Distributed data fusion: application to confidence management in vehicular networks

Véronique Cherfaoui, Thierry Denoeux
Heudiasyc UMR CNRS 6599
Université de Technologie de Compiègne, France
veronique.cherfaoui@hds.utc.fr
thierry.denoeux@hds.utc.fr

Zohra Leïla Cherfi
LTN INRETS
Institut National de Recherche sur les Transports et leur Sécurité, Arcueil, France
zohra.cherfi@inrets.fr

Abstract – This paper presents a method for the management of information dissemination in a vehicular network (VANET). Due to the particularities of the application (ad hoc network, dynamical nodes, broadcast messages), an algorithm has been developed to fuse and combine data in distributed systems. Matching spatial information is made easier by the use of a numerical map as support of a database. A model of confidence management based on the belief function framework is then described considering spatial dispersion of data sources, delays due to the multi-hop transmission and dependency between sources. Preliminary results are presented based on simulated messages referenced on a real map data.

Keywords: distributed data fusion, confidence, belief functions, ad hoc network application.

1 Introduction

In the last years, more and more communication devices have been embedded in vehicles. Many applications based on wireless communication have been developed in which the vehicles are the nodes of an ad hoc network called VANET. In the VANET, all vehicles broadcast messages and each vehicle has knowledge about its neighborhood only through the messages it receives. Most research papers deal with communication protocols, routing and congestion problems. Due to the nature of applications (driving assistance systems or emergency braking alert) recent works have been dedicated to the security mechanism in order to avoid malicious node intercepting, modifying or sending erroneous data [6][8]. Supposing these problems to be (partially) solved, we propose a method to manage and exploit message information from the receiver node point of view.

In this study we consider messages regarding safety such as accident, reduced visibility, traffic jam, etc., and we consider car-to-car (C2C) communications. It is the context of the Safespot project [18][19] and the cognitive car [9]. Each car (node) is able to detect, localize, date and characterize an event and, if necessary, broadcast it in a message. Due to the multi-hop transmission protocol, the distance between sender and receiver nodes is not limited by the transmission power of antennas. In order to localize and date the content of a message, we assume that each node is equipped with a Global Positioning System (GPS) receiver. When a node receives a message, it updates its database and, if necessary it broadcasts its updated information.

The problem of information dissemination, i.e., proposing a strategy to broadcast information is not addressed in this paper. Indeed, the road traffic can be high, the bandwidth is limited and the number of exchanging messages could have to be reduced. Different strategies have been proposed in the literature [5][1][16] concerning this problem. It should also be remarked that algorithms for combining and fuse data are very different from algorithms developed in infrastructure vehicle (V2I) communication applications. In this latter case, a centralized module combines collected data and disseminates global information.

The objective of the work reported in this paper is to develop a methodology for combining data included in messages arriving from other nodes. Since data are uncertain and could be the result of processing disseminated data, we focus on confidence management in a distributed and dynamical context. The confidence could be exploited to provide the driver with relevant information and/or to decide about the transmission of the result in the network. This contribution is intended to be a part of the information dissemination strategy to be developed in future work.

This work is based on the use of belief functions to combine degrees of confidence about events reported in exchanged messages. We first define the attributes of each message and then describe the methodology to combine data coming from distributed, dynamical and asynchronous sources.
2 Background on Belief Functions

The transferable belief model TBM [14] is a model to represent quantified beliefs based on belief functions [13]. It has the advantage of being able to explicitly represent uncertainty about an event. It takes into account what remains unknown and represents what is already known.

2.1 Knowledge Representation

Let $\Omega$ be a finite set of all possible solutions of a problem. $\Omega$ is called the frame of discernment (also called state space); it is composed of mutually exclusive elements. The knowledge held by source can be quantified by a belief function defined from the power set $2^\Omega$ to $[0,1]$. Belief functions can be expressed in several forms: the basic belief assignment (bba) denoted $m$, the credibility function $\mathcal{B}$, the plausibility function $\mathcal{P}$, and the commonality function $\mathcal{Q}$ which are in one-to-one correspondence. We recall that $m(A)$ quantifies the part of belief that is restricted to the proposition “the solution is in $A \subseteq \Omega$” and satisfies: $\sum_{A \subseteq \Omega} m(A) = 1$.

Thus, a bba can support a set $A \subseteq \Omega$ without supporting any sub-proposition of $A$, which allows accounting for partial knowledge. The complete notation of a belief function is: $m_{S,t} \{X\} \{BC\}_{S,t}(A) \subseteq \Omega$ where $S$ is the information source, $t$ the time of the event, $\Omega$ the frame of discernment, $A$ a parameter which takes value in $\Omega$ and $BC$ the evidential corpus or knowledge base. This formulation represents the degree of belief allocated by the source $S$ at time $t$ to the hypothesis that $X$ belong to $A$. The notation is simplified in the following paragraph to clarify the combination formulae. In Section 4, the complete notation including the time, source and parameter will be used.

Smets introduced the notion of open world where $\Omega$ is not exhaustive; this is quantified by a non zero value of $m(\Omega)$. Other functions can be calculated from the bba $m$ using the following formulas:

Credibility function: $\mathcal{B}(A) = \sum_{B \subseteq \Omega} m(B)$,

Plausibility function: $\mathcal{P}(A) = \sum_{B \subseteq \Omega} m(B)$,

Commonality function: $\mathcal{Q}(A) = \sum_{B \subseteq \Omega} m(B)$.

Another function that can be computed from $q$ is the conjunctive weight function [15] defined by: $w(A) = \prod_{B \subseteq A} q(B)^{-1}(1-q(B))^{-1}$.

The $w$ function is well-defined if $m$ is non dogmatic, i.e. if $m(\Omega) > 0$. Functions $\mathcal{B}$, $\mathcal{P}$, $w$ and $m$ are in one-to-one correspondence. In particular, formula to recover $m$ from $w$ are given in [15] and [4].

2.2 Information Fusion

Let $n$ distinct pieces of evidence be defined over a common frame of discernment and quantified by bbas $m_1^\Omega \cdots m_n^\Omega$. They may be combined using a suitable operator. The most common ones are the conjunctive and disjunctive rules of combination defined, respectively, as:

$$m_{\cap}^\Omega(A) = \sum_{A \cap \cdots \cap A = A} m_1^\Omega(A_1) \cdots m_n^\Omega(A_n)$$

$$m_{\cup}^\Omega(A) = \sum_{A \cup \cdots \cup A = A} m_1^\Omega(A_1) \cdots m_n^\Omega(A_n) .$$

The resulting bbas should be normalized under the closed world assumption. Dempster’s rule [2] denoted by $\oplus$ normalises the result of the conjunctive rule with $K = \frac{1}{1-m(\Omega)}$ and sets the mass on the empty set to 0.

The conjunctive and disjunctive rules of combination assume the independence of the data sources. In [3] and [4] Denoeux introduced the cautious rule of combination (denoted by $\Phi$) to combine dependent data. This rule has the advantage of avoiding double-counting of common evidence when combining non distinct bbas. In particular, the combination of a bba with itself yields the same bba: $m = m \Phi m$ (idempotence property). The cautious rule of combination can be easily computed by taking the minimum of conjunctive weights : with obvious notations, $w_1 \Phi w_2 = w_1 \land w_2$, where $\land$ denotes the minimum operator.

2.3 Reliability and Discounting Factor

The belief function framework makes it possible to model the user’s opinion about the reliability of a source [7]. The idea is to weight more heavily the opinions of the best source, and conversely for the less reliable ones. The result is a discounting of the bba $m^\Omega$ produced by the source, resulting in a new bba $m_{\alpha}^\Omega$ defined by:

$$m_{\alpha}^\Omega(A) = \alpha \cdot m^\Omega(A), \quad \forall A \subseteq \Omega, A \neq \Omega$$

$$m_{\alpha}^\Omega(\Omega) = 1 - \alpha + \alpha \cdot m^\Omega(\Omega).$$

The discounting factor $\alpha$ can be regarded as the degree of trust assigned to the sensor.

3 Exchanged Data

3.1 Level of Information

Safety applications in VANET are being investigated in order to increase the vehicle visibility area and produce useful information in view of developing ADAS.
(Advanced Driver Assistance System) functions. The level of exchanged data depends on the applications: the concept of cognitive car [9] assumes that vehicles communicate in crossing roads to avoid collision. Traffic information applications are based on the vehicle positions and speed exchange [12].

We propose to exchange data concerning a set of events in order to increase the “visibility” of the driver and to allow the anticipation of dangerous situations. These events are classified into three categories:
- Static (or slowly evolving) and localized events such as Accident (AC), working area (WA), Dangerous Object (DO);
- Dynamical (quickly evolving) and localized events such as an animal on the road (AN), a countersense vehicle (CV) or a dangerous vehicle (DV);
- Slowly evolving and diffuse events such as low visibility (LV), traffic-jam or congestion zone (CZ), low adherence area (LA). These events concern a whole geographical area.

In this work, it is assumed that vehicles are equipped with systems able to detect these events.

3.2 Spatial and Temporal References

When a node receives a message, it has to decide whether it is relevant according to the node location and the node itinerary. Knowing that an event is geo-localized with GPS, and assuming the node has a numerical map, it is possible to associate an event with a road segment. A road segment is an entity in a Geographical Information System (GIS) database. Each road segment is determined by a unique Road-ID. Figure 1 describes the geometric definition of a road segment: it is connected with other segments at the origin and end extremities.

![Figure 1: definition on a road segment in GIS](image)

This approach has three main advantages:
- The spatial data association between two events is made easier;
- The space representation by roads is discrete. Consequently, two messages geo-localized on the same Road-ID concern the same event.
- It is possible to assign prior knowledge to each road segment. For example, frequent fog reported in an area can be associated to segments in this area.

As mentioned in the introduction, the time between the creation and reception of a message can be higher than transmission delay. This is due to the multi-hop and retransmission capabilities of communicating cars. Consequently, two attributes should be defined for dating an event: one for time stamping the event when it was detected and one for specifying the date of updating if the message was created from the combination of other messages.

3.3 Confidence Attributes

In order to analyze a road situation according to set of events, we study the confidence of these events according to the confidence assigned by the vehicle that has detected the event, the redundancy of messages concerning this event and the date and location attributes.

We propose to model the confidence with belief functions in order to take into account the uncertainty of initial data. The operators defined in the belief function framework can be applied to compute a unique belief function by combining belief masses coming from many sources. A message describing the event $ev$ is assumed to contain a mass function on $\Omega = \{0, 1\}$. The value 0 represents $\neg ev$ and 1 represents $ev$. The mass function $m$ can then be represented as a quadruplet $[m(\emptyset), m(\{0\}), m(\{1\}), m(\Omega)]$. The value $m(\Omega)$ is interpreted as the degree of doubt and the value $m(\emptyset)$ represent the degree of conflict between sources.

There are two ways to initialize the mass function. First, when a vehicle $vehID$ detects an event $ev$ at time $t$ on road $roadID$, it computes a degree of certainty $d$. This value is used to define the bba $m_{vehID,roadID,t}^{\Omega}(ev)$ as:

$$m_{vehID,roadID,t}^{\Omega}(ev)(\emptyset) = 0$$
$$m_{vehID,roadID,t}^{\Omega}(ev)(0) = 0$$
$$m_{vehID,roadID,t}^{\Omega}(ev)(1) = d$$
$$m_{vehID,roadID,t}^{\Omega}(ev)(\Omega) = 1 - d$$

When a vehicle $vehID$ predicts an event $ev$ on a road segment $roadID$ and it does not detect it, it assigns a confidence value $d'$ to $\neg ev$ and builds the following mass function for $ev$:

$$m_{vehID,roadID,t}^{\Omega}(ev)(\emptyset) = 0$$
$$m_{vehID,roadID,t}^{\Omega}(ev)(0) = d'$$
$$m_{vehID,roadID,t}^{\Omega}(ev)(1) = 0$$
$$m_{vehID,roadID,t}^{\Omega}(ev)(\Omega) = 1 - d'$$

The confidence values $d$ and $d'$ are set based on the reliability of the detection system (driver observation, sensor processing).

Our aim is to improve the level of knowledge from exchanged messages. The belief of the distributed sources is expressed by their mass functions. Distributed data
fusion then consists in combining these masses with the appropriate operator. The conjunctive rule will be used when assuming the independence of messages, whereas the cautious rule will be applied in case of dependent messages. To determine which operator should be applied when combining information contained in a message with other information, an attribute src is set to 1 when the message is original and set to 0 when it results from the combination of other messages.

The content of message is summarized in Table 1.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ev</td>
<td>Type of event</td>
</tr>
<tr>
<td>roadID</td>
<td>Road Id for localisation of event</td>
</tr>
<tr>
<td>subSeg</td>
<td># of subsegments (for fine localization)</td>
</tr>
<tr>
<td>coordgps</td>
<td>GPS Coordinate</td>
</tr>
<tr>
<td>vehID</td>
<td>Vehicle ID having detected the event</td>
</tr>
<tr>
<td>src</td>
<td>Binary value indicating if the message content is the result of detection or the result of processing disseminated data.</td>
</tr>
<tr>
<td>tAcquisition</td>
<td>Time of event perception</td>
</tr>
<tr>
<td>tLastUpdate</td>
<td>Last updating date</td>
</tr>
<tr>
<td>m</td>
<td>Mass function</td>
</tr>
</tbody>
</table>

Table 1. Message attributes

Some specific attributes (subSeg) have been added for future developments.

4 Message Combination

4.1 Temporal Persistence

The observed system is composed of events evolving in time and space. The delay between emission and reception can be small or large according to the routing and propagation algorithms. The confidence in a message also depends on of the age of the received event.

Belief masses should thus be modified according to the delay between the date of data processing and the date of data emission. The objective is to maintain data consistency with or without new messages. Indeed, in order to maintain a high level of confidence about an event, new messages confirming this event are needed. Without such confirmation, the confidence should decrease. We propose to define a discounting factor γ according to the time difference \( \Delta t = t_{\text{current}} - t_{\text{lastUpdate}} \) and a value \( \rho(ev) \) depending on the event ev:

\[
\gamma = \exp(-\Delta t / \rho(ev)).
\]

The value \( \rho(ev) \) characterizes the persistence of event ev. For example, \( \rho(LV) \) is high while \( \rho(AN) \) is shorter.

Temporal extension is performed by a discounting operation:

\[
m_{\text{current}}(A) = \gamma \cdot m_{\text{lastUpdate}}(A), \quad A \neq \Omega
\]

\[
m_{\text{current}}(\Omega) = 1 - \gamma + \gamma \cdot m_{\text{lastUpdate}}(\Omega)
\]

Notice that we have simplified the notation because the referential, event, road and source are constant in this part of the algorithm.

4.2 Spatial Propagation

Message combination in ad hoc networks should also take into account spatial properties of the observed events.

We can assume that some events observed on road segment \( S_i \) hold as well at positions close enough to \( S_i \). It is true, for example, for weather observations. The size of the neighborhood depends of the type of the diffuse event. A previous approach to plausible reasoning from spatial observations was proposed by Lang and Muller in [11]. These authors consider an observation point \( o \) (for example, the current position of the vehicle) and try to infer beliefs about what holds in \( o \) from the properties of the other road segments \( x_i \) \((i=1 \ldots n)\). Their model of spatial persistence is based on an extrapolation of observation \( m_{x_i \rightarrow o} \) calculated by discounting the mass function \( m_{x_i} \):

\[
m_{x_i \rightarrow o}(A) = \exp(-\frac{d(x_i,o)}{\lambda}) \cdot m_{x_i}(A), \quad A \neq \Omega
\]

\[
m_{x_i \rightarrow o}(\Omega) = (1 - \exp(-\frac{d(x_i,o)}{\lambda})) + \exp(-\frac{d(x_i,o)}{\lambda}) \cdot m_{x_i}(\Omega)
\]

where \( d(x_i,o) \) is the distance between the focus point \( o \) and the road segment \( x_i \) and \( \lambda \) represents the degree of persistence.

The belief at the focus point \( o \) is then the Dempster combination of the \( m_{x_i \rightarrow o} \):

\[
m_o = \bigoplus_{i=1}^{n} m_{x_i \rightarrow o}
\]

However, the problem of the dispersion in space of points \( x_i \) \((i=1,\ldots,n)\) has to be considered. Figure 2 shows two different situations. Since the points \( x_1 \) and \( x_2 \) are close in the second configurations, their influence on point \( o \) should be reduced.

![Figure 2: Dispersion configurations.](image)
To remedy this problem, Lang and Muller propose to introduce a discounting factor when combining mass assignments. The discount rate grows with the proximity and thus the dependency between the points where observations have been made. The discounting factors are calculated based on geometrical criteria.

We propose in this paper to make use of the numerical map to define the dependency between points $x_i$ (defined, for example, by GPS coordinates). As previously mentioned, an event is localized on the numerical map and a road-ID is associated to it. The roadID attribute can thus be used to indicate if events are located on the same portion of road. In this case, the masses $m_{x_i \rightarrow o}$ are considered as non independent and combined using the cautious rule. If the messages containing the masses come from roads with different roadIDs, they are combined with Dempster’s rule. The algorithm is then:

$$m_o = m_{x_i \rightarrow o}$$

$$\text{tabRoadID} \leftarrow \text{roadID}(x_1)$$

for $i = 2..n$

if $\text{roadID}(x_i) \in \text{tabRoadID}$

$$m_o = m_o \otimes m_{x_i \rightarrow o}$$

else

$$\text{tabRoadID} \leftarrow \text{roadID}(x_i)$$

$$m_o = m_o \otimes m_{x_i \rightarrow o}.$$ 

Here again the notation has been simplified because the referential, parameter, time and source are constant.

4.3 Prior Knowledge

Using numerical map to support the database makes it easy to attach prior information to road segments. For example, a congestion zone can be identified in some urban area, frequent fog condition can be observed near the wet zone, etc. This information can be fused with extrapolated and combined data before a decision is made.

4.4 Global Algorithm

As the vehicle is moving continuously, two approaches can be considered for processing a received message. The first one is a message-triggered approach in which each message is processed when it arrives at the node. Since the frequency and number of messages are unknown, it is difficult to guarantee that all messages will be processed. We prefer the second approach, referred to as the road segment triggered approach, in which each received message is kept in memory. When the node moves on a road segment $S_v$, all messages in a specified neighborhood are processed to compute the belief in the situation on $S_v$. The neighborhood can be defined according to an area around the current position or according to the itinerary of the vehicle.

In the context of VANET, it is unrealistic to assume that messages are independent. Information could be relayed and completed by the nodes in the network. As mentioned previously, some message attributes can determine if two messages are independent or not. We can consider that independence can be assumed in the following cases:

- Two messages sent by two different nodes with attribute src equal to 1. ($\text{src} = 1$ means that information is acquired by the node)
- Two messages sent by the same node at two different dates are regarded as independent if the node has made two distinct acquisitions: $\text{src}$ equal to 1 and $\text{t Acquisition}$ values are different.

In the other cases, the messages are processed as coming from dependent sources. The global algorithm is described in Figure 3. The grey box has not been implemented yet and is left for future work.

5 Preliminary Results

In order to test this approach, the above algorithm has been implemented in Matlab. The messages are simulated on the basis of a real numerical map (NavTeQ). A module extracts roads from the map in a specified area [10]. It uses the Benomad kit. Examples in this section are made with a map cache extracted from the GIS with a radius equal to 4 km.
5.1 Spatial Extension

To illustrate the spatial propagation we propose to simulate seven messages concerning the event “Low Visibility” (LV). Indeed, low visibility corresponds to weather conditions like fog or hard rain. These phenomena are spatially diffuse and can be spatially propagated. The first example shows the results of the combination of messages coming from seven distinct road segments situated in two distinct areas. The results reported in Figure 4 were calculated with normalized operators (mass on empty set was used to normalize). The attenuation factor computed from the distance between the current position of the vehicle and GPS data in messages discount efficiently the $m\{1\}$ values.

![Map for LV messages](image1)

a) Map for low visibility messages. The vehicle is on segment V. The fog area is localized in the grey area. *RoadId* of messages are shown.

![bba contained in each message](image2)

b) $bba$ contained in each message (left) and spatial discounting on segment V (right).

![Resulting belief function](image3)

c) Resulting belief function for LV event on segment V with normalized operators.

Figure 4. Example of spatial propagation for LV messages distributed on real map

The next scenario concerns the problem of spatial dependency. To illustrate it, the mixed cautious/conjunctive rule was compared with conjunctive only combination in the case where messages are localized on the same segments. The results are shown in Figure 5.

![Map messages](image4)

a) Map messages. The vehicle is on segment V. *RoadId* of messages are shown.

![spatial propagation for LV](image5)

b) $bba$ contained in each message (left) and spatial discounting on segment V according the *RoadID* localization (right).

![Belief function](image6)

c) Belief function on segment V. Comparison between two different strategies

Figure 5. Example of spatial propagation for seven messages localized on only three segments on real map.

The use of the cautious rule when messages are co-localized limits the reinforcement of confidence values as compared to well distributed sources. The well distributed
sources are represented by the belief function resulting from the conjunctive rule only. This approach has lower computational complexity than the spatial dispersion method proposed in [11].

The main difficulty for the implementation of this method is the definition of parameter $\lambda$. The spatial diffusion of real events is never constant and depends on a lot of context-dependent conditions.

5.2 Source Dependency

The next scenario illustrates the global algorithm (Figure 6). It simulates messages concerning an event LV coming from a group of vehicles exchanging data in an area. Messages are relayed by other vehicles (thanks to multi-hop protocol). A subset of messages (localized at 11, 120 and 30 RoadID) are labelled with $src = 0$ (not original perception of event).

a) Map messages. The vehicle is on segment V. RoadID of messages are shown. Messages localized on 11, 120 and 30 RoadID are labelled with $src = 0$ (not initial message).

b) bba's before and after spatial discounting on segment V according the RoadID localization and $m$.

c) Belief functions on segment V

Figure 6 shows the belief functions on segment V. The result of our approach is compared with the use of the conjunctive rule only without tacking into consideration the dependence between messages. Thanks to the cautious rule, all messages can be processed according to their distinctness. The behaviour of the cautious rule is highlighted in this example.

5.3 Temporal Discounting

Figure 7 shows an example of global combination with a message localized on segment V (current road segment). The message were dated with $t_{\text{Acquisition}} = t_{\text{Current}} - 3600s$. The RoadID is the current segment V and $src=1$. Temporal discounting on $m$ was performed on the bba of this message, before combining it with the result reported in figure 6c).

Figure 7 shows the belief functions on segment V. The implementation of this algorithm in real conditions can be envisaged provided GPS data (position and global clock) are available. The discounting factor based on the decay function can be roughly estimated for different kinds of dynamical events. However, like the spatial parameter $\lambda$, we can already assume that implementation in real conditions will require fine tuning of the $\gamma$
parameter, as the life duration of an event is context-dependent.

6 Conclusion

A distributed data fusion method for uncertain reasoning in ad hoc and dynamical networks has been presented. The method is based on belief functions and implements a strategy to combine confidence in messages. The preliminary results are promising and this approach is still in the process of being developed in order to validate the principle. A decision rule and a sending message strategy have to be defined before considering more complex simulations. Coupling this program with an ad hoc network simulator will be a real added value for validation.

The method described in this paper is able to process only one type of event. We will next consider the relation between events in order to take into consideration the link between some of them (for example, a congestion event is often the consequence of an accident event).

Finally, future work will focus on coupling information reliability with node reliability, i.e., confidence in the sending vehicle. Indeed, one of possible Sybil attacks consists in multiplying false nodes relaying many times the same information and creating false data sources. In the future, the confidence management system under development will have to be robust against Sybil node attacks.

Acknowledgments

This work is a part of the Prévisibilité project supported by the French Region Picardie. The authors would like to thank M. Shawky and B. Ducourthial who are the initiators of this research.

References