim Feuerbach, Tobias Meuser, Björn Richerzhagen, and Ralf Steinmetz. "Collaborative Decentralized Resource Reservation for Emergency Communication Networks." In: *Proc. of Conference on Local Computer Networks (LCN)*. IEEE, Oct. 2018, pp. 150–158.

[23] Johannes Müller, Tobias Meuser, Ralf Steinmetz, and Michael Buchholz. "A Trust Management and Misbehaviour Detection Mechanism for Multi-Agent Systems and its Application to Intelligent Transportation Systems." In: *Proc. of International Conference on Control and Automation (ICCA)*. IEEE, June 2019, pp. 1–7.

[24] The An Binh Nguyen, Marius Rettberg-Päplow, Christian Meurisch, Tobias Meuser, Björn Richerzhagen, and Ralf Steinmetz. "Complex Services Offloading in Opportunistic Networks." In: *Proc. of IFIP Networking*. May 2018, pp. 532–540. ISBN: 978-3-903176-08-9.

[25] Tobias Rückelt, Ioannis Stavrakakis, Tobias Meuser, Doreen Böhnstedt, and Ralf Steinmetz. "Data Transmission Plan Adaptation for Connected Vehicles." In: *Proc. of International Balkan Conference on Communications and Networking (BalkanCom)*. May 2017, pp. 1–6. ISBN: 978-3-901882-94-4.

[26] Tobias Rückelt, Ioannis Stavrakakis, Tobias Meuser, Imane Horiya Brahmi, Doreen Böhnstedt, and Ralf Steinmetz. "Data transmission plan adaptation complementing strategic time-network selection for connected vehicles." In: *Ad Hoc Networks* 82 (Jan. 2019), pp. 146–154. ISSN: 1570-8705.

DEMO PAPERS

[27] Patrick Lieser, Nils Richerzhagen, Michal Lipinski, Clemens Krug, Tobias Meuser, Björn Richerzhagen, and Ralf Steinmetz. "Decentralized Disaster Area Detection in Mobile Networks." In: *Demonstrations of Conference on Local Computer Networks (LCN Demonstrations)*. IEEE, Oct. 2018, pp. 1–3.

TECHNICAL REPORTS

[28] Tobias Meuser, Oluwasegun Taiwo Ojo, Daniel Bischoff, Antonio Fernández Anta, Ioannis Stavrakakis, and Ralf Steinmetz. *Hide Me: Enabling Location Privacy in Heterogeneous Vehicular Networks*. 2020. arXiv: `2001.07170 [cs.CR]`.

ERKLÄRUNG LAUT PROMOTIONSORDNUNG

§8 Abs. 1 lit. c PromO

Ich versichere hiermit, dass die elektronische Version meiner Dissertation mit der schriftlichen Version übereinstimmt.

§8 Abs. 1 lit. d PromO

Ich versichere hiermit, dass zu einem vorherigen Zeitpunkt noch keine Promotion versucht wurde. In diesem Fall sind nähere Angaben über Zeitpunkt, Hochschule, Dissertationsthema und Ergebnis dieses Versuchs mitzuteilen.

§9 Abs. 1 PromO

Ich versichere hiermit, dass die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Quellen verfasst wurde.

§9 Abs. 2 PromO

Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 22. Oktober 2019

Tobias Meuser