Ability-aware trust for vehicular networks

Walter Bamberger*, Josef Schlittenlacher and Klaus Diepold

Department of Electrical Engineering and Information Technology,
Technische Universität München,
80290 Munich, Germany
E-mail: walter.bamberger@tum.de
E-mail: josef.schlittenlacher@mytum.de
E-mail: kldi@tum.de
*Corresponding author

Abstract: Future vehicles will exchange much information through a wireless network in order to efficiently maintain their inner model of the environment. Before they can believe received pieces of information, they must evaluate their reliability. Trust is a mechanism to estimate this reliability based on the sender. It depends on friendship-like relations between vehicles, the social structure. Our traffic and network simulation shows that such relations indeed arise, because vehicles often drive the same route. Within this simulation, all vehicles are equipped with the proposed trust model, which continuously monitors the experiences made with others. The model focuses on these direct experiences of the individual and does not depend on a central reputation unit. It continuously evaluates the performance and reputation of other vehicles and includes a feedback loop to faster adapt to changes in the other's behaviour. Since the performance of a vehicle depends on the abilities of its sensors, like different abilities in velocity measurement or in traffic sign detection, the proposed model develops trust depending on the sender and on the type of the information.

Keywords: ability aware; information security; probabilistic logic; reliability; reputation; social structure; trust model; vehicles; vehicular ad hoc network; VANET.

Biographical notes: Walter Bamberger works as a Doctoral candidate at the Technische Universität München. He is interested in interdisciplinary research on the fields of engineering and social sciences. Currently, he investigates trust between cognitive systems in alignment with trust between humans: Trust helps to handle the uncertainty in the interaction between those systems. With this idea, he raised a seedling fund in the cluster of excellence Cognition for Technical Systems (CoTeSys).

Josef Schlittenlacher studied electrical engineering at the Technische Universität München and currently a Doctoral candidate at the Institute of Psychology of the Technische Universität Darmstadt. His research interests are related to the combination of both subjects as well as to the functionality of the human brain.

Klaus Diepold received his Dr-Ing from the Technische Universität München. He spent more than ten years in industry working on digital television and video compression. Currently, he holds a Chair at the Department of Electrical Engineering and Information Technology of the Technische Universität München. His research interests include fast algorithms for signal processing of audio-visual data and for machine learning in cognitive systems. He is a Principal Investigator in the excellence cluster CoTeSys and engaged in entrepreneurial activities for the commercialisation of academic research.

1 Introduction

Consider a scenario in which cars continuously communicate with each other while driving around. These future cars can perceive, reason, learn, plan, and act in a way that they understand the surrounding traffic scene while still controlled by a driver (as in the CoTeSys satellite project MuCAR-3 in Munich). In the literature, these cars are sometimes called cooperative cognitive automobiles (Stiller et al., 2007). They are expected to improve the traffic efficiency and safety. These cars need a vehicular network to efficiently maintain the inner model of their environment. With this network, they exchange all kinds of model information as shown with the following examples (CAR 2 CAR Communication Consortium, 2007; Kranz, 2008; Matheus et al., 2004):

- *Long-term information*: structural alterations, permanent changes in the traffic regulations (e.g., traffic signs), new or removed points of interest (like petrol stations, hotels or museums).

- *Medium-term information*: temporary changes in the traffic regulation (e.g., traffic signs because of special events), hazardous locations (e.g., dirt or oil on the road), and content changes of a point of interest (e.g., prices).

- *Short-term information*: hazardous locations (e.g., fog, aquaplaning), obstacles along the road, free parking space, level of service, and traffic congestions.

- *Situational information*: green light optimal speed advisers, collision warning, collision preparation, merging assistance (to join flowing traffic), and emergency vehicles.

1.1 The problem statement

Despite the cryptographic measures already proposed for the information security in vehicular networks, the received data may still be wrong [Raya and Hubaux, (2007), especially p.65]. Reasons for this might be bad recognition capabilities or defect sensors at the sender, but also software virus infections or manipulations of the sending car. So, a car must try to handle the uncertainty associated with received data. It does so in two steps: it evaluates the sender over several interactions and it evaluates the received data with regard to the current situation. This is where the trust model proposed in this article comes in. It is a mechanism to evaluate the sender over several interactions. Models for the second step (that is, evaluate the data with regard to the current situation) are introduced, e.g., in Golle et al. (2004) and Raya et al. (2008).

So, this paper addresses the following problem: how can a car estimate an expectation, whether another car will send correct information about a certain subject? The authors of this paper call a mechanism that can do this *inter-agent trust*. Thus, the proposed model is a trust model. More specifically, it focuses only on the processing within the car; it is only the individual-level side of a trust-enabled network according to Ramchurn et al. (2004).

With the help of this expectation, the car should be able to distinguish good from bad information sources. In the end, it should have a more accurate inner model of the environment and should better decide what information should when be exchanged with whom (i.e., what information is really necessary and who would give it). The following scenarios make clearer in what situations trust is necessary and how a trust model must be designed.

1.2   Trust scenario

Cars have differently restricted sensory capabilities. Therefore, they sometimes miss to report an event (like a new sign) or spread out wrong information (e.g., a wrong degree of free parking space). The error rate depends on the car and on the involved sensors.

The trusting car must know how correct the values that the other car sends about a certain information type usually are. The authors call this property the competence of the other car regarding the given type of information (see Section 4.1). Received values must be processed with an uncertainty that is related to the sender's competence.

A car randomly generates fake messages about non-existent events. This may be, because it has been infected by a software virus or because the manipulator does this just for fun.

In this scenario, the trusting car must estimate how good it can predict the other's behaviour. If this is not possible (because the other's behaviour is sometimes very good and sometimes very bad), it should increase the disbelief in the other's trustworthiness as shown in Section 4.2.

Every Wednesday, a car reports wrong parking space values to ensure it can park on the desired car park. While it drives to the car park, it claims it has been there 30 minutes ago and there was no free parking space. It reports bad recommendations about other cars of the opposite direction that have reported free parking space. Besides of these Wednesdays, the car always acts trustworthy.

As in the previous scenario, a quasi-random misbehaviour should be considered as untrustworthy. In addition, this scenario shows that reputation can be manipulated (increased and decreased). Because of this, the proposed trust model takes recommendations with care as explained in Sections 4.3 and 4.4.

1.3   Overview and terms

So far, we clarified the problem this paper tackles and its surrounding scenarios. Next, we give an overview of the paper and the specifics of the proposed trust model.

First, we relate our model to selected other trust models from different fields. Then, we briefly introduce subjective logic, the framework we use for artificial reasoning.

In the subsequent Section 4, the model itself is presented in detail. In contrast to reputation systems, which are already in discussion for intervehicular communication (e.g., Raya and Hubaux, 2007), the proposed model focuses on the individual car and its

experiences. The opinion of others (that is, the reputation) is still considered, with lower priority though. In addition, the proposed trust model distinguishes the different abilities of the other car regarding its different sensors and their error rates (see Section 4.4). This paper further features the capacity of the binary error and erasure channel as a better indicator for trust (see Section 4.5). While trust is often related to something like a mean error in the performance of the other agent, the model proposed here additionally incorporates how good the trusting car can decide based on the information received from the other car. The authors call this property the other's *predictability* (detailed in Section 4.2). It helps to faster adapt to changes in the other's behaviour.

As pointed out so far, we understand trust as a mechanism that relies on a social structure. The 'social structure' of cars follows from their periodical trips. So, these trips must be part of an evaluation environment for trust. Section 5 introduces such an environment and proposes it as the right environment to investigate a trust model for vehicular ad hoc networks (VANETs). The results section then shows that a car indeed meets some other cars regularly. So, a car can create an image of other cars on its own – during 'social interaction'.

The following terminology is used throughout this paper: *Reputation* is considered as the opinion about someone by people in general – it represents the common opinion. It can be used as an indication when judging about the trustworthiness of another party, but it is different from trust. The proposed trust model tries to build-up a local view on the reputation by combining several received *recommendations* of others about a third car.

When a car drives around, it observes properties of its environment (*observation*). It may pass these properties to other cars. Such received information is called a *report*. Information that helps to judge about the trustworthiness of another car is *evidence*. So, a report that has been verified by the car becomes evidence.

This article extends our previous publication (Bamberger et al., 2010) at the SocialCom 2010 in the following ways. The model has been extended to distinguish not only cars but also types of information (which are related to sensors). Furthermore, the discounting operation is now base rate sensitive; with this modification, the system seems to settle down faster. (Both changes significantly enhance Section 4). For the evaluation, the simulation has been extended significantly. Now, cars maintain a complex knowledge base, and different algorithms for information integration, dissemination and trust development work on it (Section 5). As a consequence, Section 6 has been completely rewritten.

2 Related work

When people from vehicular network security talk about trust, they usually refer to trust relations in a public key infrastructure. This mechanism from information security mainly provides identification of entities but not trust in the sense of this paper. Raya and Hubaux (2007) give a good overview about security aspects and mechanisms in vehicular networks. For data verification, they first considered reputation [Raya and Hubaux, (2007), p.65], but then developed their own data-centric trust model (Raya et al., 2008). While their model focuses only on the data (disregarding the sender), the proposed model focuses only on the sender. So, both models complete each other. Golle et al. (2004) also verify the data, this time based on a model of the network.

VANETs are similar to wireless sensor networks. In both networks, mainly sensory data is exchanged. The framework of Zhang et al. (2006) served as a starting point for our work. In contrast to their scenarios, a car receives less reports about the same subject, but it can get to specific places (if the driver does so) to verify the received information. Furthermore, a car has several kinds of sensors, while Zhang et al. consider only one kind of sensor per sensor node. As a consequence, the proposed model develops trust separately for the different types of information a sender provides.

In the area of electronic marketplaces, reputation systems are well-known (e.g., eBay's feedback system (eBay Inc., 2010), the beta reputation system (Jøsang and Ismail, 2002), or even Google's PageRank (Brin and Page, 1998). Here, the trust comes from human beings. The reputation system then combines these statements to build a global view on an entity – the reputation. To get this global view, reputation systems work on the network level. This is in contrast to purely virtual market places of multi-agent systems. There, trust comes directly from the artificial agents. Ramchurn et al. (2004), and Sabater and Sierra (2005) give a good overview of the trust and reputation models in this area.

3   Subjective logic

Subjective logic (Jøsang, 2001, 2007) is a framework for artificial reasoning under uncertainty. It has a foundation in Bayesian statistics and set theory. This chapter shortly introduces the main concepts used in this paper. The reader finds more details in the given references.

3.1   Opinions

An agent can have some evidence that supports a statement and some that opposes it. This statement can regard a possible situation or state of the world, like 'There will be a solar eclipse in 2121' or 'Bill had an affair with Sarah'. More generally, the agent could argue about more disjoint situations or states of the world (that is, more than just pro and contra). Let $X = \{x_1, \ldots, x_N\}$ be the set of considered situations. In the evidence representation of an opinion, the strength of evidence for each situation is expressed in the variables $r_s \in [0, \infty[$ with $s = 1, \ldots, N$. These variables are collected in the evidence vector $r = (r_1, \ldots, r_N)^T$. In addition, an agent may have a subjective opinion about the situations without any direct evidence. The base rate vector a expresses this opinion. Its influence decreases with increasing evidence. With all this together, the tuple $\omega_X^M = (r, a)$ describes the opinion of the agent M about the set of situations X. The more evidence the agent has in total $\left(\sum_s r_s \right)$, the more certain it is about its opinion and the lower is its ignorance.

The same opinion could also be expressed with probabilities $p = (p_1, \ldots, p_N)^T$ for every possible situation and a degree of ignorance u about these probabilities. The tuple $\omega_X^M = (p, u, a)$ is a probabilistic representation of the opinion.

Finally, the opinion could also be expressed with the beliefs $b = (b_1, \ldots, b_N)^T$. Beliefs (Shafer, 1976) are subjective ratings of the situations with the constraint $\sum_s b_s + u = 1$.

So, the belief representation of an opinion consists of the tuple $\omega_X^M = (b, u, a)$.

Mappings exist between these three representations of an opinion, based on the Dirichlet distribution. Only the mapping from the evidence to the belief representation

$$b_s = \frac{r_s}{W + \sum_{s=1}^{N} r_s}$$

$$u = \frac{W}{W + \sum_{s=1}^{N} r_s}$$

(1)

is explicitly used in this paper. W denotes the weight of the base rate vector a.

3.2 Operations on opinions

When opinions from two or more different sources exist about a set of possible situations X, they can be fused into one (Jøsang, 2007). If the pieces of evidence are independent (like two rolls of the same dice), then the *cumulative fusion* $\omega_X^A \oplus \omega_X^B$ is the right fusion. This can be used, e.g., to combine the opinions about the outcomes of several interactions for trust development. If the pieces of evidence are dependent (like two concurrent, maybe conflicting observations of one dice roll), then the *average fusion* $\omega_X^A \underline{\oplus} \omega_X^B$ must be applied. This operation is appropriate to combine the opinions about several properties of a single interaction.

When a car B has the opinion ω_X^B about the proposition that there is a new traffic sign at a certain location, then A only beliefs this as far as it considers B as trustworthy in providing such information. So, if A has the opinion ω_B^A about B's trustworthiness, then both opinions can be combined with the *base rate sensitive discounting* into the opinion $\omega_X^{A:B} = \omega_B^A \otimes \omega_X^B$ (Jøsang et al., 2006). The discounted opinion states how much A beliefs the proposition based on B's opinion.

This article proposes two more operations not yet found in the literature: evidence scaling and priority fusion. When one opinion should have a lower weight in the cumulative fusion operation (i.e., is less important) than the other opinions, its ignorance must be increased. This could be necessary, e.g., because its evidence is older than that of the other opinions. Scaling the evidence vector r with the desired factor w realises exactly this functionality: $w \omega_X^M = (w\,r, a)$. The authors call this operation *evidence scaling*.

When someone knows from many experiences that a car performs well with one sensor, she would probably assume that it performs well with another sensor too – even if she does not have any experiences regarding this other sensor of the car. But when the person has much evidence for both sensors, she would only take the available evidence of each sensor to judge about its reliability. To realise this logic, the authors propose the *priority fusion* operation. It is defined as[1]
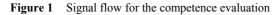
$$r_s^{PD} = r_s^P + u^P r_s^D$$
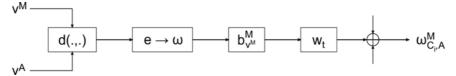$$= r_s^P + \frac{W}{\sum_j r_j^P + W} r_s^D \qquad \text{with } s = 1, \ ..., N,$$

where the superscript P denotes the prioritised opinion and D the discriminated one. W is the base rate weight as described in the previous subsection. The definition of the priority fusion shows that the influence of the discriminated opinion decreases with a decreasing ignorance u^P – which corresponds to increasing evidence in the prioritised opinion.

4   The trust model

As described in the introduction, cars exchange their knowledge about the environment in form of reports. Such a report contains the opinion ω_v^A of the sender A about the possible values v_s of a model attribute v. As a simple example, this could be the belief (v_+ = yes) and disbelief (v_- = no) whether there is a new traffic sign at a certain location. The receiver M uses this opinion (and all the opinions received from other cars) to build its own opinion ω_v^M about v. But M takes A's opinion also to develop trust in A. For this, it must judge how good A's opinion is. However, M can do this only if it knows the correct value of v. So, it must wait until it is very certain about this value. Then, it can evaluate all reports regarding v. The reports then become pieces of evidence for the trust development. This section describes how a car evaluates those reports to update its opinion $\omega_{A,i}^M$ about the trustworthiness of A concerning the information type i. The Figures 1 to 4 guide through this process.
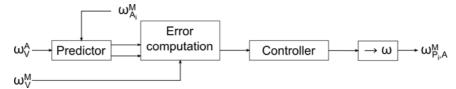
The proposed trust model incorporates three components: M's opinion $\omega_{C_i,A}^M$ about A's competence regarding the type of information i reflects the mean error of the available evidence. This is the main criteria whether reports of type i from A are good. The opinion $\omega_{P_i,A}^M$ about A's predictability indicates, whether M is able to make a right decision based on A's opinion and its trust in A to provide information of type i. It is sensitive to outliers. As its name indicates, its computation includes a predictor. Based on the predicted value, feedback control adjusts the predictability opinion. Finally, the cars exchange their opinions about other cars in addition to the reports about their environment. These opinions are called recommendations and build-up M's opinion $\omega_{R_i,A}^M$ about A's reputation in providing information of type i. In fact, this is not the opinion of all cars that know A but only M's view on the reputation of A. So, the authors call it the local reputation of A regarding the information type i. In the end, all three components of one sender regarding one type of information are fused in an intermediate opinion. Then, the intermediate opinions regarding all types of information are combined in one final trust opinion regarding the desired type of information and the desired sender.
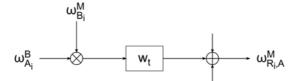
Figure 1 Signal flow for the competence evaluation

Notes: The difference between what the car M thinks is right and what the car A has sent to M determines the competence of A regarding the information type i from M's point of view.

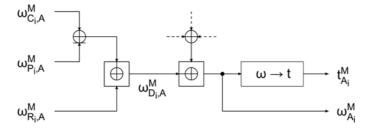**Figure 2** Signal flow for the predictability evaluation

Notes: First, only based on the opinion of A, a decision is made as described in Section 5.3. Then, this decision is compared with the car's own opinion about v. If both are different, an error measure is passed to the controller, which, in turn, adjusts $\omega_{P_i,A}^{M}$.

Figure 3 Signal flow for the reputation computation

Note: To aggregate the reputation of A regarding the information type i, M discounts all recommendations and accumulates them (weighted by their age).

Figure 4 Signal flow for the combination of all components in one trust opinion

Notes: For every information type ι, the three components competence, predictability, and reputation are combined in an intermediate opinion $\omega_{D_i,A}^{M}$. The intermediate opinions of all types are fused then in a way that the type of interest i has priority. The resulting overall opinion $\omega_{A_i}^{M}$ represents the trust in the car A regarding type i. For deciding, it is transformed in the one-dimensional trust value $t_{A_i}^{M}$.

4.1   *Evaluation of the competence*

Trust and reputation are often derived from the number of good and bad experiences in the past or from the degree how good or bad these experiences were (e.g., eBay Inc., 2010; Jøsang and Ismail, 2002; Zhang et al., 2006). The authors of this paper call this notion of trust the competence of the other car.

A car is more competent, if the information it provides is more precise. Thus, competence is inverse to the mean error of all pieces of evidence. Figure 1 shows the steps, how a car M can obtain an opinion about the *competence* of another car A regarding the type of information i.

The car M has a set of evidence $J_{A,i}$ from the car A and of the type i. For each piece of evidence, M first computes a distance between the value v^A proposed by A and the value v^M which M thinks is right. This operation can use any metric that maps to the closed range [0, 1]. In many cases, the relative error $e_{rel} = \frac{e}{e_{max}}$ might be appropriate.

This relative error indicates A's incompetence. With it, the opinion about A's competence regarding that single piece of evidence can be calculated in evidence representation as

$$r_- = e_{rel}$$
$$r_+ = 1 - e_{rel}.$$

r_- reflects A's incompetence and r_+ A's competence.

M cannot be completely sure, whether its reference value v^M is correct. So, it can also not be sure about the relative error. The belief $b_{v^M}^M$ in its assumed correct value v^M reflects the belief in the correctness of the relative error. As a consequence, M must scale the above described opinion with $b_{v^M}^M$.

Finally, the opinions about A's competence of all pieces of evidence $j \in J_{A,i}$ are combined to an overall opinion about A's competence. For this operation, the cumulative fusion operation of subjective logic is appropriate. To have an opinion about A's present competence, old evidence must be weighted lower in this operation. Therefore, all opinions are scaled with a time-dependent factor $w_{t,j}$. (It is further discussed in Section 4.6). In the end, the opinion about A's competence is

$$r_+ = \sum_{j \in J} w_{t,j} b_{v_j^M}^M \left(1 - e_{rel,j}\right)$$
$$r_- = \sum_{j \in J} w_{t,j} b_{v_j^M}^M e_{rel,j}. \tag{2}$$

So far, we have shown how several pieces of evidence with the car A regarding the information type i are transformed in one opinion about A's competence regarding i. The remaining paragraph shows that this transformation has a reasonable interpretation even in belief representation: the disbelief in A's competence results from (1) and (2) as

$$b_- = \frac{\sum_{j \in J} w_{t,j} b_{v_j^M}^M e_{rel,j}}{\sum_{j \in J}^N w_{t,j} b_{v_j^M}^M + W}.$$

W is the weighting factor for the base rate vector (the a priori probability) in subjective logic. When only the observed evidence is taken into account without a priori considerations (*W* = 0), then b_- is a weighted mean of the relative error. So, $\omega_{C_i,A} = (b,u)$ reflects the competence as proposed in the beginning of this section: competence is inverse proportional to the mean error. The weighting factors are due to the age of the evidence and *M*'s own certainty about each computed error. Both are reasonable.

4.2 Evaluation of the predictability

The competence measure given above describes the mean behaviour of the other car. It does not account for outliers or quick changes in the other's behaviour. As a consequence, this paper proposes to include a feedback control that is sensitive to outliers. The idea is that the car *M* behaves as if it received three times the same message from three other cars that have about the same trust value. Would it decide right, then? So, it predicts its decision based on the message of the other car *A* and finally verifies whether the predicted decision was right.

Figure 2 shows the computation schema with the predictor and the controller. There $\omega_{A,i}^M$ is the total trust opinion about *A* regarding the information type *i*. It is taken from the output of the trust system and fed back here to control the prediction. So, the predictability component of the trust model implements a feedback loop together with a controller to benefit from the methods of control theory.

The prediction goes in the same way as the decision-making (e.g., as described in Section 5.3). In general, the decision-making is represented by an operator *D* from a set of opinions $\{\omega_v^1,...,\omega_v^N\}$ about a certain statement *v* to a tuple, which contains the decision which value to believe (may also be none) and the associated opinion ω_v^M of the car *M*. More specifically, if the car *M* applies the operator *D* on three pieces of *A*'s opinion (ω_v^A), the result can be to accept *A*'s opinion or not. This can then be compared with *M*'s real opinion about *v*.

If the decision to accept or not accept *A*'s value was right, then *A* was predictable. So, $\omega_{P_i,A}^M$ is right; there is no prediction error (*e* = 0). Otherwise, if the decision to accept or not accept was wrong, *M* must compute a prediction error *e* and readjust the relation $c = \frac{b_+}{b_-}$ between the belief and disbelief of $\omega_{P_i,A}^M$ to obtain a new $\omega_{P_i,A}^M$.

The prediction error is not simply 0 or 1 if the decision was right or wrong. Instead, the degree of the error depends on the predicted belief ($\hat{b}_{v^A}^M$) and *M*'s own belief ($b_{v^A}^M$) about the value v^A that is proposed by *A*:[2]

$$e = b_{v^A}^M - \hat{b}_{v^A}^M.$$

The belief in the value of *A* is taken instead of that in the value of *M*, because the high belief of *A* in the value v^A let to the wrong decision. If it would have been a bit lower, the prediction could have been not to believe at all.

From the above equation, a more precise condition for the correctness of the prediction can be derived. Since the strength of the own belief has an influence on the sign of the error, it must be considered as well. First, if *A*'s value has been accepted and

v^A is different from v^M and $e < 0$, then the decision to accept A's value was wrong. Second, if A's value has been rejected and v^A is equal to v^M and $e > 0$, then the decision to reject A's value was wrong. In these two cases, M's opinion about A's predictability must be readjusted.

For the readjustment, this paper proposes a PI controller (proportional and integral controller) with the output

$$y_k = K_P e_k + K_I \sum_{i=1}^{k} e_i,$$

where k is the step counter. k is increased with every piece of evidence. (Thus, it represents the number of evaluated reports). With the controller output, a new c for the next piece of evidence can be computed by

$$c_{k+1} = \begin{cases} (1 + y_k) c_k & \text{for } y_k \geq 0 \\ \dfrac{1}{1 - y_k} c_k & \text{for } y_k < 0. \end{cases}$$

To derive $b_{P_i,A}^M$ from c, the ignorance $u_{P_i,A}^M$ is necessary. It reflects the amount of evidence; so it is known and always equal to $u_{C_i,A}^M$. With both together, the new predictability opinion can be obtained with

$$\omega_{P_i,A}^M = \left(\left(\frac{c(1 - u_{P_i,A}^M)}{1 + c}, \frac{1 - u_{P_i,A}^M}{1 + c} \right)^T, u_{P_i,A}^M \right).$$

The predictability opinion should make the model more sensitive to outliers in the other car's behaviour. So, the control must be more sensitive to wrongly accepted than to wrongly rejected reports. Therefore, the simulation for this paper has been performed with $K_P = 0.2$ and $K_I = 0.1$ if the report has been wrongly accepted and $K_P = 0.1$ and $K_I = 0.1$ in the other case. This is the only place, where the model distinguishes between the trust constitution and the trust reduction. (In the competence component, positive and negative experiences have the same influence on the trust development).

When choosing these controller variables, the following considerations must be taken into account: high values, especially for K_p, make the system very responsive to single observations. This counteracts the competence component. Small values, in contrast, lead to slow system dynamics similar to that of the competence component. All in all, a balance must be found that meets the expected dynamics of the interaction partners.

4.3 Evaluation of the local reputation

Reputation reflects what people say about another person or object. It condenses the general opinion of a group or society – not only of one or two individuals. So in technical systems, it is mostly realised as a feature of the network, which collects all recommendations to compute a reputation from them. For example, eBay's feedback system (eBay Inc., 2010) and the beta reputation system (Jøsang and Ismail, 2002) do so. This paper focuses on individual-level trust, though. Here, the individual collects

recommendations from others to build an opinion about another car's reputation. This individual reputation opinion may be different from the 'real' global reputation. So, the authors call it *local reputation*, when it is helpful to distinguish both.

Reputation systems have to face the problem of statistically dependent recommendations (Jøsang et al., 2006; Ramchurn et al., 2004). This is especially necessary in decentralised systems. The proposed reputation mechanism takes three measures to handle this: first, only the most recent recommendation of one source about a third car is saved. Second, a car should only spread a recommendation about cars, with which it had at least one direct interaction. And third – which is the most important measure – reputation plays only a subordinate role in the opinion formation as detailed in the next section.

With these measures, the reputation can be computed as in other reputation systems (see Figure 3): every recommendation is discounted with the opinion about its source. So, the more the car M trusts the car B regarding the given information type, the more weight B's opinion gains in the fusion process. This is done by the base rate sensitive discounting as discussed in Jøsang et al. (2006). This operation weights the opinion of trustworthy cars higher in the subsequent opinion fusion. Further, the reputation opinions are weighted according to their age as described in Section 4.6. This whole process finally ends in a single estimate $\omega_{R_i,A}$ about A's reputation regarding the information type i.

4.4 Combining everything together in a trust value

With the previous three sections, the car M has opinions about three properties of A: its competence, its predictability and its reputation. It can compute these opinions for every information type. This section describes how these opinions are transformed into a trust value which quantifies the sender-related uncertainty of values received from A.

When looking at these three properties, competence and predictability are obtained by the car itself from direct interactions. As both are based on the same evidence, they must be combined with the averaging fusion. This is shown in the upper left branch of Figure 4.

However, only little evidence exists to evaluate competence and predictability. In contrast, recommendations of many cars are usually available resulting in an opinion about another car's reputation with small ignorance. But reputation has the disadvantage that it is prone to attacks of others (see the requirements in Section 1.2). This is why a fusion method is needed that takes advantage of the dense network offered by the reputation mechanism and of the reliability that comes along with the competence and predictability opinions. The priority fusion as introduced in Section 3.2 fits very well here. As long as a car has no own experiences, it should take the reputation. The more experiences it has, the less reputation should influence the own trust. So, the own experience – gained by the car itself as a combination of competence and predictability – is the prioritised opinion and the local reputation – given as recommendations from other cars – is the discriminated opinion.

Now, the car M has a trust opinion originating only from the evidence regarding the given information type i. This opinion is useful, if enough evidence is available. But if there is not enough evidence available, it would be reasonable to see if the car M has evidence about the car A regarding other information types. Considering this evidence

seems useful. (It assumes, though, that all sensors of a car have a similar quality). As a consequence, the computations that have been introduced up to here are performed for every type of information ι. This results in an intermediate opinion $\omega_{D_\iota,A}^M$ for all these types of information. The intermediate opinions of all types $\iota \neq i$ are fused cumulatively. Weighting factors in this fusion determine how strong the evidence of other information types should influence the trust for the information type i. (For example, the similarity between the type i and the type ι may determine what factor is chosen). The result is then taken as the discriminative opinion in a priority fusion with the intermediate opinion $\omega_{D_i,A}^M$ of type i. This finally results in M's opinion $\omega_{A_i}^M$ about the trustworthiness of A regarding the type of information i. As long as evidence of type i is available, this evidence dominates the trust. Otherwise, the mean evidence from all other information types is taken.

At this point, the car M has computed its opinion about the trustworthiness of the other car A. An opinion is a two-dimensional magnitude, though. Sometimes – e.g., to make a decision – a one-dimensional trust value is preferable. So, the final computation step in Figure 4 transforms the trust opinion $\omega_{A_i}^M$ in a trust value $t_{A_i}^M$. Such a transformation already exists: e.g., the expectation value of subjective logic. Another one with roots in information theory is proposed in the following section.

4.5   Opinions and the binary error and erasure channel

The opinion of the car M about the trustworthiness of another car A describes whether M expects that A sends reliable information or not. The belief b^+ expresses the degree of evidence whether M expects to receive correct information, the disbelief b_- whether M expects to receive wrong information. The ignorance u expresses the absence of an expectation (due to a lack of evidence). This setting is similar to the setting in the model of the binary error and erasure channel as shown in Figure 5 (Cover and Thomas, 2006). Its channel capacity

$$C = (1-u)\left(1 - H_b\left(\frac{b_+}{1-u}\right)\right)$$

with

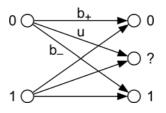$$H_b(x) = -x\log_2(x) - (1-x)\log_2(1-x)$$

expresses how much information can be correctly transmitted through such a channel. The trust opinion can be interpreted as the description of such a channel. Then, the equation above maps the trust opinion $\omega_{A_i}^M = ((b_+, b_-)^T, u)$ to a trust value $t_{A_i}^M$. Only one modification must be included: The channel capacity is symmetric with respect to b^+ and b_-; both could be exchanged without changing C. For trust development, b_+ and b_- cannot be exchanged. As a consequence, a negative sign must be placed in front of the above equation, if the disbelief is higher than the belief. With it, the trust value has a range from -1 to 1. It can be linearly mapped to the range $[0, 1]$ in the end.

What is the advantage of using the channel capacity C over the (commonly used) expectation value E? The ratio between both

$$\frac{C}{E} = \frac{b_+ + b_-}{b_+ - b_-}\left(1 - H_b\left(\frac{b_+}{1-u}\right)\right)$$

has the following properties: It remains constant if the ration $\frac{b_+}{b_-}$ remains constant with an arbitrary ignorance. Otherwise, the channel capacity punishes high disbelief and rewards low disbelief stronger than the expectation value. As Section 6 shows, this property makes it easier to separate agents with different error rates. So, it helps in the decision process.

Figure 5 The binary error and erasure channel

Note: The disbelief b_- in the trustworthiness of another car A can be interpreted as an error rate of a transmission channel, the ignorance u as an erasure rate.

4.6 Temporal weighting of opinions

Recent pieces of evidence should influence a present decision more than old ones. This lets a system better adapt to changes in the environment: trustworthy cars can become untrustworthy, and untrustworthy cars must have a chance to become trustworthy. So, a car must weight new opinions higher than older ones, when it fuses them. This weight must depend on the age of the opinion. For the weighting operation, the evidence scaling (Section 3.2) fits well when the opinions are fused with the cumulative fusion.

The mapping from the age of an opinion to its weight should have the following properties:

1   at age $t = 0$, the mapping should be 1

2   for $t \to \infty$, it should tend to zero

3   in between, it should be monotonic.

Three mappings with these properties are common: from the set of exponential functions, only $w_t = b^{\frac{t}{1year}}$ (and its equivalent formulations) have the required properties. The rational polynomials $w_t = \frac{a}{t+a}$ and the piecewise functions
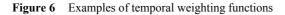
$$w_t = \begin{cases} mt + 1 & \text{for } t < -\dfrac{1}{m} \\ 0 & \text{else} \end{cases}$$
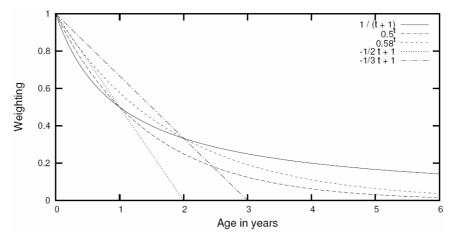
with $m < 0$ fit here well, too.

Figure 6 compares these weighting functions. It shows that the rational polynomial decreases faster than the exponential functions in the beginning, but becomes flatter,

then, and tends more slowly to zero in the end. The linear function with $m = -1/3$ is a good approximation for the exponential functions. The described properties are already significant within the lifetime of a car.

All in all, the rational polynomial does not suppress old evidence sufficiently. The exponential functions show a suitable form. In addition, they can be used in a recursively updating algorithm, because $w(t) = w(\Delta t_1) \cdot w(\Delta t_2)$ for $t = \Delta t_1 + \Delta t_2$. (A recursive update algorithm can help to save memory). On the downside, they are computationally expensive. In contrast, the linear functions can be computed quickly and can be adjusted to have a sufficiently good form. They cannot be updated recursively, though.

Figure 6   Examples of temporal weighting functions

Notes: The exponential functions (dashed lines) decrease slower in the beginning, but show lower weighting values in the end compared to the rational polynomial (solid line). The linear functions approximate the exponential functions from a global view.

5   Method of the model evaluation

In order to test the performance of the trust model the authors have created an environment that simulates the cars' driving and communication behaviour. The simulation code is publicly available on our project web page (Bamberger, 2010) for extension or verification by the research community.

5.1   *The social structure in the simulation*

As pointed out in the introduction, a simulation environment for trust models must reflect the underlying social structure. In the case of a vehicular network, this structure is constituted by the movements of the cars and their communication. Because these movements are not random, they lead to regular meetings of cars. For example, some cars could use the same parking site and have similar trip times. The regular meetings are important so that trust can develop.

We created a suitable traffic scenario, which simulates an urban population of a town with about 15,000 inhabitants. It only contains trips to work and back home; however, these incorporate the majority of the regular ones and also allow random meetings. The duration of 23 weeks is pretty long compared to other simulations but it is necessary because developing trust takes some time. The simulation is performed with the microscopic traffic simulator SUMO (Krajzewicz et al., 2002).

The cars' communication is then simulated with the network simulator Shawn (Kröller et al., 2005). It is very efficient and suitable for our case, because it simulates the effects caused by a phenomenon, not the phenomenon itself. When cars meet, they only exchange own observations about the environment or recommendations that origin from own experiences. Information of a third party is not yet relayed.

Both together, the cars' movements and their communication, represent the social structure.

5.2   The information model in the simulation

As described in Section 1, there is much information vehicles can exchange. In the simulation, every car is equipped with a sensor that asks its environment for such events. An event is of one of three different information types (for long-, medium- and short-term information). The value of every event is an integer in the range of 0 to 4. The events are randomly generated on the map and exist only for a limited time. Because of memory restrictions, the amount of events has been limited. In reality, if all the possible kinds of information mentioned in the introduction are considered, more events can be expected; this would then increase the trust model's performance.

5.3   The cars' behaviour in the simulation

Cars have sensors and every sensor continuously checks for events in the environment. If it finds one, it reports it to the car, but adds a Gaussian random error first. The underlying Gaussian distribution is mean free with a configurable variance (see the table in Figure 9). There is a separate random variable for every type of information.
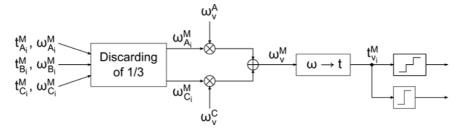
Every observation is saved in the knowledge base and spread out to other cars then. A car only sends own observations, not reports of other cars. This makes the information integration and decision simpler. (This paper focuses on the trust model only).

Reports from other cars are also saved in the knowledge base. The process to determine knowledge from received information must incorporate the sender's trustworthiness. A car does this in a way that is illustrated in Figure 7. It is similar to the method in Zhang et al. (2006). First, the car takes all messages regarding a specific event and discards the third of them that comes from the senders with the lowest trust values. Afterwards, every remaining opinion is discounted with the trust value of its sender regarding the given information type. This has the effect that well-known and well behaving vehicles get a higher weight. The discounting is similar to the base rate sensitive discounting as given in Section 3.2, but depends on the mapping from the opinion to the trust value (see Section 4.5). Finally, the opinions about the event (discounted reports and own observations) are combined into one using cumulative fusion.

All reports and observations are kept in the knowledge base; they are never forgotten in the simulation. This makes it possible for a car to change its opinion about an event.

The trust value is computed every time it is needed from scratch. This way, a previous trust judgement can be revised when more evidence is available. This is in alignment with the observation that artificial reasoning is non-monotonic (Pearl, 1988). This means, when new evidence comes in, it can lead to a completely different judgement of the situation.

Figure 7   Signal flow of the decision process

Notes: First, only the most trustworthy cars are extracted. Then, their opinions are discounted and combined. This also includes M's own observations. The result is M's opinion about v. It can be transformed in the trust value t_v^M for decision-making.

6   Evaluation results and discussion

The simulation gives insights of the social structure and the trust model. Figure 8 illustrates how often cars meet each other. The values are mean values over all cars; they represent a prototypical car. For example, after 23 weeks this prototypical car has met about five other cars at least 20 times. These are something like 'good friends', which it meets once a week. The figure also shows how the system settles down over time.

Figure 9 visualises the competence component of the trust model. In the simulation, cars have three different sensors for three types of information (see Section 5.2). The left plot is for one, the right plot for another sensor. The sensors have different error variances. To show the effect of the sensor error on the trust development, four types of cars have been configured. Every car of the same type has the same error variance for each of the sensors. The table in Figure 9 gives the error variances for both sensors. The plots show that the trust model in the receiver can well distinguish cars with good sensors from cars with bad sensors. And it can develop trust in each sensor separately, although the evidence of both sensors is always combined in the last priority fusion as described in Section 4.4. To show this, only trust values are taken that are based on at least six pieces of evidence. In contrast, if only little evidence is available for one sensor and much evidence for the other, then the simulation logs show that indeed the evidence for the other sensor dominates also the trust value for the first one. This is as intended.

Both plots also show that cars of the different types can equally well judge about the trustworthiness of the sender. This is in contrast to the previous simulation results in (Bamberger et al., 2010). The reason for the change is an improved decision module in the car. This new decision module finds a better balance between own observations and received reports. It can better cope with bad sensors in the own car.
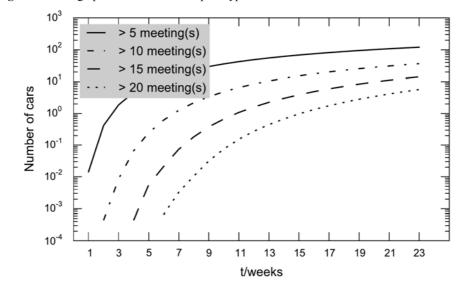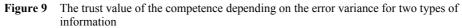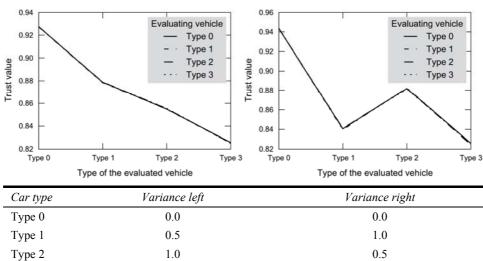
Figure 8 This graph shows how often a prototypical car meets other cars

Note: For example, on average a car meets about two cars at least once a week.

Figure 9 The trust value of the competence depending on the error variance for two types of information

Car type	Variance left	Variance right
Type 0	0.0	0.0
Type 1	0.5	1.0
Type 2	1.0	0.5
Type 3	1.5	1.5

Notes: Four types of cars have two types of sensors each. The different line styles
indicate the type of the receiving car. The *x*-axis shows the type of the sending
car. A receiving car can well distinguish the different error variances of the
sender's sensors based only on the competence component. The table gives the
error variances of both sensors for all car types.

Figure 10 illustrates what happens if the channel capacity is taken for the trust value computation instead of the expectation value (see Section 4.5). For these plots, the same simulation configuration is taken as for those in Figure 9. Only the algorithm for the trust value computation has been exchanged. Two main differences can be read from the figures. First, the channel capacity assigns a wider range of trust values to the different types of cars; while the range is only about 0.12 with the expectation value, it is about 0.25 when the channel capacity is used. This makes it easier for a decision algorithm to separate trustworthy from untrustworthy information and to apply a threshold. Second, expectation values seem to be quite high, even for high error variances. As a consequence, the trust value can hardly reflect high trust in good cars with a long interaction history. All in all, the channel capacity exploits the range from 0 to 1 of the trust value better for the given application than the expectation value. On the downside, the value of the channel capacity has a fixed base rate of 0.5. (This is the default trust in absence of any evidence). This is quite low compared to the trust values shown in Figure 10. For the expectation value, the base rate can be chosen freely; it should be around the trust values for medium quality cars.
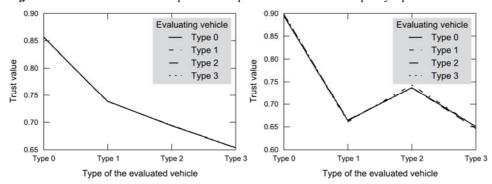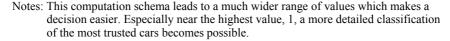
Figure 10 Trust values of the competence computed with the channel capacity equation

Notes: This computation schema leads to a much wider range of values which makes a decision easier. Especially near the highest value, 1, a more detailed classification of the most trusted cars becomes possible.

To visualise the predictability component, single sender-receiver combinations must be considered. Mean values cannot well describe its behaviour; that was its purpose. Figure 11 shows for one receiving car how it develops trust in one specific sending car. This sending car has been configured to mimic a manipulation. During the simulation, it suddenly sends manipulated messages only. The plot on the left hand side shows the trust development with the competence component only. In contrast, both, the competence and the predictability component, have been used for the plot on the right hand side. The plots show that the trust decreases faster if the predictability component is active. This effect would be even stronger, if more positive evidence would have been available before the sender got manipulated. The reason for this is that the competence component reflects a mean value and means change the more slowly the more values were included before. Furthermore, the competence component treats positive and negative evidence equally; in contrast, the predictability component is more sensitive to negative than to positive evidence.
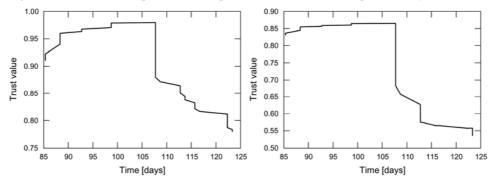
Figure 11 The trust development in a manipulated car without and with predictability

Notes: The plots show how one car develops trust in a specific other car over time. This other car suddenly begins to send only wrong information. The adaptation to this new behaviour happens faster when the predictability component is active as the right plot shows.

7 Conclusions and future work

In this article, we systematically modelled trust as a social mechanism to handle the uncertainty in the information exchange. It integrates different components to make it suitable for different situations. The model assigns one trust value to every combination of sender and information type. This is necessary since different types of information may come from sensors with different error variances. The trust model also features the capacity formula of the binary error and erasure channel as an appropriate transformation of a two-dimensional trust opinion in a one-dimensional trust value. Furthermore, we proposed a traffic simulation that partly represents the 'social structure' between the cars. Thus, it is suited to investigate a social phenomenon like trust. The simulation results show that the competence component reflects the mean performance of a car, while the predictability component helps to faster react on sudden changes in the other's behaviour.

Currently, we are about to apply learning algorithms to the trust development. We want to handle even more complex trust situations and we are interested in the dynamical behaviour of those algorithms. In addition, we want to generalise the trust model to various kinds of cognitive systems with different tasks.

Acknowledgements

References

Bamberger, W. (2010) 'Fidens: trust for cognitive vehicles', Project web page, available at http://www.ldv.ei.tum.de/fidens.

Bamberger, W., Schlittenlacher, J. and Diepold, K. (2010) 'A trust model for intervehicular communication based on belief theory', in *2010 IEEE Second International Conference on Social Computing (SocialCom)*, pp.73–80, doi:10.1109/SocialCom.2010.20.

Brin, S. and Page, L. (1998) 'The anatomy of a large-scale hypertextual web search engine', in *Computer Networks and ISDN Systems*, Vol. 30, pp.107–117, doi:10.1016/S0169-7552(98) 00110-X, Special issue with the *Proceedings of the Seventh International World Wide Web Conference*.

CAR 2 CAR Communication Consortium (2007) 'Manifesto', Technical report, version 1.1, August, available at http://www.car-to-car.org/index.php?id=31.

Cover, T.M. and Thomas, J.A. (2006) *Elements of Information Theory*, Wiley, Hoboken, NJ, USA, ISBN 978-0-471-24195-9.

eBay Inc. (2010) 'All about feedback', available at http://pages.ebay.com/help/feedback/allaboutfeedback.html (accessed on 2010/03/02).

Golle, P., Greene, D. and Staddon, J. (2004) 'Detecting and correcting malicious data in VANETs', in *VANET '04: Proceedings of the 1st ACM International Workshop on Vehicular Ad Hoc Networks*, pp.29–37, ACM, New York, NY, USA, doi:10.1145/1023875.1023881.

Jøsang, A. (2001) 'A logic for uncertain probabilities', in *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, June, Vol. 9, No. 3, pp.279–311, doi:10.1142/S0218488501000831.

Jøsang, A. (2007) 'Probabilistic logic under uncertainty', in Gudmundsson, J. and Jay, B. (Eds.): *Thirteenth Computing: The Australasian Theory Symposium (CATS2007)*, CRPIT, Vol. 65, pp.101–110, Australian Computer Society, Inc., Darlinghurst, Australia, available at http://crpit.com/confpapers/CRPITV65Josang.pdf.

Jøsang, A. and Ismail, R. (2002) 'The beta reputation system', in *15th Bled Electronic Commerce Conference*, available at http://persons.unik.no/josang/papers/JI2002-Bled.pdf.

Jøsang, A., Marsh, S. and Pope, S. (2006) 'Exploring different types of trust propagation', in *Proceedings of 4th International Conference on Trust Management, iTrust 2006*, pp.179–192, Springer, Berlin, doi:10.1007/11755593 14.

Krajzewicz, D., Hertkorn, G., Rössel, C. and Wagner, P. (2002) 'SUMO (simulation of urban mobility): an open-source traffic simulation', in *Proceedings of the 4th Middle East Symposium on Simulation and Modelling (MESM 2002)*, pp.183–187, SCS European Publishing House, available at http://www.eurosis.org/cms/?q=taxonomy/term/28.

Kranz, M. (2008) 'Foundations of and applications for car-to-car communication', in *RADCOM 2008 Radar, Communication and Measurement, Gerotron Communication GmbH*, Martinsried, Germany, available at http://elib.dlr.de/53824.

Kröller, A., Pfisterer, D., Buschmann, C., Fekete, S.P. and Fischer, S. (2005) 'Shawn: a new approach to simulating wireless sensor networks', in *Proceedings of the 2005 Design, Analysis, and Simulation of Distributed Systems Symposium (DASD'05)*, pp.117–124, available at http://arxiv.org/abs/cs/0502003.

Matheus, K., Morich, R. and Lübke, A. (2004) 'Economic background of car-to-car communication', in *Proceedings of the 2nd Braunschweiger Symposium Informationssysteme für mobile Anwendungen (IMA 2004)*, Braunschweig, Germany.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, CA, USA, ISBN 0-934613-73-7.

Ramchurn, S.D., Huynh, D. and Jennings, N.R. (2004) 'Trust in multi-agent systems', in *The Knowledge Engineering Review*, Vol. 19, No. 1, pp.1–25, doi:10.1017/S0269888904000116.

Raya, M. and Hubaux, J.P. (2007) 'Securing vehicular ad hoc networks', in *Journal of Computer Security*, Vol. 15, No. 1, pp.39–68, ISSN 0926-227X.

Raya, M., Papadimitratos, P., Gligor, V. and Hubaux, J.P. (2008) 'On data-centric trust establishment in ephemeral ad hoc networks', in *INFOCOM 2008. The 27th Conference on Computer Communications*, IEEE, April, pp.1238–1246, doi:10.1109/INFOCOM.2008.180.

Sabater, J. and Sierra, C. (2005) 'Review on computational trust and reputation models', in *Artificial Intelligence Review*, Vol. 24, No. 1, pp.33–60, doi:10.1007/s10462-004-0041-5.

Shafer, G. (1976) *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, ISBN 0-691-08175-1.

Stiller, C., Färber, G. and Kammel, S. (2007) 'Cooperative cognitive automobiles', in *Intelligent Vehicles Symposium, 2007 IEEE*, June, pp.215–220, doi:10.1109/IVS.2007.4290117.

Zhang, W., Das, S.K. and Liu, Y. (2006) 'A trust based framework for secure data aggregation in wireless sensor networks', in Das, S.K. (Ed.): *Proceedings of the 2006 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks (SECON '06)*, IEEE, September, Vol. 1, pp.60–69, doi:10.1109/SAHCN.2006.288409.

Notes

1   The definition of the priority fusion differs from that in Bamberger et al. (2010); it scales the evidence of the discriminated opinion. This new definition has better theoretical properties, while the results are similar to our previous proposition.

2   In the equation for the prediction error, the order of operands was wrong in Bamberger et al. (2010). (It was only wrong in the paper, not in the simulation).