



Politecnico di Bari

Repository Istituzionale dei Prodotti della Ricerca del Politecnico di Bari

A Knowledge Fusion Approach for Context Awareness in Vehicular Networks

This is a post print of the following article

Original Citation:

A Knowledge Fusion Approach for Context Awareness in Vehicular Networks / Ruta, M., Scioscia, F., Gramegna, F., Ieva, S., Di Sciascio, E., Perez De Vera, R.. - In: IEEE INTERNET OF THINGS JOURNAL. - ISSN 2327-4662. - ELETTRONICO. - 5:4(2018), pp. 2407-2419. [10.1109/JIOT.2018.2815009]

Availability:

This version is available at <http://hdl.handle.net/11589/123277> since: 2022-06-08

Published version

DOI:10.1109/JIOT.2018.2815009

Publisher:

Terms of use:

(Article begins on next page)

A Knowledge Fusion Approach for Context Awareness in Vehicular Networks

Michele Ruta, *Member, IEEE*, Floriano Scioscia, Filippo Gramegna, Saverio Ieva, Eugenio Di Sciascio, Raffaello Perez De Vera



Abstract—Vehicular Ad-hoc Networks (VANETs) are a challenging IoT scenario. While research is proposing increasingly sophisticated hardware and software solutions for on-board context detection, probably high-level context information sharing has not been adequately addressed so far. This paper proposes a novel logic-based framework enabling a contextual data management and mining in VANETs. It grounds on a knowledge fusion algorithm based on non-standard, non-monotonic inference services in Description Logics, adopting standard Semantic Web languages. Ontology-referred context annotations produced by individual VANET nodes are merged with automatic reconciliation of inconsistencies. An efficient information dissemination protocol complements the proposal. The approach has been implemented in a vehicular network simulator and early experimental results proved its effectiveness and feasibility.

Index Terms—Vehicular Networks; Knowledge Management; Information Fusion; Semantic Web of Things; Multi-Agent Systems

1 INTRODUCTION

In the Internet of Things (IoT), processing and communications capabilities, often wireless and ad-hoc, borrowed from heterogeneous pervasive micro-devices attached to objects augment their basic properties and allow information exchange. Embedding sensing and computing hardware in everyday things adds an immaterial substratum to them materializing data to which applications should give sense. Hence, cognitive IoT enables meaningful multi-object coordination, cooperation and control. Knowledge Representation and Reasoning (KRR) techniques and technologies allow information modeling based on formal and rigorous interpretation of its meaning (semantics). This enables not only greater interoperability across different HW/SW platforms, but also reasoning tasks (*inferences*) to derive new implicit insight from information explicitly asserted in a Knowledge Base (KB). Among the many available KRR languages and tools, those born from the Semantic Web

initiative enjoy global adoption and high optimization of algorithms and implementations. The convergence of the Semantic Web and the IoT has led to the so-called *Semantic Web of Things* (SWoT), aiming at intelligent information processing and exchange in mobile and pervasive environments.

Vehicular Ad-hoc Networks (VANETs) are a fascinating and challenging IoT context. Cars and vans are increasingly equipped with various sensor types to monitor in real time both the state of internal components/subsystems and external stimuli coming from other vehicles, pedestrians, surrounding objects and environmental conditions. Advanced Driving Assistance Systems (ADAS) are already deployed in mass-market car models, exploiting a Controller Area Network (CAN) bus infrastructure distributing sensor data to multiple processing cores for mining, in order to support a range of driving tasks (*e.g.*, lane keeping, distance keeping, parking) in an automated way. This is Level 2 in the 0-5 driving automation scale from the Society of Automotive Engineers (SAE) [1]. Aggressive industrial research programs aim to bring Level 4 to the market within the next decade, *i.e.*, autonomous driving in the most severe conditions. In order to reach this goal, progress is required not only on cognitive computing architectures performing on-board sensing and intelligent information analysis, but also on vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) information exchange. Just like human drivers use turn signals and horns to communicate and receive indications from traffic lights, VANETs have been conceived and Dedicated Short Range Communications (DSRC) technologies have been designed to allow data exchange among vehicles (*a.k.a.* On-Board Units, OBUs) as well as with fixed infrastructure elements (Road-Side Units, RSUs). Nevertheless, VANET information exchange is still quite simplistic. Current DSRC standards –coming to the market in the next few years– are limited to either basic emergency signals –with no further intelligent interaction– for urgent safety alerts, or multimedia streams for comfort services. Contextual data gathering and mining approaches in literature have been based on custom models and formalisms, which all vehicles must agree upon. Instead, standard high-level formal languages provide abstraction and interoperability, by adopting non-ambiguous reusable vocabularies (ontologies) to model contextual knowledge and support automated inference services to implement high-level cognitive processes devoted

Manuscript received March 31, 2017; revised November 8, 2017; accepted February 26, 2018 (Corresponding author: Floriano Scioscia).

Authors were with the Department of Electrical and Information Engineering, Polytechnic University of Bari, Bari, I-70125, Italy (e-mail: michele.ruta@poliba.it, floriano.scioscia@poliba.it, filippo.gramegna@poliba.it, saverio.leva@poliba.it, eugenio.disciascio@poliba.it, raffaello.perez@hotmail.it).

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Digital Object Identifier XXXXXX

to ambient consciousness.

This paper introduces a distributed collaborative framework for context awareness in vehicular networks. It is based on two main contributions: (i) a non-monotonic inference service for knowledge fusion, with resolution of inconsistencies and resilience against spurious information; (ii) a dissemination protocol for efficient bidirectional knowledge propagation in VANETs. SWoT research efforts [2] showed how a smartphone on a vehicle can exploit the On-Board Diagnostics (OBD-II) port to extract real-time data, perform efficient mining and generate a semantically rich description of the context it is dipped in. A VANET composed by such kind of vehicles, could (i) quickly adapt to fast context condition evolution and (ii) recover in very few steps from detection mistakes of individual agents. Furthermore, the approach proposed here is grounded on the Open World Assumption (OWA), which makes it well suited to processing even incomplete information (due to *e.g.*, missing sensor device types). While the dissemination protocol was designed specifically for VANETs, results concerning knowledge fusion have general applicability to MAS in a wide range of IoT contexts. To the best of our knowledge, no other study is similar enough to our proposal [3] to allow a direct comparison. Hence, the approach was fully implemented in the NCTUns network simulator [4] and experimental analysis was conducted to early assess its feasibility and effectiveness.

The remainder of the paper is as in what follows. Section 2 provides a general background on VANETs and KRR, while Section 3 discusses on most relevant related work. The proposed approach is described in Section 4, then Section 5 includes an illustrative case study to clarify the proposal and introduce the first experiments. Finally, concluding remarks closes the paper.

2 BACKGROUND

In the following subsections a general background on vehicular networks and knowledge management tools and theory is provided to make the paper self-contained.

2.1 Basics of VANET and V2X Communications

A VANET is a network consisting of moving vehicles (OBUs) and fixed stations (RSUs), both equipped with wireless interfaces, sharing information about the context where they are plunged, as depicted in Figure 1.

The U.S. FCC (Federal Communication Commission) reserved 5.850-5.925 GHz frequency spectrum exclusively for DSRC V2V and V2I applications adopting IEEE 802.11p [5], [6], [7], the most widely recognized standard for V2X (vehicle-to-anything, *i.e.*, V2V or V2I) communication. Recently, fifth generation (5G) networks have been proposed as a cloud-friendly, high-performance alternative to DSRC [8]. In either case, VANET nodes (*actors* or *agents*) exchange information about the environment where they operate, gathered through their on-board devices, with the aim to detect in advance dangerous situations and assist the driver.

In VANETs, a *push* (broadcast) communication model is preferred w.r.t. a *pull* (on-demand) one, since it allows to avoid routing algorithms at the network layer, which would

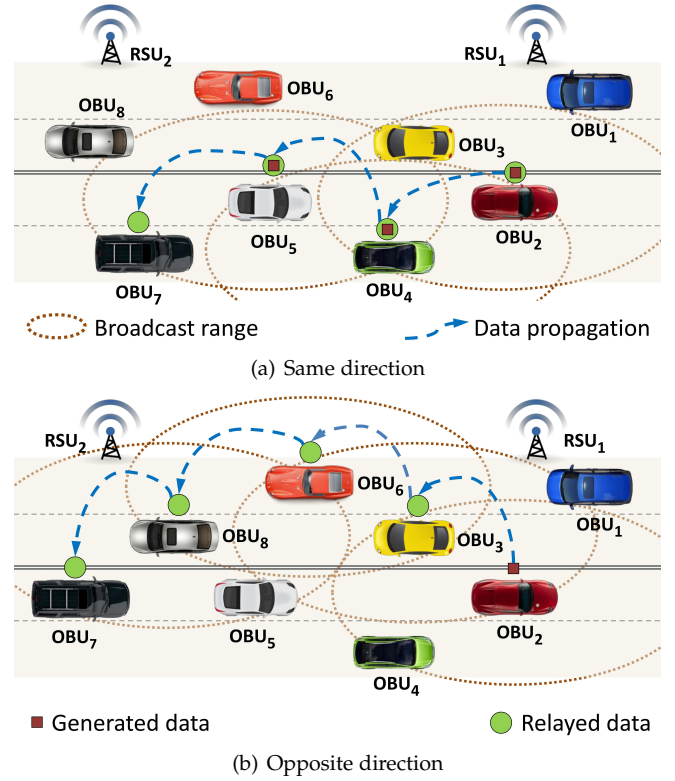


Fig. 1. VANET bidirectional data dissemination strategies

introduce unacceptable latencies in high-speed or high-traffic scenarios. In the push communication model, data is sent to all agents within the radio range of the sender, in order to have a fast and wide data propagation. To achieve this goal, it is important to determine when to send the collected information. Two basic strategies exist: *flooding* and *dissemination*. In flooding, each node periodically broadcasts its data; recipients store this information and immediately forward it. It is fairly intuitive that such a mechanism may incur in network load problems, particularly in scenarios with high vehicle density. A more scalable data exchange mechanism is *dissemination*, where data produced (*generated data*) and received (*relayed data*) by a node are periodically sent together after a so-called Broadcast-Period (BP). This approach is more efficient in terms of number of packets sent over the network and reduces the general collisions. In literature different dissemination models exist [9]. The framework proposed here adopts a V2V data push communication model with a *bidirectional* dissemination scheme [10]. This choice allows exploiting the relative motion between vehicles travelling in different directions (Figure 1), which increases the distance covered by a packet in one hop w.r.t. other models.

2.2 Knowledge Representation and Reasoning Basics

The terminological substratum in the proposed knowledge fusion approach is given by the Description Logics (DLs) family of logic languages for KRR [11]. Basic DL elements are: *concepts* (*a.k.a.* *classes*) representing sets of objects; *roles* (*a.k.a.* *properties*), linking pairs of objects in different concepts; *individuals*, named instances belonging to concepts.

They can be combined using logical *constructors* to build concept *expressions*. Each DL has a different set of constructors; including more constructors enhances expressiveness of a DL, but it also increases complexity of *inference* tasks, possibly making them not computationally tractable or even undecidable, so a careful trade-off is required. Every DL uses the *conjunction* constructor, usually denoted as $\sqcap$; some DLs include also disjunction $\sqcup$ and complement $\neg$. Roles can be combined with concepts using *existential role quantification* and *universal role quantification*. Other constructs may involve counting, as *number restrictions*. Concept expressions can be used in *inclusion* and *definition* axioms, which impose restrictions on possible models of the domain according to elicited knowledge. Sets of such axioms are called *TBox* (Terminological Box, *a.k.a.* *ontology*). A TBox without cyclic (groups of) axioms is called *acyclic*. DL semantics is grounded on set-based *interpretations* which tie expressions to sets of objects in a domain. A *model* of a TBox $\mathcal{T}$ is an interpretation satisfying all inclusions and definitions in $\mathcal{T}$. In this paper, both framework and examples refer to the *Attributive Language with unqualified Number restrictions* ($\mathcal{ALN}$) DL, a subset of OWL 2 (Web Ontology Language) [12]).

Cognitive IoT contexts and scenarios envisage entities with a distributed intelligence to gather knowledge, act upon it and exchange with other entities. It is needed, then, the capability to fuse (partially) redundant information and/or retract consequences when new conflicting knowledge becomes available. Standard inference services like *Subsumption* and *Satisfiability* [11] are not enough for that, as they provide just a Boolean answer. Therefore the proposed approach exploits non-monotonic inference tasks originally defined for belief revision. In particular, *Concept Abduction* (CA) [13] provides an explanation when Subsumption does not hold: given a DL language $\mathcal{L}$, an ontology $\mathcal{T}$ in $\mathcal{L}$ and two concepts C and D satisfiable in $\mathcal{T}$, if $\mathcal{T} \not\models C \sqsubseteq D$, then CA finds a concept H (for *Hypothesis*) such that $\mathcal{T} \models C \sqcap H \sqsubseteq D$. Basically, H represents what is in D but is missing from C , *e.g.*, when two entities (vehicles in case of VANETS) V_C and V_D are in the same area but only V_D detects wind speed by having an anemometer. Notice that in general C can have information missing from D as well, otherwise the subsumption relation $\mathcal{T} \models D \sqsubseteq C$ would be true. For this purpose, the *Bonus* inference service [14] extracts a concept B from C , denoting what it presents beyond D . On the other hand, if the conjunction $C \sqcap D$ is unsatisfiable w.r.t. the ontology, one can retract some requirements G (for *Give up*) from D to obtain a concept K (for *Keep*) such that $K \sqcap C$ is satisfiable in $\mathcal{T}$. This is called *Concept Contraction* (CC) [13] and occurs when two entities produce partially conflicting descriptions of a given area. For both CA and CC, minimality criteria are defined, since one usually wants to hypothesize or give up as little as possible. The proposed framework requires also a way to subtract information in a description from another one. This is accomplished by means of the *Concept Difference* reasoning service [15]. Also for Difference many valid solutions exist, but a maximality criterion should be adopted in this case (*i.e.*, subtract as much as possible).

3 RELATED WORK

Cognitive computing starts from human cognition models to define artificial intelligence systems capable of complex tasks on large volumes of heterogeneous data [16]. Prevalent approaches include knowledge-based ones –like KRR– and machine learning ones (such as neural networks) [17]. In addition to high accuracy levels, predictably fast computation and support for large data flows are often required, since many applications work in real time on Big Data streams [18]. Information fusion techniques are adopted for that. A basic distinction in literature exists between High-Level Information Fusion (HLIF) and Low-Level Information Fusion (LLIF) [19]. LLIF is typical of multi-sensor fusion, dealing with numerical data and aiming at classification, identification and tracking. HLIF concerns abstract symbolic information and aims at situation awareness and information refinement. The framework proposed here refers to HLIF, therefore sensor data fusion is beyond the scope of the paper. We suppose vehicles are equipped with on-board sensing and data mining capabilities for LLIF to produce logic-based annotated descriptions of the environment and the vehicle itself, such as in [2]. A similar approach to vehicular context detection is in [20], adopting Complex Event Processing (CEP) to analyze patterns of elementary events; nevertheless, the output is not formal and symbolic, hence it cannot be easily exploited for HLIF.

Cevolani [21] discussed a framework for belief merging, which has theoretical similarities with the approach of the present paper. In particular, that work emphasized the importance of *weak disagreement* between peers in order to improve truth approximation. The author, however, focused on majority agreement among several peers, which requires *one-shot* merging; for their nature, cognitive frameworks in VANETs and other Multi-Agent Systems (MAS) for Internet of Things contexts appear more amenable to decentralized *iterative* approaches like the one presented here. This marks a difference also with logical *arbitration*, where all arguments are of equal importance [22], situating the work in the area of *belief revision*, where latest arguments are considered more important and their truth must be preserved by using non-monotonic inference services. In particular, the knowledge fusion approach extends and generalizes the simpler case introduced in [23].

Ilarri *et al.* [3] surveyed extensively the data management aspects in VANETs. In wireless sensor network and VANET literature, *semantic data aggregation* often refers to lossy compression or approximate techniques to summarize data from multiple sources; in this paper actual information aggregation at semantic-based representation level occurs instead. Furthermore, assessing relevance of incoming data was marked as an important issue: the approach adopted here falls under the *space-time relevance function* category, albeit as a simple binary function to decide packet drop or storage. Finally, the authors pointed out that complexity of VANET scenarios defy the development of generic solutions and that ontology-based approaches can help by increasing interoperability between different vehicular data management systems and allowing high-level situation description, assessment and inference, but their full application possibilities in VANETs are yet to be explored; in this regard, the

proposed work aims to provide an early contribution.

Intelligent HLIF is clearly a key problem in VANETs, as information is often vital for safety. The network must be robust against imprecise or incorrect data injected by vehicles due to faults or misbehaviors [24]. The survey [25] provides an overview of issues and techniques on information fusion for context awareness in VANETS, including ontology-based context models. Nevertheless, in reported works inference was used only for service discovery, not for HLIF (confirming the findings of [3]). Examples include the work in [26] and S-Aframe [27], a multi-agent framework for vehicular social networks which includes a semantic-based context model for vehicles, exploiting a rule-based engine to support service discovery. More recently, methods for coping with imprecision and uncertainty were proposed for VANETs in decentralized collaborative HLIF implementations. Bayesian approaches are quite popular, such as [28], which enables a vehicle to improve its GPS location estimate by incorporating GPS data from other vehicles and inter-vehicle distance measurements. Another popular class of methods is based on the Dempster-Shafer theory (*a.k.a.* belief function theory). Variants have been applied to VANETS *e.g.*, in [29], [30], assessing results in selected case studies. Unfortunately, the above approaches cannot be easily extended to work with semantic-based formalisms for context characterization. A recent approach is in [31] for collision threat assessment in VANETs, adopting Multi-Entity Bayesian Networks, which extend entities in a Bayesian network with semantic relationships, in order to improve probabilistic event evaluation.

4 MULTI-AGENT KNOWLEDGE MANAGEMENT IN COMPLEX CONTEXTS

This section introduces the proposed cognitive framework in the context of vehicular complex networks. Both knowledge fusion strategy and data dissemination approach will be presented in the following subsections.

4.1 Knowledge Fusion via Deductive Intelligence

In what follows, an operational –rather than axiomatic– definition of the information fusion framework is provided for easier understanding. A swarm of independent, mobile entities (agents) exchanging semantically annotated *packets* is envisioned. Each packet contains the sender’s current knowledge of the environment and context, expressed in a DL language w.r.t. a shared reference ontology. The problem of preliminary ontology agreement among agents is beyond the scope of the paper. Packets contain a 4-tuple of annotations, each providing a particular point of view on the context, particularly:

- *C (Confirmed)*: elements observed by both the sender and other agents;
- *X (Clash)*: observations of the sender, not consistent with observations of others;
- *M (My)*: elements observed by the sender, but not by other agents;
- *E (External)*: elements observed by other agents, but not by the sender.

The approach is grounded on the *Open World Assumption* (OWA), stating the absence of information is not a constraint of negation, but simply underspecified or unknown. This implies *X* contains only explicitly conflicting knowledge fragments, whereas an agent considers *E* as possibly valid information, even though it cannot confirm it directly. By contrast, most database management systems adopt the *Closed World Assumption*, where information not explicitly stored is considered to be false.

When an agent completes every data gathering and annotation round, it generates a semantic annotation *N* representing fresh detected information. A collaborative protocol for periodic information dissemination, detailed in Section 4.2, exploits information fusion with reconciliation of inconsistencies. Meanwhile, incoming packets are stored in a cache. Three basic cases can occur:

A. Generation: it occurs if an agent generates an annotation and it has no valid packets in its cache.

B. Relay: it occurs if an agent has one packet in its cache and it is not able to produce its own semantic annotation.

C. Integration and relay: it occurs when an agent has received at least one valid packet, which must be fused with locally produced information before propagating the updated context snapshot.

In any case, the result is a new packet *P'*, to be sent in broadcast.

In the above case A, the resulting *P'* is (OWL Manchester syntax [32] is adopted for easier readability): $P'_A = \langle C \equiv \top; X \equiv \top; M \equiv N; E \equiv \top \rangle$. The annotation *N* will be placed in the packet field *M* (observation of the sending entity), while other fields will be set to $\top$ (*Top* or *Thing*), *i.e.*, the most generic possible observation¹. Conversely, in case B the agent relays a received packet *P*, hence the output *P'* will contain only observations made by other agents: $P'_B = \langle C \equiv \top; X \equiv \top; M \equiv \top; E \equiv C_P \text{ and } X_P \text{ and } M_P \text{ and } E_P \rangle$. The *E* field is populated with the conjunction of the four annotations in *P*; the other ones will be set to $\top$. For any exchanged envelope *P*, its four annotations will benefit from mutual semantic compatibility, so that their conjunction will always be satisfiable, as long as the annotations generated by new observations are satisfiable. This property is guaranteed by the algorithm for case C whose details are reported hereafter. The last case (C) requires the fusion of generated and received information. For this purpose, a complex inference service was devised, so as to preserve the semantics of the fields of both input and output packets. The problem is named *Concept Integration* and is formalized as follows:

Definition 1. Let $\mathcal{L}$ be a DL language, $\mathcal{T}$ a set of axioms in $\mathcal{L}$, *N* a concept expression in $\mathcal{L}$ and $\mathcal{P} = (P_1, P_2, \dots, P_n)$ a chronologically sorted set (from the newest to the oldest) of 4-tuples of concept expressions $P_i = \langle C_i, X_i, M_i, E_i \rangle$ in $\mathcal{L}$, where *N* is the current knowledge of agent *A*, *P_i* is the knowledge accumulated by the agent *A_i*; *N*, *C_i*, *X_i*, *M_i*, *E_i* $\forall i$ are satisfiable in $\mathcal{T}$. A *Concept Integration Problem* (CIP), identified by $\langle \mathcal{L}, N, \mathcal{P}, \mathcal{T} \rangle$, is finding a 4-tuple of concept expressions

1. Due to the OWA, unspecified observations are denoted as “anything” instead of “nothing”, as one might think.

$P' = \langle C', X', M', E' \rangle$, which is the integration of N and $\mathcal{P}$ and represents knowledge accumulated by agent A' .

The simplest CIP case, for $|\mathcal{P}| = 1$, was solved in [23] exploiting the Concept Contraction, Abduction, Difference and Bonus non-standard and non-monotonic inference services. Algorithm 1 solves the CIP for the general case $|\mathcal{P}| \geq 1$. Figure 2 depicts the corresponding flowchart. The *Contract*, *Abduce*, *Bonus* and *Difference* (or $-$) operators refer to Concept Contraction, Abduction, Bonus and Difference, respectively. Lines 1-6 of the algorithm compute auxiliary variables. The new M' field is computed in lines 8-11: it is computed by eliminating iteratively older observations from new ones. Lines 12-16 compute the new C' field, by removing from the new observation N all the fragments which are either unconfirmed (i.e., M') or semantically inconsistent with older observations. Lines 17-21 compute the new X' field, which is simply the conjunction of all incompatible fragments with the new observation, as computed by CC in line 4. Finally, the E' field is computed. Pairs of received observations are processed iteratively, from the oldest to the newest: for each pair, mutually inconsistent fragments are removed via Contraction and Difference. The remaining consistent part of the newest received observation is used to compute the Bonus for the newly generated observation N , obtaining the knowledge fragment which are neither implied by N nor clashing with it. In this way, CIP allows to integrate an “older” snapshot of the context with “new” information to produce an “updated” snapshot, which reflects the viewpoint of the agent that has performed the integration.

The proposed knowledge fusion approach complies the following postulates by Dubois *et al.* [33]: unanimity, information monotonicity, (para)consistency enforcement, optimism, minimal commitment. Conversely, by design it does not have fairness (as it privileges more recent inputs), insensitivity to vacuous information (as it deals with “testimonies” provided by independent agents, so that “empty input” –i.e., $\top$ – is treated differently from “no input”) and commutativity (as it always integrates information from the oldest to the newest) properties. As illustrated by example in Section 5.1, features of the CIP reasoning task include management of incomplete information, reconciliation of inconsistencies in input data, quick adaptation to changes and robustness against spurious or inaccurate information. Actually, it reaches a steady state in just two rounds of observation and fusion after every variation in information [23]. Finally, Algorithm 1 shows that the computational complexity of Concept Integration depends directly on the one of Abduction, Contraction, Bonus and Difference inference tasks. In particular, the experimental evaluations with the *Mini-ME* mobile reasoner [34] worked on the $\mathcal{ALN}$ DL with acyclic TBoxes, for which structural algorithms [11] exist with PTIME complexity in all these inferences: in that case it follows also CIP has PTIME complexity.

4.2 Dissemination Protocol

The above knowledge fusion approach fully complies with various cognitive Internet of Things scenarios. Vehicular networks recall IoT general elements and principles with a severe complexity, hence they have been chosen as reference

ALGORITHM 1: Concept Integration Problem solution

Algorithm: *Integrate*($\langle \mathcal{L}, N, \mathcal{P} = (P_1, \dots, P_n), \mathcal{T} \rangle$)

Require:

$\langle \mathcal{L}, N, P_1 = \langle C_1, X_1, M_1, E_1 \rangle, \dots, P_n = \langle C_n, X_n, M_n, E_n \rangle, \mathcal{T} \rangle$
with $N, C_i, X_i, M_i, E_i \forall i \in \mathcal{L}$ satisfiable in $\mathcal{T}$

Ensure: $P' = \langle C', X', M', E' \rangle$ with $C', X', M', E' \in \mathcal{L}$ satisfiable in $\mathcal{T}$

```

// Preliminary computations
1: for  $i := 1$  to  $n$  do
2:    $Q_i := C_i$  and  $X_i$  and  $M_i$  // All fragments but  $E_i$ , to stop
   propagation of external (bonus) information
3:    $S_i := Q_i$  and  $E_i$ 
4:    $(K_i, G_i) := \text{Contract}(N, S_i)$ 
5:    $H_i := \text{Abduce}(K_i, S_i)$ 
6: end for
// Compute  $M'$ 
7:  $T_0 := H_1$ 
8: for  $i := 1$  to  $n - 1$  do
9:    $T_i := T_{i-1} - (H_i - H_{i+1})$ 
10: end for
11:  $M' := T_{n-1}$ 
// Now compute  $C'$ 
12:  $V := M'$  // temporary variable
13: for  $i := 1$  to  $n$  do
14:    $V := V$  and  $G_i$ 
15: end for
16:  $C' := N - V$ 
// Now compute  $X'$ 
17:  $V := G_1$ 
18: for  $i := 2$  to  $n$  do
19:    $V := V$  and  $G_i$ 
20: end for
21:  $X' := V$ 
// Finally, compute  $E'$ 
22: for  $i := n - 1$  to  $1$  do
23:    $(K_i, G_i) := \text{Contract}(Q_i, Q_{i+1})$ 
24:    $(\text{Inv}K_i, \text{Inv}G_i) := \text{Contract}(Q_{i+1}, Q_i)$ 
25:    $Q_i := K_i - (K_i - \text{Inv}K_i)$  // temporary variables
26: end for
27:  $E' := \text{Bonus}(N, Q_1)$ 

```

testbed. The protocol to disseminate data among moving entities focuses on an efficient dissemination of annotations, preserving compliance with VANET standards like 802.11p. The bidirectional strategy described in Section 2.1 is adopted as baseline. A **Semantic Packet (SP)** encapsulates the four semantic annotations managed by a vehicle V according to the fusion algorithm in Section 4.1. Figure 3 shows the SP structure. The header is as follows.

- *Message type*: its value can be (i) *Generated*, including new information observed by V in the current Broadcast Period (BP), and (ii) *Relayed*, when V merges and propagates knowledge received from other vehicles. As per the bidirectional model, nodes in the same direction broadcast both generated and relayed data, whereas nodes in the opposite direction propagate relayed data only.
- *Message ID*: unique message identification.
- *Sender ID*: unique ID of the sending node.
- *Timestamp*: time of SP generation.
- *Time To Live (TTL)*: time before SP expiration.
- *Vehicle data*: include latitude/longitude coordinates, acceleration (in m/s^2), speed (in m/s) and direction (in degrees).

SP body consists of the 4-tuple expressions described in Section 4.1. Proper compression techniques are adopted to

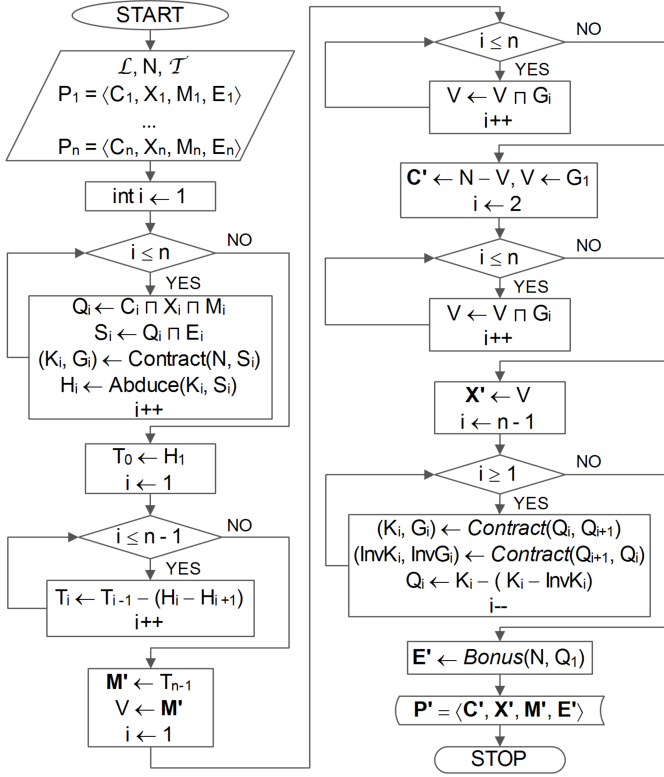


Fig. 2. Concept Integration Problem flowchart

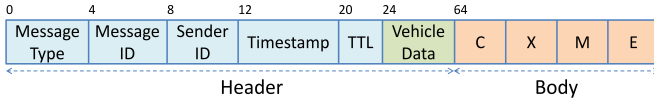


Fig. 3. Semantic Packet structure

cope with the verbosity of OWL annotations.

Figure 4 shows flowcharts for the dissemination protocol executed by every vehicle in each broadcast period. A detailed explanation is in what follows.

1. Listening. When vehicle V receives a SP, it checks the following conditions are matched: (i) Message Type is valid; (ii) the SP has not expired; (iii) the packet complies with the dissemination model, according to the motion direction of V (basically, the SP must come either from a node ahead in any direction or from a node behind, if in the same direction). If at least one of these conditions is not fulfilled, the message is discarded. Otherwise, the received SP is cached. In particular, a $Store\$$ cache holds the Generated SPs (GSPs) referring to road sections in the vehicle's path and the Relayed SPs (RSPs) in the opposite direction of V . A distinct $Relay\$$ cache holds the GSPs received from vehicles travelling in the opposite direction of V and the RSPs relayed by vehicles behind V in the same direction. The rationale is that $Store\$$ will contain knowledge of interest for nodes behind V and $Relay\$$ for nodes ahead of V in either direction.

2. Processing. During every broadcast period, V performs on-board data gathering, mining and annotation, producing a new semantic description N of its context. At the end of the BP, V generates a new GSP exploiting the fusion

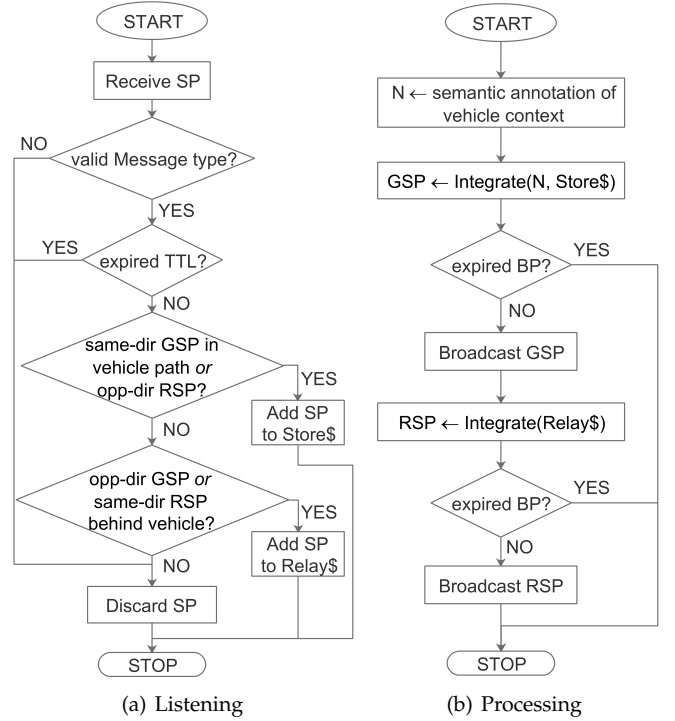


Fig. 4. Dissemination protocol flowcharts

algorithm in Section 4.1 to merge N with annotations in packets selected from the $Store\$$, if the location contained in the header is within a radius R w.r.t. the current position of V . On the other hand, an RSP is generated by running the fusion algorithm on the annotations stored in the $Relay\$$. In this case, the role of N is played by the conjunction of all four semantic annotations in the latest packet in $Relay\$$ (i.e., as if the generator of the latest packet performed the Concept Integration).

The proposed protocol requires vehicles to send the GSP before the next BP expires, whereas they will calculate and send the RSP only if there is enough available processing time, as the knowledge fusion algorithm is the most computationally expensive procedure in the framework.

5 SIMULATIONS

A straightforward quantitative comparison of the proposed cognitive approach for information fusion and dissemination with the state of the art is not completely possible, due to the not negligible differences in adopted techniques, reference scenarios and experimental setups. Anyway, Table 1 provides a feature comparison with the closest and most relevant works from Section 3. The present work is not the only one based on DL ontologies, but other similar approaches did not exploit non-monotonic reasoning and thus do not allow a "controlled" knowledge fusion and resiliency toward incomplete or inconsistent information. In general, only inconsistencies management is accomplished. Further frameworks include machine learning (neural networks, particularly), Bayesian networks (probabilistic) and belief function theory. The latter two support HLIF and are also robust toward missing or erroneous data, whereas neural networks were applied only to low-level features. Most

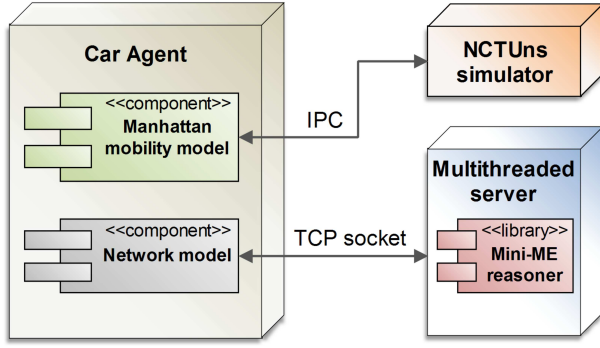


Fig. 5. Component diagram of the simulation architecture

works are oriented toward multi-agent systems, but few of them take into account specific (and strict) requirements of the IoT.

Implementation and performance evaluation of the proposed framework were carried out exploiting the NCTUns network simulator [4]. Each modeled scenario represents a VANET for cooperative monitoring of driving risk factors, where every agent runs on a smartphone within a car and gathers data from its own sensors as well as from the OBD-II (On-Board Diagnostics) port integrated in the vehicle. Raw data are analyzed and annotated with concept expressions referred to an ontology which models road features, traffic, weather and driving style [2] (not reported here due to lack of space; besides, the approach is independent from ontology details). In particular, each simulated vehicle has been configured to run a C++ process named *Car Agent*. As shown in Figure 5, the Car Agent manages both the Manhattan mobility model [35] and a network model complying the dissemination schema described above in Section 4.2. The Car Agent interacts with both the NCTUns simulation engine and the multithreaded reasoner server, whose elementary inference services are invoked to implement the semantic knowledge fusion algorithm (Section 4.1) whose non-standard inferences are provided by the *Mini-ME* mobile reasoner [34]. Agent-server communication occurs through a TCP socket connection; Google Protocol Buffer² is used to serialize the high-level structured data to a platform-independent binary format. In a real implementation, the reasoner would be integrated directly within the agent software. The following subsections present an illustrative case study to clarify the framework and introduce an early experimentation on it, respectively.

5.1 Case Study

The toy example sketched here defines a cooperative risk monitoring application even if it is not a complete solution (which should require further modules as data gathering from embedded vehicle devices, mining to generate annotations, feedback controls to provide advanced driving assistance, suitable user interfaces). Anyway, it can better highlight most relevant features of the proposed approach.

Let us consider a road consisting of two lanes where vehicles V_1 and V_2 travel from west to east, with V_2 in front

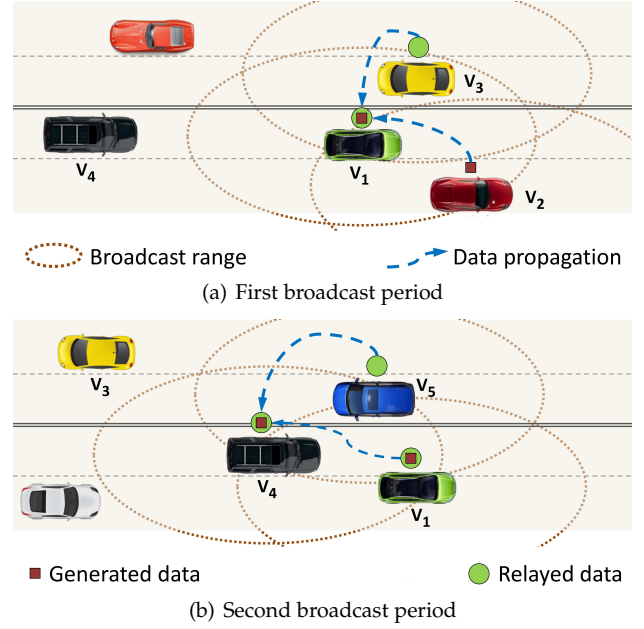


Fig. 6. Case study scenario

of V_1 , while vehicle V_3 is in the opposite direction (Figure 6). All vehicles compose and exchange semantic information over the network, according to the dissemination protocol exposed before. At the beginning of its BP (Figure 6(a)), vehicle V_1 detects strong wind, an uphill road, low traffic and rain, and generates the SP $P1$ accordingly. Let us suppose that during its last BP V_1 also received a GSP $P2$ from V_2 and a RSP $P3$ from V_3 , whose content is:

$P1 = \langle C \equiv T; X \equiv T; M \equiv N \equiv (\text{hasSlope } \textcolor{red}{\text{only}} \text{ Ascent_Slope}) \text{ and } (\text{hasWeather } \textcolor{red}{\text{only}} \text{ Rain}) \text{ and } (\text{hasTraffic } \textcolor{red}{\text{only}} \text{ Low_Traffic}) \text{ and } (\text{hasWind } \textcolor{red}{\text{only}} \text{ Strong_Wind}); E \equiv T \rangle$

$P2 = \langle C \equiv T; X \equiv T; M \equiv (\text{hasSlope } \textcolor{red}{\text{only}} \text{ Ascent_Slope}) \text{ and } (\text{hasWeather } \textcolor{red}{\text{only}} \text{ Rain}) \text{ and } (\text{hasTraffic } \textcolor{red}{\text{only}} \text{ High_Traffic}); E \equiv T \rangle$

$P3 = \langle C \equiv T; X \equiv T; M \equiv T; E \equiv (\text{hasSlope } \textcolor{red}{\text{only}} \text{ Ascent_Slope}) \text{ and } (\text{hasWeather } \textcolor{red}{\text{only}} \text{ Rain}) \text{ and } (\text{hasTraffic } \textcolor{red}{\text{only}} \text{ Low_Traffic}) \text{ and } (\text{hasRoad_Type } \textcolor{red}{\text{only}} \text{ Secondary_Road}) \rangle$

After the expected BP, V_1 executes the knowledge fusion algorithm as in what follows:

$K_1 \equiv (\text{hasSlope } \textcolor{red}{\text{only}} \text{ Ascent_Slope}) \text{ and } (\text{hasWeather } \textcolor{red}{\text{only}} \text{ Rain}) \text{ and } (\text{hasWind } \textcolor{red}{\text{only}} \text{ Strong_Wind})$

$G_1 \equiv (\text{hasTraffic } \textcolor{red}{\text{only}} \text{ Low_Traffic})$

$K_2 \equiv (\text{hasSlope } \textcolor{red}{\text{only}} \text{ Ascent_Slope}) \text{ and } (\text{hasWeather } \textcolor{red}{\text{only}} \text{ Rain}) \text{ and } (\text{hasWind } \textcolor{red}{\text{only}} \text{ Strong_Wind}) \text{ and } (\text{hasTraffic } \textcolor{red}{\text{only}} \text{ Low_Traffic})$

$G_2 \equiv T$

$H_2 \equiv (\text{hasWind } \textcolor{red}{\text{only}} \text{ Strong_Wind})$

$H_1 \equiv T_1 \equiv M \equiv (\text{hasWind } \textcolor{red}{\text{only}} \text{ Strong_Wind})$

$V \equiv (\text{hasWind } \textcolor{red}{\text{only}} \text{ Strong_Wind}) \text{ and } (\text{hasTraffic } \textcolor{red}{\text{only}} \text{ Low_Traffic})$

$C \equiv (\text{hasSlope } \textcolor{red}{\text{only}} \text{ Ascent_Slope}) \text{ and } (\text{hasWeather } \textcolor{red}{\text{only}} \text{ Rain})$

$X \equiv (\text{hasTraffic } \textcolor{red}{\text{only}} \text{ Low_Traffic})$

$K_1 \equiv (\text{hasSlope } \textcolor{red}{\text{only}} \text{ Ascent_Slope}) \text{ and } (\text{hasWeather } \textcolor{red}{\text{only}} \text{ Rain}) \text{ and } (\text{hasTraffic } \textcolor{red}{\text{only}} \text{ High_Traffic})$

$G_1 \equiv T$

$InvK_1 \equiv T$

$InvG_1 \equiv T$

$E \equiv T$

At the end of the procedure, a new GSP $P1'$ is produced and disseminated over the network by vehicle V_1 .

2. <https://github.com/google/protobuf>

TABLE 1
Comparison of the proposed cognitive approach with the state of the art (works sorted by publication year)

Source	Approach	Context Representation	Application Domain	Information Fusion	Incomplete/Inconsistent Information Resiliency	MAS-oriented	IoT-oriented
Santa <i>et al.</i> [26]	KRR	Description Logics	VANETs	No (points of interest discovery only)	No (rule-based reasoning)	Yes (overlay net)	Yes (cloud-assisted)
Terroso-Sáenz <i>et al.</i> [20]	Complex Event Processing	Probabilistic graph	Vehicles	LLIF (fuzzy rule-based)	Yes	No	No
Hu <i>et al.</i> [27]	KRR	Description Logics	VANETs	No (service discovery only)	No (rule-based reasoning)	Yes	Yes (mobile devices)
Rohani <i>et al.</i> [28]	Bayesian	Position estimation	VANETs	LLIF (probabilistic estimator)	Yes	Yes	No
Radak <i>et al.</i> [29]	Belief function theory	Belief mass distribution	General (probabilistic)	HLIF (distributed confidence calculation)	Yes (<i>discounting</i> function)	Yes	No
Farah <i>et al.</i> [30]	Belief function theory	Belief mass function	VANETs (possibly general)	HLIF (mass function combination)	Yes (<i>discounting</i> function)	Yes	No
Golestan <i>et al.</i> [31]	Bayesian networks	Multi-entity BN theories	General (probabilistic)	HLIF (hybrid MEBN inference)	Yes	Yes	No
Chen, Shi <i>et al.</i> [17]	Convolutional Autoencoder Neural Network	Opaque (CANN features)	Healthcare	LLIF (CANN training)	No	No	No
Chen, Hao <i>et al.</i> [16]	Convolutional Neural Network	Opaque (CNN features)	Healthcare	LLIF (multimodal CNN training)	Yes (latent factor model)	No	No
This work	KRR	Description Logics	General (ontology-based)	HLIF (non-monotonic reasoning)	Yes	Yes	Yes (mobile reasoning)

$P1' = \langle C \equiv (\text{hasSlope } \textcolor{red}{\text{only}} \text{ Ascent_Slope}) \text{ and } (\text{hasWeather } \textcolor{red}{\text{only}} \text{ Rain}); X \equiv (\text{hasTraffic } \textcolor{red}{\text{only}} \text{ Low_Traffic}); M \equiv (\text{hasWind } \textcolor{red}{\text{only}} \text{ Strong_Wind}); E \equiv T \rangle$

Afterwards, as shown in Figure 6(b), the same road section is crossed also by vehicles V_4 and V_5 ; with V_4 moving from west to east and behind V_1 , V_5 in the opposite direction. In particular, vehicle V_4 detects low traffic, rain, an uphill road, and classifies that road as secondary. Furthermore, it receives packets $P1'$ and $P5$ (a RSP forwarded by vehicle V_5):

$P4 = \langle C \equiv T; X \equiv T; M \equiv (\text{hasSlope } \textcolor{red}{\text{only}} \text{ Ascent_Slope}) \text{ and } (\text{hasWeather } \textcolor{red}{\text{only}} \text{ Rain}) \text{ and } (\text{hasTraffic } \textcolor{red}{\text{only}} \text{ Low_Traffic}) \text{ and } (\text{hasRoad_Type } \textcolor{red}{\text{only}} \text{ Secondary_Road}); E \equiv T \rangle$

$P5 = \langle C \equiv (\text{hasWind } \textcolor{red}{\text{only}} \text{ Strong_Wind}); X \equiv T; M \equiv T; E \equiv (\text{hasTraffic } \textcolor{red}{\text{only}} \text{ Low_Traffic}) \rangle$

After its BP, vehicle V_4 runs the fusion algorithm again, producing a new GSP $P4'$:

$P4' = \langle C \equiv (\text{hasSlope } \textcolor{red}{\text{only}} \text{ Ascent_Slope}) \text{ and } (\text{hasWeather } \textcolor{red}{\text{only}} \text{ Rain}) \text{ and } (\text{hasTraffic } \textcolor{red}{\text{only}} \text{ Low_Traffic}); X \equiv T; M \equiv (\text{hasRoad_Type } \textcolor{red}{\text{only}} \text{ Secondary_Road}); E \equiv (\text{hasWind } \textcolor{red}{\text{only}} \text{ Strong_Wind}) \rangle$

As evident, the information detected by vehicle V_4 is complemented by data obtained from all other vehicles. This avoids loss of information due to lacking of data sources, as in the case of strong wind not detected by vehicle V_4 (not equipped with an anemometer on-board). In addition, incorrectly detected information is reabsorbed by the system in at most two runs of the fusion algorithm, as in the case of traffic condition: the low traffic sensed by V_1 is first placed in the X annotation of packet $P1'$ (because of the high traffic formerly sensed by vehicle V_2) and then is properly included in the C annotation of packet $P4'$, when V_4 confirms V_1 's observation.

5.2 Experiments

As shown in Figure 7, three scenarios were built, differing for road topology and environmental conditions. Maps were divided in zones each having peculiar road and weather characteristics. Pre-defined annotations pertaining to the different map zones simulated the environmental conditions extraction.

TABLE 2
Scenario A environment description

x/y	$y \leq 1000$	$1000 < y \leq 2000$
$x \leq 2000$	high traffic, weak wind, fog, ascent slope	high traffic, weak wind, fog
$2000 < x \leq 4000$	high traffic, weak wind, fog	high traffic, weak wind, fog, descent slope
$4000 < x \leq 6000$	high traffic, high wind, cloudy, ascent slope	high traffic, high wind, cloudy
$6000 < x \leq 8000$	low traffic, high wind, rain	high traffic, high wind, rain, descent slope
$8000 < x \leq 10000$	low traffic, high wind, rain, ascent slope	low traffic, high wind, rain

Scenario A simulates a typical extra-urban section, with a homogeneous road condition. Scenario B represents a mixed urban and extra-urban path, with a less uniform road configuration w.r.t. the previous one. Finally, Scenario C depicts a typical modern urban district (3x2 km), with a checkerboard arrangement and many intersections with secondary roads. In each map obstacles for wireless signal were deployed, represented as orange segments in Figure 7. In real scenarios such obstacles can be constituted by buildings, walls or other filtering structures. Detailed characteristics of the three scenarios are in Table 2, 3 and 4, respectively.

Various parameters were measured and evaluated, in different operating conditions, to assess the performance of the proposed approach. For each scenario, the simulation was run three times by varying the number of vehicles (25, 50, 75) and mobility "profiles" (as defined by NCTUns and reported in Table 5) randomly assigned to vehicles. In Scenario A, Profile5 (fitting an extra-urban context) was assigned to all vehicles, the broadcast period was set to 7s and the simulation time to 300s. In Scenario B, Profile1 and Profile2 were each assigned to the 30% of vehicles, Profile3 to the remaining 40%; the broadcast period and simulation time had the same value as in Scenario A. Finally, in Scenario C, Profile1 was assigned to the 35% of vehicles, Profile2 to

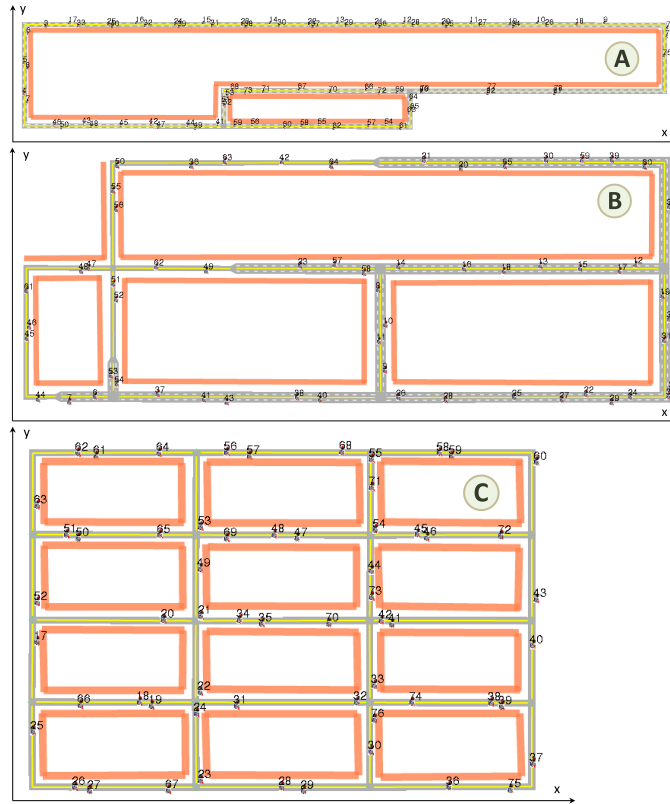


Fig. 7. Map representation of each scenario

TABLE 3
Scenario B environment description

x/y	$y \leq 666$	$666 < y \leq 1333$	$1333 < y \leq 2000$
$x \leq 1666$	secondary road, clear, ascent slope, weak wind	secondary road, clear, descent slope, weak wind	secondary road, clear, ascent slope
$1666 < x \leq 3333$	primary road, cloudy, descent slope, weak wind	primary road, cloudy, ascent slope, weak wind	secondary road, rain, descent slope
$3333 < x \leq 5000$	primary road, rain, ascent slope, weak wind	primary road, cloudy, descent slope, weak wind	primary road, rain, ascent slope

TABLE 4
Scenario C environment description

x/y	$y \leq 666$	$666 < y \leq 1333$	$1333 < y \leq 2000$
$x \leq 1500$	secondary road, rain, ascent slope	secondary road, rain, ascent slope, weak wind	secondary road, cloudy, ascent slope
$1500 < x \leq 3000$	secondary road, rain, descent slope	secondary road, rain, descent slope, high wind	secondary road, cloudy, descent slope

TABLE 5
NCTUns vehicle mobility profiles

	Profile1	Profile2	Profile3	Profile4	Profile5
MaxSpeed (m/s)	18	36	20	8	50
MaxAcceleration (m/s ²)	1	3	10	3	10
MaxDeceleration (m/s ²)	4	5	3	1	20

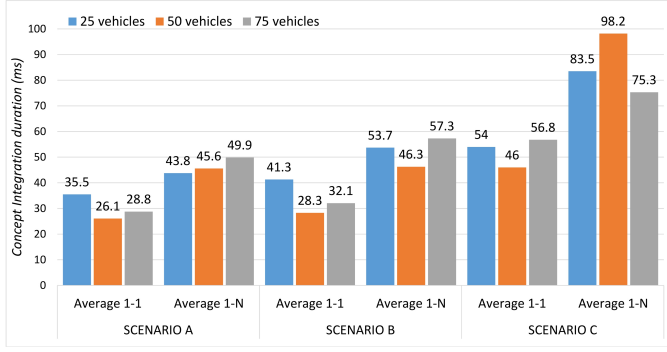
TABLE 6
Average bandwidth usage [kB/s]

	Scenario A			Scenario B			Scenario C		
N. of vehicles	25	50	75	25	50	75	25	50	75
Input BW	0.45	1.46	2.41	1.18	3.53	5.63	0.61	2.86	4.74
Relay BW	1.27	3.26	4.99	3.11	5.93	7.67	1.13	4.31	5.51
Output BW	0.51	0.51	0.52	0.6	0.6	0.6	0.49	0.49	0.49
BW gain	0.89	2.83	4.64	1.98	5.82	9.21	1.25	5.77	9.67

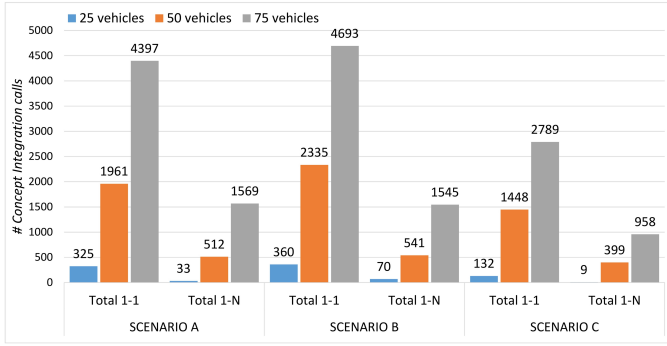
the 10% and Profile3 to the remaining 55%; the broadcast period was set to 9s and the simulation time to 500s. Simulations were performed on a laptop PC with Intel Core i5-3230M CPU at 2.60 GHz, 4 GB of RAM and Windows 8.1 (64bit) operating system. VMWare 6.0.5 build-2443746 virtualization software runned a virtual machine equipped with dual-core processor with 3 GB of RAM and Fedora 12 (Constantine) operating system. The Linux kernel is 2.6.31.6-nctuns20091227, a version customized by NCTUns to enable transparent networking between the VANET simulator and processes on the host machine [4].

Bandwidth usage and network load. First of all, test measured the rate of data exchanged by each vehicle through the VANET in every scenario. As shown in Table 6, the average *output* bandwidth usage was almost constant for all vehicles, as they transmitted just one message at the end of each broadcast period. Conversely, the average *input* and *relay* bandwidth usage –corresponding to incoming packets cached in Store\$ and Relay\$, respectively– depended on the network topology and was more variable, increasing as the density of vehicles grew. Overall network load appeared as rather limited in all scenarios, though. Relay packets always used more bandwidth than the input traffic. Results highlight the benefits of the proposed approach, merging messages received from other vehicles with the self-generated annotation to create a single packet to be sent over the network. Defining the *bandwidth gain* coefficient as the ratio between the input and the output bandwidth produced by a vehicle, its average values for each scenario were reported in Table 6. The knowledge fusion algorithm ensured the vehicle output bandwidth was significantly lower than the input bandwidth in all scenarios with more than 25 vehicles, up to a 9-fold saving in scenarios B-75 and C-75. Scenario A-25 was the only case with slightly inefficient use of bandwidth, due to a very low vehicle density.

Reasoning. With reference to knowledge fusion processing, it was evaluated the average response time of the multithreaded reasoner server. Figure 8(a) describes fusion time (on average and for each scenario), which changes in dependence on the number of vehicles. The *Average 1-1* columns report on the mean time computed limiting the fusion algorithm to the case when a vehicle receives and stores only one packet during the BP; in this special circumstance the fusion algorithm is simplified [23]. The



(a) Mean CIP solution time



(b) CIP calls

Fig. 8. Concept Integration Problem performance results

Average 1-N columns represent the mean time computed with the general version of the fusion applied when an agent receives and stores more than one packet. We obtained an overall average reasoning time of 0.038 s for Scenario A, 0.043 s for Scenario B and 0.069 s for Scenario C. Results highlight that the time spent for information manipulation is compatible with the scheduled BP of each scenario and inversely proportional to their size: this is easily explained with the smaller number of SPs to be processed in case of a lower density of vehicles. Figure 8(b) reports on the calls to the knowledge fusion algorithm. As expected, the number of vehicles is the most important factor, although scenario configuration has some influence. The *Total 1-1* results demonstrate that, limiting the fusion algorithm to the 1-1 case, the number of needed calls would grow very fast w.r.t. vehicle density, making the approach hardly manageable; scalability improves significantly by exploiting multi-packet fusion described in Section 4.1.

Knowledge diffusion. In order to evaluate the effectiveness of the proposed approach, it has been measured the percentage of vehicles which received messages containing correct information about a zone they are moving toward. When a packet containing correct observations about a zone of the map is received, we measured how many seconds in advance it arrives before the vehicle reached that zone. Specifically, we computed the percentage of vehicles that have received at least one message containing correct information before entering the zone of interest. This analysis is interesting to assess the effectiveness of the proposed framework. In fact, vehicles are only interested in useful knowledge characterizing a map zone, but within a VANET

TABLE 7
Percentage of vehicles receiving at least one packet with proper information

N. of vehicles	Scenario A			Scenario B			Scenario C		
	25	50	75	25	50	75	25	50	75
% at 2s	90.0	100	97.3	77.3	91.5	94.4	64.3	78.1	87.9
% at 4s	0.0	0.0	2.7	13.6	4.3	1.4	14.3	9.8	7.6
% at 6s	0.0	0.0	0.0	4.6	2.1	2.8	0.0	4.9	4.6
% at 8s	5.0	0.0	0.0	0.0	0.0	1.4	14.3	4.9	0.0

0% ≤ P < 2%	2% ≤ P < 4%	4% ≤ P < 6%	6% ≤ P < 8%
8% ≤ P < 10%	10% ≤ P < 15%	15% ≤ P < 20%	20% ≤ P

(a) Colors legend

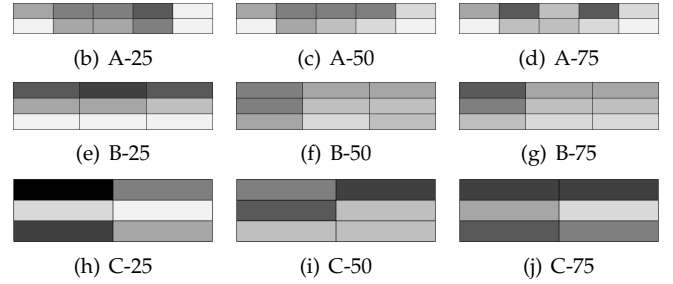


Fig. 9. Percentage of irrelevant packets

incorrect information may be introduced due to erroneous data gathering or other spurious phenomena [24]. The proposed approach aims to ensure that at runtime the execution of the semantic-based fusion algorithm will globally reconcile conflicts due to one-off observations that are not confirmed by other vehicles. To speed up this typically cognitive process, it is appropriate to generate the highest number of messages containing proper information, as soon as possible. Table 7 summarizes the results of such analysis for each reference scenario. The cumulative percentage tends to 100% and increases with the number of vehicles deployed on the map when the vehicle is in close proximity to the target zone (except for Scenario A). This means that the higher the vehicle density and the number of received packets, the greater the probability of receiving at least one packet containing proper and useful information.

Irrelevant information propagation. A final aspect to consider is the percentage of received messages which contain irrelevant information about a zone, defined as annotation fragments in conflict with the description assigned to the zone itself during the map configuration in the NCTUns simulation. In particular, for each map zone it has been measured the percentage (P) of packets containing not useful information with respect to the total number of packets. As the color changes from light grey to black in Figure 9(a), it means a higher number of packets with irrelevant data was found. Figure 9 evidences how Scenario C is the most critical. In fact, only in the central sections of the configuration with 25 vehicles there are light grey zones (Figure 9(h)). The top sections, in all configurations, have at least one zone with critical performance. Globally, it can be observed that scenarios A and B are less critical since they are wider and the exchange of packets is reduced. A clear correlation does not emerge between the number of vehicles in the scenario and the percentage of not relevant packets:

probably the road topology and the vehicle mobility model, which affect data dissemination, are more significant factors and require further investigation to improve performance. Although the knowledge fusion approach grants quick recovery from spurious information propagation, minimizing such phenomenon is of paramount importance for robust knowledge-based IoT scenarios as vehicular ones.

Preliminary large-scale tests. The analysis of systems with higher vehicle number and density is very important to assess performance scalability. A new simulation testbed is being set up in the *Veins* framework [36], capable of managing large-scale simulations efficiently and importing maps from OpenStreetMap for more realistic scenarios. This is an ongoing work whose full implications are still under investigation. Anyway, in early results, with 9 scenarios including 100, 150 and 200 vehicles per kilometer of road (up to 696 vehicles overall), network load did not increase significantly w.r.t. the NCTUns testbed; bandwidth gain was always between 1.03 and 3; the average number of cached semantic packets per broadcast period was between 0.43 and 6.32; knowledge fusion required on average less than 20 ms (due to an optimized simulator-reasoner interface).

6 DISCUSSION

The proposed cognitive computing approach for context awareness in the Semantic Web of Things provides the following peculiar contribution:

- Rich and structured context representation, based on KRR technologies. Descriptions of relevant conditions and phenomena are annotated w.r.t. ontologies, which provide shared vocabularies about reference knowledge domains, endowed with formal machine-understandable semantics.
- A new Concept Integration reasoning task, performing knowledge fusion among observations of independent entities with reconciliation of inconsistencies. It provides quick adaptation to changes and robustness against spurious events due to missing values or inaccuracies in input data. In addition to algorithmic specification, formal properties and complexity considerations were discussed in Section 4.1.
- The above knowledge-based framework was purposely integrated in a bidirectional data dissemination protocol for vehicular networks, able to propagate the information for situation awareness among OBUs and RSUs.

The proposed knowledge fusion algorithm is inherently general-purpose. It applies to intrinsically unpredictable scenarios populated by moving, variable and volatile entities as typically happens in the Internet of Things. Application areas may include complex cyber-physical systems, robot team coordination, industrial IoT, affective intelligence in human-machine interaction and more. Since multiple ontologies could be managed at the same time and fusion can occur only among annotations referring to the same ontology, some of the above applications may require preliminary ontology agreement among entities (agents) as well.

The main uncovered issue is security. Classical VANET security approaches are based on public key infrastructures [37]. Recently, blockchain has been proposed in VANETs as a more flexible and scalable way to guarantee data integrity through trustless cooperation of nodes [38]. The proposed framework does not prescribe any specific solution. In principle, any security protocol could be transparently added allowing datagram exchange, without loss of functionality. Nevertheless, experimental activities will be required mostly to assess the impact on performance.

Scalability study in very large and dense vehicular networks is ongoing. The NCTUns simulation testbed allowed to verify the correctness of the proposal and satisfactory performance results in low and medium density scenarios. Inherent architectural limitations of the tool make the setup of much larger simulations difficult. A new testbed is being finalized within *Veins*. A summary of preliminary results was included in Section 5.2 to show the approach has promising scalability properties.

7 CONCLUSION AND FUTURE WORK

The paper proposed a novel knowledge-based approach for information fusion in the Internet of Things. An application to enhance context awareness in VANETs is presented as proof of concept. It includes knowledge integration and alignment through inferences on semantic annotations formalized in Description Logics. Is is granted a general applicability to dynamic scenarios. A dissemination protocol designed for vehicular networks enables efficient information propagation. Peculiarities of the proposal make it suitable to robust distributed context monitoring, even in the presence of incomplete and/or inaccurate information detection by individual entities. Early experiments demonstrated the effectiveness of the approach.

Future work encompasses several aspects. First of all a proper context annotation framework exploiting semantic-enhanced data mining and machine learning techniques must be devised to setup a full technological stack. Furthermore, integration of RSUs in the framework is straightforward with the already defined elements, and will enable further services such as real-time information update on back-end map repositories to provide web-based rich information monitoring. Finally, performance improvements are ongoing, driven by a novel extended simulation campaign, to sustain even larger and more complex scenarios. They will be continuously tested before the expected real implementation of the full framework.

REFERENCES

- [1] On-Road Automated Driving (Orad) Committee, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," *SAE International Standards*, 2016.
- [2] M. Ruta, F. Scioscia, F. Gramegna, and E. Di Sciascio, "A mobile knowledge-based system for on-board diagnostics and car driving assistance," in *UBICOMM 2010, The Fourth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, 2010, pp. 91–96.
- [3] S. Ilarri, T. Delot, and R. Trillo-Lado, "A data management perspective on vehicular networks," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2420–2460, 2015.
- [4] S. Wang and C. Chou, "Nctuns tool for wireless vehicular communication network researches," *Simulation Modelling Practice and Theory*, vol. 17, no. 7, pp. 1211 – 1226, 2009.

- [5] WG802.11 - Wireless LAN Working Group, "IEEE Standard for Information technology – Local and metropolitan area networks – Specific requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments," *IEEE Standard*, pp. 1–51, 2010.
- [6] C. Campolo, A. Molinaro, A. Vinel, and Y. Zhang, "Modeling prioritized broadcasting in multichannel vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 687–701, 2012.
- [7] Q. Wang, S. Leng, H. Fu, and Y. Zhang, "An IEEE 802.11 p-based multichannel MAC scheme with channel coordination for vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 449–458, 2012.
- [8] R. Yu, J. Ding, X. Huang, M.-T. Zhou, S. Gjessing, and Y. Zhang, "Optimal resource sharing in 5g-enabled vehicular networks: A matrix game approach," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 7844–7856, 2016.
- [9] M. Chaqfeh, A. Lakas, and I. Jawhar, "A survey on data dissemination in vehicular ad hoc networks," *Vehicular Communications*, vol. 1, no. 4, pp. 214–225, 2014.
- [10] T. Nadeem, P. Shankar, and L. Iftode, "A Comparative Study of Data Dissemination Models for VANETs," in *2006 3rd Annual International Conference on Mobile and Ubiquitous Systems - Workshops*, July 2006, pp. 1–10.
- [11] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. Patel-Schneider, *The Description Logic Handbook*. Cambridge University Press, 2002.
- [12] W3C OWL Working Group, "OWL 2 Web Ontology Language Document Overview (Second Edition)," *W3C Recommendation*, 2012, <https://www.w3.org/TR/owl2-overview/>.
- [13] M. Ruta, E. Di Sciascio, and F. Scioscia, "Concept Abduction and Contraction in Semantic-based P2P Environments," *Web Intelligence and Agent Systems*, vol. 9, no. 3, pp. 179–207, 2011.
- [14] M. Ruta, F. Scioscia, E. Di Sciascio, and G. Piscitelli, "Semantic matchmaking for location-aware ubiquitous resource discovery," *International Journal On Advances in Intelligent Systems*, vol. 4, no. 3/4, pp. 113–127, 2012.
- [15] G. Teege, "Making the Difference: A Subtraction Operation for Description Logics," *Proceedings of the Fourth International Conference on the Principles of Knowledge Representation and Reasoning (KR94)*, pp. 540–550, 1994.
- [16] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease Prediction by Machine Learning over Big Data from Healthcare Communities," *IEEE Access*, 2017.
- [17] M. Chen, X. Shi, Y. Zhang, D. Wu, and M. Guizani, "Deep Features Learning for Medical Image Analysis with Convolutional Autoencoder Neural Network," *IEEE Transactions on Big Data*, 2017.
- [18] M. Chen, J. Yang, Y. Hao, S. Mao, and K. Hwang, "A 5G cognitive system for healthcare," *Big Data and Cognitive Computing*, vol. 1, no. 1, p. 2, 2017.
- [19] E. P. Blasch, D. A. Lambert, P. Valin, M. M. Kokar, J. Llinas, S. Das, C. Chong, and E. Shahbazian, "High level information fusion (HLIF): Survey of models, issues, and grand challenges," *IEEE Aerospace and Electronic Systems Magazine*, vol. 27, no. 9, pp. 4–20, 2012.
- [20] F. Terroso-Sáenz, M. Valdés-Vela, F. Campuzano, J. A. Botia, and A. F. Skarmeta-Gómez, "A complex event processing approach to perceive the vehicular context," *Information Fusion*, vol. 21, pp. 187–209, 2015.
- [21] G. Cevolani, "Truth approximation, belief merging, and peer disagreement," *Synthese*, vol. 191, no. 11, pp. 2383–2401, 2014.
- [22] P. Liberatore and M. Schaerf, "Arbitration (or how to merge knowledge bases)," *IEEE Transactions on Knowledge and Data Engineering*, vol. 10, no. 1, pp. 76–90, 1998.
- [23] F. Scioscia, M. Ruta, and E. Di Sciascio, "A swarm of Mini-MEs: reasoning and information aggregation in ubiquitous multi-agent contexts," in *4th OWL Reasoner Evaluation Workshop (ORE 2015)*, ser. CEUR Workshop Proceedings, vol. 1207. CEUR-WS, jun 2015, pp. 15–22.
- [24] F. Ghaleb, A. Zainal, and M. Rassam, "Data verification and misbehavior detection in vehicular ad-hoc networks," *J. Teknologi*, vol. 73, no. 2, pp. 37–44, 2015.
- [25] F. Sattar, F. Karray, M. Kamel, L. Nassar, and K. Golestan, "Recent advances on context-awareness and data/information fusion in ITS," *International Journal of Intelligent Transportation Systems Research*, vol. 14, no. 1, pp. 1–19, 2016.
- [26] J. Santa, A. Muñoz, and A. F. Gómez-Skarmeta, "Providing adapted contextual information in an overlay vehicular network," *International Journal of Digital Multimedia Broadcasting*, vol. 2010, 2010.
- [27] X. Hu, J. Zhao, B.-C. Seet, V. C. Leung, T. H. Chu, and H. Chan, "S-Aframe: agent-based multilayer framework with context-aware semantic service for vehicular social networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 1, pp. 44–63, 2015.
- [28] M. Rohani, D. Gingras, V. Vigneron, and D. Gruyer, "A new decentralized Bayesian approach for cooperative vehicle localization based on fusion of GPS and VANET based inter-vehicle distance measurement," *IEEE Intelligent Transportation Systems Magazine*, vol. 7, no. 2, pp. 85–95, 2015.
- [29] J. Radak, B. Ducourthial, V. Cherfaoui, and S. Bonnet, "Detecting road events using distributed data fusion: Experimental evaluation for the icy roads case," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 184–194, 2016.
- [30] M. B. Farah, D. Mercier, F. Delmotte, and É. Lefèvre, "Methods using belief functions to manage imperfect information concerning events on the road in VANETs," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 299–320, 2016.
- [31] K. Golestan, B. Khaleghi, F. Karray, and M. S. Kamel, "Attention assist: A high-level information fusion framework for situation and threat assessment in vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 5, pp. 1271–1285, 2016.
- [32] W3C OWL Working Group, "OWL 2 Web Ontology Language Manchester Syntax (Second Edition)," *W3C Working Group Note*, 2012, <https://www.w3.org/TR/owl2-manchester-syntax/>.
- [33] D. Dubois, W. Liu, J. Ma, and H. Prade, "The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks," *Information Fusion*, vol. 32, pp. 12–39, 2016.
- [34] F. Scioscia, M. Ruta, G. Loseto, F. Gramegna, S. Ieva, A. Pinto, and E. Di Sciascio, "A Mobile Matchmaker for the Ubiquitous Semantic Web," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 10, no. 4, pp. 77–100, 2014.
- [35] F. Bai, N. Sadagopan, and A. Helmy, "IMPORTANT: a framework to systematically analyze the Impact of Mobility on Performance of Routing Protocols for Adhoc Networks," in *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428)*, vol. 2, March 2003, pp. 825–835 vol.2.
- [36] C. Sommer, R. German, and F. Dressler, "Bidirectionally coupled network and road traffic simulation for improved IVC analysis," *IEEE Transactions on Mobile Computing*, vol. 10, no. 1, pp. 3–15, 2011.
- [37] H. Hasrouny, A. E. Samhat, C. Bassil, and A. Laouiti, "VANet security challenges and solutions: A survey," *Vehicular Communications*, 2017.
- [38] P. K. Sharma, S. Y. Moon, and J. H. Park, "Block-VN: A Distributed Blockchain Based Vehicular Network Architecture in Smart City," *Journal of Information Processing Systems*, vol. 13, no. 1, pp. 184–195, 2017.