Adaptive Decision Making based on Temporal Information Dynamics

Tobias Meuser, Martin Wende, Patrick Lieser, Björn Richerzhagen and Ralf Steinmetz
Multimedia Communications Lab, Technische Universitaet Darmstadt, Germany

Keywords: Decision-making, Distributed, Quality of Information, Information Lifetime, Information Accuracy.

Abstract: To increase road safety and efficiency, connected vehicles rely on the exchange of information. On each vehicle, a decision-making algorithm processes the received information and determines the actions that are to be taken. State-of-the-art decision approaches focus on static information and ignore the temporal dynamics of the environment, which is characterized by high change rates in a vehicular scenario. Hence, they keep outdated information longer than necessary and miss optimization potential. To address this problem, we propose a quality of information (QoI) weight based on a Hidden Markov Model for each information type, e.g., a road congestion state. Using this weight in the decision process allows us to combine detection accuracy of the sensor and the information lifetime in the decision-making, and, consequently, adapt to environmental changes significantly faster. We evaluate our approach for the scenario of traffic jam detection and avoidance, showing that it reduces the costs of false decisions by up to 25% compared to existing approaches.

1 INTRODUCTION

In recent years, vehicles have become increasingly connected. Consequently, an increasing number of assistance functions relies on information that is provided by other vehicles, e.g., intelligent route planning. With ongoing research towards autonomous vehicles, the amount of shared information and functions relying on this information is expected to grow.

However, as the information is sensed by other vehicles with their onboard sensors, its quality can vary significantly. Furthermore, information received from multiple vehicles can be contradicting or even wrong. In a conventional vehicle with a human driver, the driver validates and rates information intuitively and makes a decision based on prior knowledge. In comparison to that, autonomous and partly autonomous vehicles lack human intuition for information rating and decision-making. Hence, analytical methods need to be developed to make decisions in light of ambiguous or even contradictory information.

Assuming that the majority of information is correct, vehicles can use approaches that rely on majority voting with simple static thresholds (Kakkasageri and Manvi, 2014). Depending on the selected threshold, these approaches adapt either slowly making them unsuitable for dynamic conditions or fast, making them vulnerable to false information. However, as information about road and traffic conditions changes frequently—and measurements are not 100% reliable—this threshold needs to be adapted for an optimal solution under dynamic conditions.

In this work, we propose a decision-making process based on an information quality rating method that can cope with ambiguous or contradictory information. We focus on two information quality factors: the false detection rate describing the percentage of erroneous measurements and the expected event lifetime of the information type. We combine both factors using an exponential function to decide on the quality of information. To this end, the event lifetime and the false detection rate are modeled using a Hidden Markov Model (HMM). Based on the HMM, we derive a weighting function that is then used in a weighted majority voting. Consequently, information of high quality has a higher impact on the decision than low-quality information. As a result, we can drastically decrease the adaption time for information with high detection rate by lowering the impact of old information in the voting procedure.

We evaluate our approach for the scenario displayed in Figure 1. The vehicles in the Area of In-

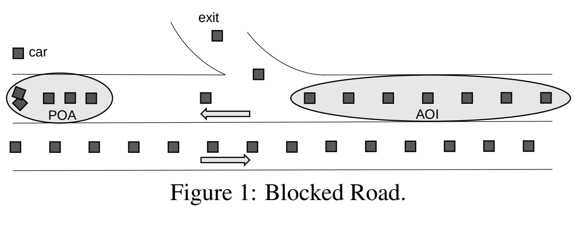


Figure 1: Blocked Road.

Meuser, T., Wende, M., Lieser, P., Richerzhagen, B. and Steinmetz, R.
Adaptive Decision Making based on Temporal Information Dynamics.
DOI: 10.5220/0006687900910102
In *Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2018)*, pages 91-102
ISBN: 978-989-758-293-6
Copyright © 2019 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

terest (AOI) drive on the road and may still take the exit, leading to a longer overall route. Consequently, if there is a traffic jam at the Place of Action (POA), the vehicles should take the exit to achieve optimal routing. However, the information about the state of the road (*jammed* or *not jammed*) needs to be distributed from vehicles in the POA to those in the AOI. We evaluate the impact of false information and the time it takes to adapt to changed road state relying on an accurate model of vehicular mobility. Our evaluation shows that our decision-making process outperforms state-of-the-art approaches significantly, reducing the amount of false decisions by up to 25%.

The remainder of the paper is structured as follows: We provide relevant background on HMMs in Section 2, followed by a discussion of existing caching systems and their handling of contradicting information in Section 3. We present our contribution, the freshness-based majority voting approach for decision-making under ambiguous or contradictory information in Sections 4 and 5. Section 6 contains an in-depth evaluation of our approach, comparing its performance against state-of-the-art decision-making processes and the optimal solution derived numerically. The paper is concluded in Section 7.

2 HIDDEN MARKOV MODEL

A Hidden Markov Model (HMM) is a statistical model in which the system states cannot be observed directly. The hidden states depend on the observable ones. Thus, the value of the hidden states cannot be assured.

We model the road conditions and the associated detection as a HMM. Figure 2 displays a general system model. There are two reasons for modeling the detection of road conditions as a HMM:

2.1 Measurement Error

The connections between the Observable and the Hidden Layer symbolize the measurement process. The real state (Hidden State (HS)) of the road ($HS_1 \ldots HS_n$) is hidden from the vehicles. The vehicles cannot directly measure the hidden states due to the restrictions of their onboard sensors. They can only measure the Observable State (OS) ($OS_1 \ldots OS_m$) on the observable layer, which maps to the associated hidden state with a certain probability. The solid lines symbolize a high probability for the mapping. If a vehicle measures a state OS_i, there is a high probability that the real state of the road is HS_i. However, this cannot be assured. If the measurement of the vehicle is erroneous, the real

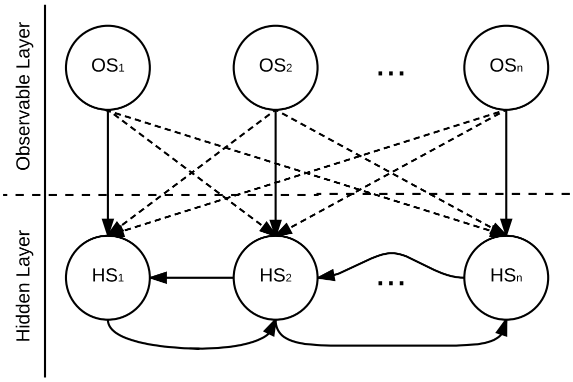


Figure 2: Hidden Markov Model for a general system.

state of the road differs from HS_m. This error is symbolized by the dotted line. The number of observable states n and the number of hidden states m can differ. We assume the measurement error to be equal for all vehicles. Thus, the number of observable states and hidden states are equal.

2.2 State Change

The hidden state of the road changes over time. Each state has a probability to stay the same state and a probability of a state change. The arrows between the hidden states symbolize the transitions between the hidden states. If the probability of a state change is high, the event is highly dynamic. If this probability is low, the event is considered static.

We will use this specific behavior for the optimization of our decision algorithm.

3 RELATED WORK

In this section, we summarize the previous works in the context of this paper. As our approach is based on the quality of the information for decision making, we first provide an overview of the respective literature. After that, we provide an overview of previous works towards decision making in distributed systems.

3.1 Quality of Information (QoI) in Distributed Networks

In the literature, Quality of Information (QoI) assessment is a repetitive topic. QoI consists of different dimensions, each dimension describing a specific property of the information. The importance of each dimension depends on the application. Not every dimensions is applicable useful for all applications.

Wang and Strong (Wang and Strong, 1996) surveyed data consumers on essential quality dimensions for information management systems. Based on this work, other researchers adapted the QoI dimensions for their applications. Chae et al. (Chae et al., 2002) adapted the concept of QoI for mobile internet applications. They took four dimensions into account, which describe the connection, content, interaction and contextual quality. They survey people to determine how the different quality dimensions combine to an overall QoI metric.

In vehicular networks, QoI is pivotal for correct decision making in vehicular applications (Kakkasageri and Manvi, 2014). Each vehicle performs the information validation by itself. The idea of Fawaz et al. (Fawaz and Artail, 2013) is to choose the Time to Live (TTL) dynamically dependent on the history of changes. With their work, it is possible to estimate the TTL of an information type. For vehicular networks, three dimensions are most important: the content quality, the trust between the vehicles and the spatiotemporal relevance of information. The necessary meta-information are available for every vehicle. Delot et al. (Delot et al., 2008) estimated the geographical relevance of information in vehicular networks. They calculated the geographical relevance using the encounter probability of the vehicle and the information. For the temporal quality, Kuppusamy et al. (Kuppusamy and Kalaavathi, 2012) published an approach called Cluster Based Data Consistency (CBDC). They concentrated on increasing the data consistency and accessibility in clustered Mobile Ad-hoc Networks (MANETs). They assured the freshness of information using a TTL value. After the expiration of the TTL, the information is considered invalid and removed from the cache. These metrics are made for their respective use cases. Though, to the best of our knowledge, there is no metric for decision making available, which can handle uncertainty. For this, the temporal relevance, the content quality and the trust between vehicles are pivotal. We extend the work of Meuser et al. (Meuser et al., 2017) with an approach to explicitly model the decrease of information value based on the TTL of the information.

3.2 Decision Making under Uncertainty

In most vehicular applications, vehicles rely on a threshold for the number of messages required to update their decision (Kakkasageri and Manvi, 2014). Molina et al. (Molina-Gil et al., 2010) researched on the security consideration in vehicular networks. They proposed a probabilistic signature validation scheme to reduce computational overhead while preventing incorrect messages. Hsiao et al. (Hsiao et al., 2011) modeled the validation of message based on their quality implicitly. Although their approach focuses on trust, it can be used for inaccurate information likewise. They validated messages of other vehicles using the already received messages. The vehicles only perform an adaptation if the message amount is sufficiently high.

In previous work, Meuser et al. (Meuser et al., 2017) used a HMM to model information with discrete event space. Using the spatiotemporal relation between information, they were able to aggregate information of different time and location. In their work, the impact of old information decreases exponentially. Moreover, they took the content quality into account and decreased the impact of inaccurate information. In their work, they did not mention how to derive the spatiotemporal dependency between information.

To our best knowledge, there is still a gap in rating QoI for dynamic information in vehicular networks. Previous work focused either on static information or provided non-optimal solutions for dynamic information. Thus, we will focus on a freshness- and accuracy-aware validation scheme for information in vehicular networks.

4 PRELIMINARIES

Vehicles can exchange information using multiple communication technologies. Available communication technologies are the cellular network and the wifi-based 802.11p standard. In general, 802.11p is used for emergency communication, while non-safety-related services need to be performed via mobile communication, as 802.11p is not suitable for high distances due to its multihop behavior. An example for non-safety-related services is the distribution of jam information.

Non-safety-related information contains meta-information to enhance the information. This meta-information are the detection time, the detection place and the expected lifetime. That information is essential for other vehicles to rate the information.

This information is distributed among the affected vehicles using a Publish/Subscribe system. For this system, we assume that every vehicle is equipped with a cellular network connection. A Publish/Subscribe server manages subscriptions and publications.

4.1 Publish/Subscribe System

The Publish/Subscribe system used is an attribute-based Publish/Subscribe system. The attributes are the ids of the road segments on which the information is located. These ids can, e. g., be extracted from OpenStreetMap[1].

While driving on the streets, each vehicle perceives its environment and shares the information with interested vehicles. For that, the vehicle publishes the information with the id of the affected road segment.

Interested vehicles subscribe to road segments to receive this information. Those road segments are parts of the planned route of the vehicle. Once a vehicle receives information, this information is stored in the cache until the information lifetime expires.

4.2 Scenario Description

In this work, we focus on an example scenario, which is visualized in Figure 3. It can be divided into 4 different phases.

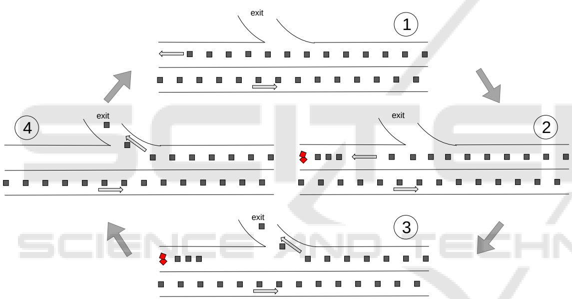


Figure 3: Visualization of the different road Phases.

In the first phase, there is no traffic jam, and the traffic flows as usual.

In the second phase, an obstacle blocks the road, e. g., a broken car. Due to the road blockage, the traffic jams. Several hundred meters distant from this point, there is an exit to bypass the accident. However, the drivers near this exit do not know about this incident. Hence, they do not leave the road and drive into the traffic jam. As the vehicles in the jam know about the blockage, they publish this information. A vehicle near the exit receives this information. After it believes the other vehicles that there is a jam, the system changes to the third phase.

In the third phase, the vehicles take the exit. We assume that under normal traffic conditions the exit of the road is a diversion. However, during the blockage of the road, the detour is the fastest route.

In the fourth phase, once the road blockade is over, the drivers still take the detour because they have no information about the jam dissolution. Thus, the vehicles at the former traffic jam publish the information that the jam has resolved. After the vehicles near the exit are confident in the received information, they stay on the road and do not take the detour anymore.

[1]http://www.openstreetmap.org

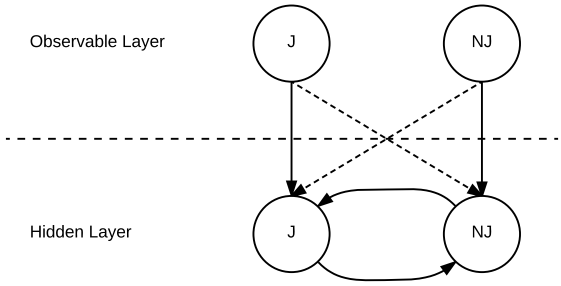


Figure 4: Hidden Markov Model for a traffic jam.

4.3 Traffic Jam Modeling

The example scenario uses a traffic jam as an example for road blockage. To make decisions based on the information type, we need to model the information. We use the model of a HMM as already used in (Meuser et al., 2017). With the HMM, the transition between states can be predicted easily. This model is trained with historic data.

The HMM for a traffic jam is shown in Figure 4. We assume there are two states for this information type: either a road segment is *jammed* or *not jammed*.

Once a vehicle tries to measure the state of a road segment, it has a certain probability to measure the correct state, i. e., measuring the road is *jammed*, and the road is *jammed*. With a low probability, the measurement is wrong, i. e. the vehicles measure the road is *jammed*, but the road is *not jammed*. The solid line between the observable layer and the hidden layer is of high probability, while the dotted line is of low probability.

The change of a road naturally changes over time. If a vehicle has measured the state of the road in the past, this measurement cannot predict the future state with certainty. We model this behavior with the state transition in the hidden layer.

4.4 Decision Making

Every time a vehicle has the chance of a detour, it checks the information in its cache. If there is a road-related information for the road segments after the current and the next exit, the vehicle evaluates the available information. If the vehicle expects the information to be correct and valid, it takes the exit.

A traffic jam is a dynamic information in the vehicular context. As vehicles can only observe their close environment, the quality of a jam detection is varying.

5 QUALITY OF INFORMATION BASED DECISION-MAKING

Decision-making algorithms can benefit many vehicular applications. In this paper, we investigate the example of jam detection. Like most other information required by vehicular applications, the jam state of the road changes regularly.

Existing approaches from literature do not use the full potential of the information, as they do not consider these information-specific properties and thus, adapt either too slow or too fast.

Slow adaptation leads to decisions based on false knowledge. If the environment changes, the vehicle still considers the old information as correct. This misinformation produces costs for the vehicles, which is, e. g., the unnecessary rerouting in case of a traffic jam.

On the opposite, fast adaptation is very sensitive to false information and creates costs through incorrect information. The costs of slow and fast adaptation are obviously contrary. The costs through false information rise if an approach adapts very fast to incoming information. On the other hand, the costs rise with increasing change rate of the environment if an approach adapts slowly.

In the following, we derive a formula for the costs of both fast and slow adaptation. We solve the resulting optimization problem to achieve the lowest possible cost.

5.1 Problem Formulation

For convenience, Table 1 provides an overview of the used variables.

We minimize the total costs $c_{total}(n)$ for wrong decisions as shown in Equation 1. The variable n is the number of messages after which the vehicle adapts to incoming information and updates its decision. Changing the value of n influences the adaptation speed of the algorithm.

$$\min c_{total}(n) \tag{1}$$

The total costs $c_{total}(n)$ for a wrong decision consist of two costs, the costs of slow and fast adaptation. They are shown in Equation 2.

Table 1: Overview of used Variables.

Variable	Description
c_{total}	Total costs of wrong decisions
c_{slow}	Costs of slow adaptation
c_{fast}	Costs of fast adaptation
n	Number of messages for adaptation
n_{opt}	Optimal number of messages for adaptation
p_f	Rate of incorrectly sensed information
p_c	Change probability of the sensed environment per time interval
T	Time to Live (TTL) for the information type
C_{fast}	Costs of an incorrect change of decision per time interval
C_{slow}	Costs of an incorrect keep of decision per time interval
t_i	Age of information i
s_i	State of information i
$f(t_i)$	Impact function of information i
$I(s)$	Impact of all information of state s

$$c_{total}(n) = c_{fast}(n) + c_{slow}(n) \tag{2}$$

The first summand is the costs $c_{fast}(n)$ for a too fast adaptation. Too fast adaptation leads to a high impact for erroneous measurements. Thus, for low accuracy measurements, the adaptation is required to be slow. The second summand is the costs $c_{slow}(n)$ for a too slow adaptation. If the real variable value changes, but the vehicle does not adapt to this change, the vehicle makes the wrong decision. For high accuracy information, the adaptation time can be low to decrease this costs.

The costs $c_{fast}(n)$ are calculated in Equation 3. They consist of the probability for a vehicle receiving a sufficiently high number of wrong information to adapt to the false information. The vehicle calculates this probability using the false detection rate p_f derived from the HMM. For this, the average false detection rate is used. The variable C_{fast} represents the costs that describe the negative impact of the decision. These costs depend on the additional costs that emerge for the vehicle in case of a false adaptation. They are the difference between the costs of the adaptation and the costs of the correct decision.

$$c_{fast}(n) = p_f^n * C_{fast} \tag{3}$$

The costs $c_{slow}(n)$ are shown in Equation 5. These costs consist of the number of messages required for the change n, the probability for a change p_c and the costs of the wrong decision C_{slow}. n states the number of messages that a vehicle requires to update its decision. As long as the vehicle has not received the required amount of messages, it will make the wrong decision. We derive the probability for a change p_c from the rate r of incoming messages per second. Vehicles individually measure this rate, but consider this rate to be constant. We calculate p_c under the assumption that the message is invalid after the TTL T. A message is invalid once the probability for any state is equal to the probability for the current state. Equation 4 shows the value for p_c with $|S|$ being the number of possible states.

$$p_c = 1 - \sqrt[T]{1/|S|} \tag{4}$$

The costs C_{slow} are calculated similarly to the costs C_{fast}, using the difference in cost of the best decision and the decision that has been made.

$$c_{slow}(n) = n * p_c * C_{slow} \tag{5}$$

5.2 Optimization Problem

As we want to minimize the costs of wrong decisions, we minimize the costs c_{total}. We search for this minimum by deriving the costs $c_{total}(n)$ for n and set it equal to 0 as shown in Equation 6. We transform this equation to Equation 7.

$$\frac{\delta}{\delta n} c_{total}(n) = 0 \tag{6}$$

$$p_f^n * \ln(p_f) * C_{fast} + p_c * C_{slow} = 0 \tag{7}$$

Solving Equation 7 results in the optimal number of messages n_{opt}. If a vehicles adapts to incoming information after n_{opt} messages, the total costs for this value is minimal, which can be derived from the behavior of the cost function. Equation 8 shows the optimal value n_{opt}. We require the number n_{opt} to be integral, thus round it.

$$n_{opt} = \left\lfloor \frac{\ln\left(-\frac{p_c}{ln(p_f)} * \frac{C_{slow}}{C_{fast}}\right)}{\ln(p_f)} \right\rceil \tag{8}$$

We need to develop an algorithm that adapts to new information after n_{opt} messages. An intuitive approach uses the approach from the literature, which adapts after a certain amount of information. This approach is robust to false information. However, its adaptation is still slower than possible.

This slow adaptation is justified by the algorithm behavior, which requires n_{opt} messages in a row to perform the adaptation. Assuming a vehicle receives $n_{opt} - 1$ messages with the new information and afterward one message with the old information, it cancels the adaptation and needs to restart it. Thus, we develop an algorithm that solves this problem.

5.3 Quality of Information-based Majority Voting

We propose a freshness-based majority voting algorithm which optimizes the costs. In the existing literature, two main approaches are proposed for decision making:

A conventional approach is to decide after a certain amount of information. This approach considers information to be correct if the vehicle has received a certain amount of messages with that information in a row. The issue with this approach is the determination of the exact message amount. For low amounts, this approach is very prone to false information.

The other standard approach decides using the amount of available information. This approach considers the information as correct, of which it has stored the most messages in the cache. This approach is resilient to incorrect information but adapts to changes slowly.

Our approach is based on majority voting and combines the advantages of both these approaches. In conventional majority voting, every vote has equal weight. Majority voting by itself is very resilient to incorrect information but adapts to changes slowly. We solve this problem by changing the weights for the information in the voting process. The weight the information is chosen in a way that a vehicle adapts after an optimal amount n_{opt} of information.

Our approach considers the freshness and accuracy of the information and works as follows: Given a set of messages M for a particular edge, the vehicle can calculate the voting score using the age t_i and the state s_i of the messages $i = 1..|M|$ as shown in Equation 9. M_s is the subset of messages containing messages of the state s. The function $f(t)$ is an impact function, which adapts to the information type. The parameter t is the age of the information in the cache.

$$I(s) = \frac{\sum_{i|s=s_i} f(t_i)}{\sum_{i=1}^{|M|} f(t_i)} \tag{9}$$

The vehicle chooses the state with the highest impact score $I(s)$. The advantage of our approach is that it adapts faster to environmental changes than conventional majority-voting, as old information are assigned smaller weights. Compared to always adapting to the newest available information, our approach is less prone to false information and can, thus, ensure a higher percentage of correct decisions.

The impact function $f(t)$ weights information in the cache. This function describes the tradeoff between fast adaptability and resilience to false information. In the next part, we will derive the function $f(t)$.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

5.4 The Impact Function $f(t)$

The impact function $f(t)$ depends on the expected rate of false information p_f and the change rate of the information p_c. As described in section 4, we model the road information using a Markov chain. Thus, $f(t)$ is a general exponential function as shown in Equation 10.

$$f(t) = a * e^{bt} + d \tag{10}$$

Based on $f(t)$, any exponential function can be created using the appropriate values for a, b and d. In the following, we will derive the values for the parameters a, b and d using the three requirements of this function.

5.4.1 Impact of New Information

The initial weight of detected information needs to be equal to the expected accuracy of this information. As t is the age of the information, Equation 11 must be true.

$$f(0) = a + d = 1 - p_f \tag{11}$$

5.4.2 Invalidation of Information After the TTL

A vehicle removes information from the cache after the TTL has expired. The weighting function gradually decreases the impact of the information. Thus, the impact of the information at the expiration of the TTL equals 0 and Equation 12 must hold true.

$$f(T) = a * e^{bT} + d = 0 \tag{12}$$

Using these two requirements, we can derive the family of parametric functions with the parameter b in Equation 13. For this, Equation 11 and Equation 12 are inserted to replace the values of a and d. Thus, this family of parametric functions ensures that the two requirements of Equation 11 and Equation 12 are satisfied regardless of the value of b.

$$f_b(t) = \frac{(1 - p_f) * \left(e^{bt} - e^{bT}\right)}{1 - e^{bT}} \tag{13}$$

Figure 5 displays the impact of the parameter b on the behavior of the respective functions. For variables with a low detection accuracy, the impact function needs to stay at its start value for a long time to compensate for the high amount of measurement errors. For variables with high detection accuracy, the impact function reduces the impact drastically after a short time to utilize the high reliability of the detected information.

To determine the exact value for the parameter b, the third and final requirement to this function is used. For this, we developed a trial-and-error based heuristic to approximate b.

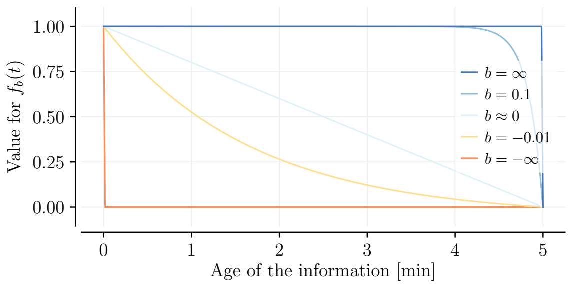


Figure 5: Visualization of the family of functions for different parameters b.

5.5 Approximation of b

We use the message amount n_{opt} with the minimal cost value to choose the appropriate value for b. We choose b such that the vehicle updates its decision after n_{opt} messages. However, the vehicle cache contains only a certain amount of messages with certain timestamps. To map the amount n_{opt} to the local knowledge of the vehicle, we derive the rate of messages r per second from the information already stored in the cache. Using the rate r, we assume a uniform distribution of information in the cache.

For the approximation of b, we look at a consistent cache with uniformly distributed messages. This consistent cache is a simplification, but in the evaluation, we can show that it performs well in simulations. We cannot use the actual distribution of the information in the cache as we cannot evaluate the correctness of the information in the cache and have no insights on future messages.

Equation 14 shows the impact $I(s_o)$ for the former state s_o. This state has been active before a change. The impact of the state sums the results of the impact function for all information in the cache starting from message n_{opt} to the last message T/r. We use n_{opt} as the first message because the messages from 0 to n_{opt} are the messages with the new state.

$$I(s_o) = \sum_{n=n_{opt}+1}^{\lfloor T/r \rfloor} f_b(n * r) \tag{14}$$

In addition to the impact of the old state, Equation 15 shows the impact $I(s_c)$ of the current state s_c. The change happened at the time $t = n_{opt} * r$. As the information with the new state comes in with same rate r as before the change, we sum the impact of all messages between 0 and n_{opt}

$$I(s_c) = \sum_{n=0}^{n_{opt}} f_b(n * r) \tag{15}$$

We want to have a decision update after n_{opt} messages. For all n smaller than n_{opt}, the vehicle does not

SCIENCE AND TECHNOLOGY PUBLICATIONS

change its decision. After the vehicle has changed its decision, the impact of the new decision raises constantly and thus it sticks to that decision. To find the optimal value of b, at which the adaptation is performed after n_{opt}, the impact of the old and the new state need to be similar. Equation 16 shows this equality.

$$\sum_{n=n_{opt}+1}^{\lfloor T/r \rfloor} [f_b(n*r)] = \sum_{n=0}^{n_{opt}} [f_b(n*r)] \qquad (16)$$

We solve this equation by trying out different values for b until we find a b for which this equality holds. Once we find b, we have completed our impact function, which a vehicle uses for its decision making. We define the two extreme cases separately. For $n_{opt} = 0$, b is equal to ∞ and for $n_{opt} + 1 > T/r$, b is equal to $-\infty$. Using b, we can exactly calculate the impact of old information and thus make good decisions based on that information.

5.6 Uniform Distribution of Messages

For the described approach, the messages in the cache need to be uniformly distributed. This distribution is automatically achieved for information which the vehicles detect bypassing, as vehicles are naturally driving over the road segment one after the other. However, if the vehicles get stuck at the information location, each vehicle transmits the information on detection and retransmits it after the expiration of the TTL. This approach leads to many messages at roughly the same time.

We solve this problem by adding a random factor to the first retransmission interval. Instead of retransmitting after the expiration of the TTL, the vehicle performs the first retransmission after a random time, which is between 0 and the TTL T. This way, the messages are distributed more uniformly.

6 EVALUATION

For the evaluation of our developed decision approach, we simulate a traffic jam on a highway. In front of the traffic jam, each vehicle has the possibility to leave the road. Figure 6 displays the road scenario.

Initially, there is no information about traffic jams in the network. At a certain point in time, the traffic congests and the vehicles distribute the information about the traffic jam using a Publish-Subscribe system. As there were no messages in the network before, the adaptation for the initial jamming of the road is fast. After a certain amount of time, the traffic

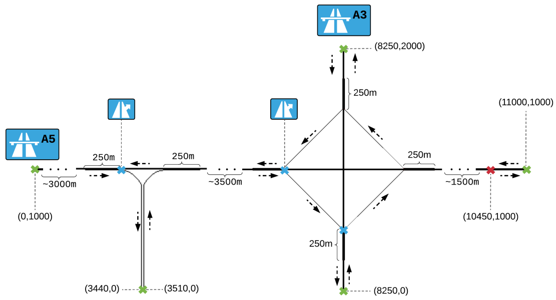


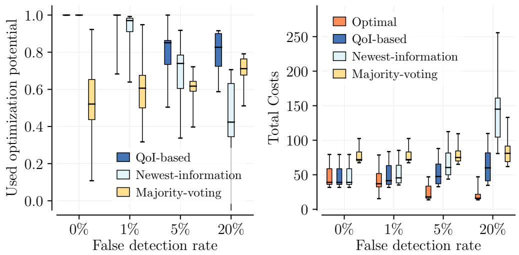
Figure 6: Schematic overview of the road scenario.

jam resolves and thus, the vehicles at the jam location publish the information that the traffic jam has resolved. However, there are already information in the vehicle's cache indicating that there is a traffic jam. Hence, the adaptation to the new requirements is far more challenging as each vehicle needs to decide on the correctness of the cached information. We consider this scenario as an appropriate scenario for decision-making, as each vehicle decides to leave or stay on the road.

We implement this scenario in the event-based Simonstrator framework (Richerzhagen et al., 2015). The Simonstrator is a network simulator, which supports different communication technologies (mobile and ad-hoc communication) and, beyond others, the Publish/Subscribe paradigm. As the movement models in the Simontrator are not suitable for our vehicular usecase, we extend the Simonstrator with Simulation of Urban Mobility (SUMO) (Behrisch et al., 2011). The connection to SUMO is accomplished using the TraCI interface of SUMO.

In the evaluation, we used two metrics to compare our approach to the state-of-the-art approaches: the costs of wrong decisions and the ratio of correct decisions. We put special focus on the reduction of costs induced by wrong decisions. We assume that costs occur every time a vehicle would have made a wrong decision. Thus, we observe the cache and make a decision every second. If this decision is wrong, we add the appropriate costs to the total costs.

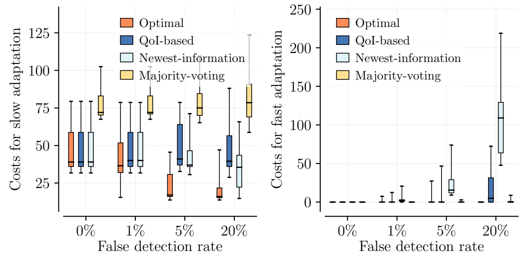
For comparison, we implemented both an optimal strategy and a random strategy. The optimal strategy chooses the correct information out of the cache using global knowledge, but can still make the wrong decision if there is no correct information in the cache. The random strategy chooses information out of the cache randomly and considers this information as correct. We expect none of the described approaches to perform better than the optimal or worse than a random approach. Thus, these are suitable bounds. We combine those two strategies to determine the used optimization potential of every approach. We



(a) Used optimization potential

(b) Total costs

Figure 7: Total costs for different false detection rates.



(a) Costs of slow adaptation

(b) Costs of fast adaptation

Figure 8: Costs by cause for different false detection rates.

calculate the used optimization potential as shown in Equation 17.

$$\text{opt} = \frac{c_{approach} - c_{random}}{c_{optimal} - c_{random}} \tag{17}$$

Table 2 gives an overview of the considered parameters and their values. We varied the false detection rate, the jam duration and the costs of a wrong detour. The bold values in the table are the default ones. The costs are calculated based on the cost ratio. To ensure a comparability of costs, we set $C_{fast} + C_{slow} = 2$. The evaluation source code is freely available[2].

Table 2: Overview of used variables.

Evaluation parameter	Value
False Detection Rate	[0%, 1%, **5%**, 20%]
Jam Duration	[200s, **300s**, 400s]
Cost Ratio between the costs of a wrong stay and a wrong exit	[0.1, **1**, 10]

We first have a look at the metrics for the different false detection rates, as the false detection rate has the highest impact on the results. For the other parameters, we only investigate on the costs of wrong decisions.

6.1 Impact of the False Detection Rate for the different Approaches

At first, we investigate on the costs and after that on the percentage of correct decisions for different false detection rates.

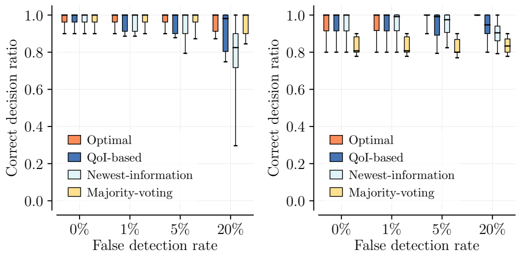
6.1.1 Costs of False Decisions

Figure 7(a) shows the used optimization potential of our approach compared to the approaches in the literature dependent on the false detection rate. For a false detection rate of 0%, our approach and the newest-information approach perform as well as the optimal approach, while the majority-voting based approach performs much worse due to its slow adaptation to environmental changes. The used optimization potential of our approach and the newest-information approach drops with increasing false detection rate. In contrast, the optimization potential of the majority-voting based approach increases. Our approach converges towards the majority-voting based approach for high false detection rates. The reason for the increased performance of the majority-voting based approach is the decrease of time for adaptation to environmental changes, as the number of correct messages in the cache is lower.

Figure 7(b) displays the total costs. Regarding total costs, our QoI-based approach has almost equal total costs regardless of the false detection rate. Compared to the approach selecting the newest information, our approach has up to 56% reduced overall costs dependent on the false detection rate and up to 43% reduced costs compared to the conventional majority-voting based approach. The total costs of the optimal approach decrease with increasing false detection rate, as false messages lower the adaptation time after the traffic jam. We explain this behavior in more detail in the next paragraph.

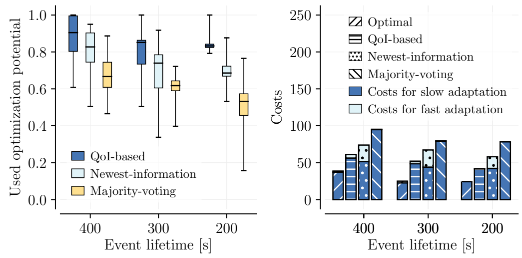
In Figure 8(a) the costs of slow adaptation are shown. We can observe that the costs of our QoI-based approach are almost equal to the costs of the fastest approach, which immediately adapts to new information. This is a great result, as our approach is far more robust to false information. The majority-voting approach has high costs of slow adaptation, as expected, since many message are required for a decision change. Interestingly, some of the approaches have lower costs of slow adaptation with higher false detection rate. Normally, it is expected that those costs increase with increasing false detection rate. However, this behavior happens if a vehicle performs a false measurement during the traffic jam, i. e., a message stating that the traffic jam has resol-

[2] https://dev.kom.e-technik.tu-darmstadt.de/simonstrator/



(a) Percentage of correct decisions during the jam (b) Percentage of correct decisions after the jam

Figure 9: Percentage of correct decisions for different false detection rates.



(a) Used optimization potential (b) Total costs of wrong decisions with cause

Figure 10: Costs for different jam durations.

ved. Although that measurement was incorrect at this point, it becomes true as the traffic jam resolves shortly after. As this originally false message is also used in the decision process, the adaptation time and thus the costs can be lower with increasing false detection rate.

Figure 8(b) shows the costs of fast adaptation. Those costs are 0 for all approaches for a false detection rate of 0%, as there are no false measurements in the system. With increasing false detection rate, the costs of fast adaptation of the algorithm increase. This is caused by the increasing number of false information in the system, which challenges the robustness of each approach. The newest-information approach is not robust to false information, thus the costs of fast adaptation increase drastically. Compared to that, the costs of fast adaptation of our QoI-based approach increase only slightly and are much lower than the costs of the newest-information approach. In contrast, our approach has higher costs of fast adaptation than the majority-voting based approach, but this behavior is intentional, as the higher costs of the fast adaptation produce lower costs of slow adaptation. Interestingly, even the optimal approach has costs of fast adaptation. This is not intuitive, as the optimal approach never makes a wrong decision. However, even the optimal algorithm makes wrong decisions, if there are only false information in the cache. This happens if the last correct message is removed from the cache due to its age.

6.1.2 Percentage of Correct Decisions

This metric investigates the effects that the decision approach has on the vehicles. Every time, a vehicle makes a decision, this decision is stored and influences the metric. Two phases are investigated. The first phase is for the decision-making during the jam. Meaning, this metric is the percentage of vehicles that have successfully taken an exit. The second phase is for the decision-making after the jam. This metric is for the percentage of vehicles that have stayed on the road after the traffic jam has resolved.

Figure 9(a) shows the correct decision ratio during the traffic jam. At the beginning of the jam, there is no information in the cache, as the vehicles do not share any information prior to the jam. Thus, the ratio is generally higher compared to the situation after the jam. However, a high correct decision ratio during jam means that only a few vehicles are in the jam and able to measure the road state. Thus, the ratio of correct decisions drops with increasing false detection rate, as the impact of false information is higher with fewer messages in the system. Moreover, we can see that our QoI-based approach perform better than the newest-information approach for every false detection rate and only slightly worse than the majority-voting based approach.

The ratio of correct decisions after the jam is shown in Figure 9(b). Our QoI-based approach is equally good as the newest-information approach with a false detection rate of 0%, but while the newest-information approach drops with increasing false detection rate, our QoI-based approach is robust to false information and thus stays almost at the same level. The performance of the majority-voting based approach also does not drop with increasing false detection rate. However, its correct decision ratio is always lower than the one of our QoI-based approach.

6.2 Impact of the Jam Duration to the different Approaches

Figure 10 shows the costs and the cost distribution for the different jam durations. With decreasing jam duration, the used optimization potential decreases likewise as shown in Figure 10(a). However, the ratio between the different approaches is not affected significantly.

To explain the decrease of costs with decreasing jam duration, we use the distribution of costs shown in Figure 10(b). There are two reasons for this decrease. Firstly, the number of messages in the cache

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

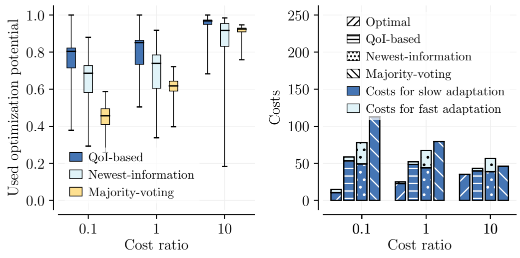
is lower, as fewer vehicles published the information. Thus, the number of messages required for adaptation is also lower. This is true for our QoI-based approach as well as for the majority-voting based approach. Secondly, the possibilities for wrong decisions during the jam is lower, as the jam duration is shorter. The newest-information approach is one of the examples, where the costs of fast adaptation decrease due to the shorter jam duration. However, the costs of fast adaptation does not decrease strongly. This is due to the lacking robustness of the newest-information approach. This leads to a very time until the system has recovered after the jam, which increases the costs of this approach.

Conclusively, we derive that our QoI-based approach outperforms the other approaches independently of the jam length.

6.3 Impact of the Costs Ratio to the different Approaches

Figure 11 shows the costs for different cost ratios. In this example, a cost ratio of 10 means that a wrong stay on the road is 10 times more expensive than a wrong leave. Similarly, a cost ratio of 0.1 means that it is 10 times more expensive to leave the road than to stay. Figure 11(a) shows the used optimization potential of the different approaches. Our QoI-based approach performs best for all of the considered cost ratios. It is also the only approach that reaches the full optimization potential in some cases. With increasing cost ratio, the used optimization potential of all approaches increases. This behavior is analyzed using Figure 11(b).

For the optimal approach, the costs of slow adaptation are almost 0. As this approach adapts fast to the jam resolving, the main costs arise through the road jamming. Thus, the costs of falsely staying on the road are the most significant factor for the optimal approach. As these costs increase with increasing cost ratio, staying on the road is punished additionally. Thus, the overall costs of the optimized approach increase with increasing cost ratio. For all the other approaches, the overall costs decrease as a part of the costs are caused by leaving the road wrongly. This is justified by the evaluation scenario, as the outdated information in the cache slows down the adaptation process. Thus, the costs of slow adaptation decrease if the costs of leaving the road wrongly are small. We can observe that our QoI-based approach outperforms the other two approaches for a cost ratio of 0.1 and 1 significantly. For a cost ratio of 10, in which slow adaptation after the jam is not costly, our approach performs only slightly better than the majority-



(a) Used optimization potential

(b) Total costs of wrong decisions with cause

Figure 11: Costs for different cost ratios.

voting based approach, as the slow adaptability of the majority-voting based approach is balanced by its robustness.

6.4 Evaluation Results

The evaluation shows that our approach reduces the total costs compared to both of the existing approaches dependent on the scenario by up to 25%. Additionally, our approach has never a higher total cost value than the other approaches. It achieves that improvement by balancing its robustness and fast adaptability to environmental conditions to achieve optimal results. For the extreme cases, our algorithm converges to the newest-information and the majority-voting based approach respectively. Moreover, we observed that the jam duration has no impact on the performance improvements of our approach, which makes it usable for decisions based on arbitrary information.

7 CONCLUSION

In this paper, we proposed a Quality of Information (QoI)-based decision making process. In this decision making process, false decisions produces costs for the deciding vehicle. False decisions have two reasons, missing robustness to false measurements and slow adaptation to environmental changes. Our novel decision making process considers information-specific properties, to make decisions inducing the lowest possible costs.

This decision making process is based on a weighted majority-voting. The used weighting determines the impact of an information and considers information-specific properties. Those properties are modeled using a Hidden Markov Model (HMM) considering the false detection rate and the information lifetime. Those two properties most important for the information impact function, as missing sensor accuracy and outdated information are common challenges in distributed networks.

To chose the appropriate impact function $f(t)$, we construct an optimization problem to minimize the costs of incorrect decisions. The resulting weighting function is an exponential function and takes the age of information as an input to calculate the weighting of that specific information. The weighting function itself depends on the information-specific properties information lifetime and false detection rate.

In the decision-making process, we perform a weighted majority-voting with the weights calculated by our weighting function. In the evaluation, we show that our approach significantly outperforms comparable approaches by up to 25% and dynamically adapts to the information-specific properties.

As future work, we aim to investigate the possibilities to filter out wrong information and consider the individual false detection rate of each sensor instead of the average into account to increase the quality of the decisions further.

ACKNOWLEDGEMENTS

The work presented in this paper was partly funded by the LOEWE initiative (Hessen, Germany) within the NICER project and by the German Research Foundation (DFG) as part of project C2 within the Collaborative Research Center (CRC) 1053 - MAKI.

SCIENCE AND TECHNOLOGY PUBLICATIONS

REFERENCES

Behrisch, M., Bieker, L., Erdmann, J., and Krajzewicz, D. (2011). Sumo–simulation of urban mobility: an overview. In *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*. ThinkMind.

Chae, M., Kim, J., Kim, H., and Ryu, H. (2002). Information quality for mobile internet services: A theoretical model with empirical validation. *Electronic Markets*, 12(1):38–46.

Delot, T., Cenerario, N., and Ilarri, S. (2008). Estimating the relevance of information in inter-vehicle ad hoc networks. In *Mobile Data Management Workshops, 2008. MDMW 2008. Ninth International Conference on*, pages 151–158. IEEE.

Fawaz, K. and Artail, H. (2013). DCIM: Distributed cache invalidation method for maintaining cache consistency in wireless mobile networks. *IEEE Transactions on Mobile Computing*, 12(4):680–693.

Hsiao, H.-C., Studer, A., Dubey, R., Shi, E., and Perrig, A. (2011). Efficient and secure threshold-based event validation for vanets. In *Proceedings of the fourth ACM conference on Wireless network security*, pages 163–174. ACM.

Kakkasageri, M. and Manvi, S. (2014). Information management in vehicular ad hoc networks: A review. *Journal of Network and Computer Applications*, 39:334–350.

Kuppusamy, P. and Kalaavathi, B. (2012). Cluster based data consistency for cooperative caching over partitionable mobile adhoc network. *American Journal of Applied Sciences*, 9:1307.

Meuser, T., Lieser, P., Nguyen, T. A. B., Böhnstedt, D., and Steinmetz, R. (2017). Adaptive information aggregation for application-specific demands. In *Proceedings of the 1st BalkanCom*.

Molina-Gil, J. M., Caballero-Gil, P., Hernández-Goya, C., and Caballero-Gil, C. (2010). Data aggregation for information authentication in vanets. In *Information Assurance and Security (IAS), 2010 Sixth International Conference on*, pages 282–287. IEEE.

Richerzhagen, B., Stingl, D., Rückert, J., and Steinmetz, R. (2015). Simonstrator: Simulation and prototyping platform for distributed mobile applications. In *Proceedings of the 8th International Conference on Simulation Tools and Techniques*, SIMUTools '15, pages 99–108. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Wang, R. Y. and Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumer. *Journal of management information systems*, 12:5–33.