Journal of Information Hiding and Multimedia Signal Processing ©2015 ISSN 2073-4212
Ubiquitous International Volume **6**, Number **6**, November **2015**

An Efficient Traffic Data Aggregation Scheme for WSN Based Intelligent Transportation Systems

Ziyi You, Shiguo Chen and Yi Wang

Department of Physics & Electronic Science
Guizhou Normal University
Guiyang, 550001, China

Key Laboratory of special Automotive Electronics
technology of the Education Department of Guizhou Province
Guiyang, Guizhou, 550001, China
357534271@qq.com; 740461512@qq.com; wyigz@126.com

Received May, 2015; revised August, 2015

Abstract. *The data aggregation with multi-resources traffic information is a very important issue for intelligent vehicle systems (ITS). It can not only collect the most critical data from the traffic environment, but also prevent sensor network congestion. Unfortunately, we so far have no suitable solutions for WSN based ITSs his paper presents an efficient approach for traffic data aggregation, applied in ITSs. At first, our approach adopts a hybrid network structure which is the combination of the chain structure and the unequal clusters structure. In this structure, all sensor nodes from the same cluster use parameters such as leading code and geographical position etc. to guarantee secure data aggregation. Furthermore, a method is introduced depending on credibility evaluation and reliability allocation so that the application layer can calculate aggregation results accurately, and then makes the decisions. Finallythe performance of the proposed scheme has been verified using simulation, showing that it is superior to similar protocol VLEACH (an improvement on LEACH) and ESDA(Efficient and Secure Data Aggregation protocol) such as the sensor nodes energy consumption and aggregation precision. The analysis result indicates that our scheme is effective and feasible in the next generation of sensor technologies of ITSs.*

Keywords: Intelligent vehicle system, Data aggregation, credibility, Evidence theory, Security analysis, Performance evaluation.

1. **Introduction.** The traffic information collection becomes the key problem of ITSs. Acquiring accurate traffic parameters, i.e. traffic volume and the velocity etc., is the fundament of traffic management. Original intelligent traffic systems generally depend on traditional monitoring sensors including inductive loops, video cameras and ultrasonic sensors .However, due to the limitation of these sensors, some disadvantages appear, which affect the efficiency of the entire sensor network and the scalability of ITSs So far, wireless sensor nodes are small-sized, densely deployed, power-efficient, and self-configured, which can work in autonomous manner to sense the surroundings. Therefore, incorporating wireless sensor networks (WSNs) into ITSs can overcome the problems associated with traditional wired sensors. Multi-resources traffic information fusion based on WSNs will provide higher accuracy of traffic statics than traditional wired sensors. It will help to realize more efficient traffic applications including electronic toll collection, parking management, intersection traffic guidance, energy conservation etc [1].

As the development of WSN architecture of ITSs, it emerges the hybrid type, which is the combination of roadway infrastructure based on WSN and selforganizing adhoc networks formed among vehicles. There are also two types of information communications namely, Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I). In V2V vehicles equipped with sensors exchange information between each other in order to avoid severe situations like traffic jam avoidance While in V2I sensor nodes installed on roadway infrastructure are deployed at important places to collect / provide information from / to vehicles This communication is very important on the roadway, especially for the timely feedback on traffic conditions on the freeway.

So far, many data aggregation schemes have been proposed for WSNs. A secure data aggregation algorithm based on pattern codes (ESDA) was designed by Liu Wei [2], where pattern codes are used to make the data aggregation in sensor networks to ensure the security of collected data .The data status is obtained by the sensor nodes through fuzzy algorithm, and sensor nodes only transfer the pattern codes of emergent data and ordinary data to the cluster head for process, which realizes the minimization of network energy consumption. Hussein T. Mouftah, et al [3] proposed a novel framework for secure information aggregation in large sensor networks to ensure the security of collected data. In the novel framework, certain nodes in the sensor network, called aggregators, help aggregating information requested by query, which not only perform the aggregation tasks correctly, but also reduces the communication overhead substantially. Leach [4] firstly introduces the cluster-based approach. It utilizes randomized rotation of local cluster heads to evenly distribute the energy load among the sensors which minimizes energy dissipation in sensor networks.

In the scenario of WSN based ITSs, the traffic information is detected from many different kind of sensors installed on the roadside or on the vehicles, and it contains large volumes of raw data like sound, airflow and temperature and so on. What is more, vehicles have the characteristic of high mobility which makes vehicle sensor networks more frequently changes in topology compared to the traditional static sensor networks. Hence, we can see advantages of those data aggregation methods mentioned above, but all of them are not suited for ITSs as their inadequacies i.e. relatively complicated algorithm, large energy consumption and longtime delay and so on.

This paper contributes by presenting a dynamic clustering data aggregation scheme based on WSN, and applied in the collection and management of multi-resources traffic information in ITSs. The scheme meets the demand of temporal and spatial distribution characteristics in traffic travel [5, 6]. In addition, it reduces the node energy consumption, and improves the authenticity of sensing data by using node credibility evaluation method. After security and performance analysis, we conclude that the proposed scheme is valid and feasible.

The rest part of the paper is organized as follows. Section 2 discusses the evidence theory based on data aggregation. The scheme is described in detail in section 3. We analyze its security in Section 4 and followed by reliability test in Section 5. Section 6 focuses on the performance simulation analysis of the scheme. Finally, in Section 7, the conclusion is given.

2. **Data aggregation theory based on evidential reasoning.** Dempster-Shafer [7] theory which is closer to people's thinking logical and natural decision-making process offers an alternative to traditional probabilistic theory for the mathematical representation of uncertainty. Therefore, in this paper, we devised the scheme for ITSs based on Dempster-Shafer evidential reasoning method.

In the field of data aggregation, Depmster Shafer theory is usually used for target identification and classification. Considering this target identification method, each sensor separately finish the local decision which will be aggregated together and sent to the information centre to get the final decision-making. The related basic concept is described as follows.

Let L be a finite propositional language. Denote by $\Theta = \{w_1, w_2, ..., w_n\}$ the non-empty world set of n possible outcomes (alternatives) of the event of interest. Where, $w_i (i = 1, 2, ..., n)$ is an explanation of L.

Symbol θ' usually can be expressed as a subset of set Θ. The set Θ has the following characteristics.

1. The finite nature.
2. All the inner elements are mutually exclusive.

There are three important functions in Dempster-Shafer theory: the basic probability assignment function (*bpa* or m), the Belief function (*Bel*), and the Plausibility function (*Pl*).

Definition 2.1. *Let Θ be framework for identifying, the basic probability assignment function is a mapping $m : 2^{\Theta} \to [0, 1]$ such that $m(\varphi) = 0$ and $\sum_{\theta \in 2^{\Theta}} m(\theta) = 1$. If $m(\theta) > 0$, we will call $m(\theta)$ the mass of θ, and the pairing of set θ with its corresponding masses $m(\theta)$ the body of evidence of m.*

Definition 2.2. *the Belief function (Bel) is a mapping $Bel : 2^{\Theta} \to [0, 1]$. Belief of a set θ is the sum of the masses assigned to all subsets (θ) of θ : $Bel(\theta) = \sum_{\rho \subseteq \theta} (\rho)$. A pixel or object is assigned to class θ if the amount of belief in support of θ is larger than that supporting its negation and the other single class hypotheses.*

Definition 2.3. *the Plausibility function (PI) is a mapping $PI : 2^{\Theta} \to [0, 1]$. Plausibility of a set θ is the sum of the masses of all sets (ρ) having non-empty intersection with $\theta PI(\theta) = \sum_{\theta \cap \rho \neq \varphi} m(\rho)$.*

We can conclude that it is not required for the sum of all the Belief measures to be 1 and similarly for the sum of the Plausibility measures, and these two measures can be derived from each other: $PI(\theta) = 1 - Bel(\overline{\theta})$, where $PI(\theta) \geq Bel(\theta)$.

A key feature of the Dempster-Shafer theory is the rule for combining bodies of evidence. The combination (called the joint m_{12}) is calculated from the aggregation of two probability assignments m_1 and m_2 in the following manner:

$$\begin{aligned} & m_{12}(\theta) = K \sum_{\theta i \cap \rho j = \theta} m_1(\theta_i) m_2(\rho_j) \\ & \text{when } \theta \neq \varphi \text{ and } m_{12}(\varphi) = 0 \\ & \text{where } K = [\sum_{\theta i \cap \rho j \neq \varphi} m_1(\theta_i) m_2(\rho_j)]^{-1}. \end{aligned} \tag{1}$$

K represents basic probability mass associated with conflict. The finite sequences $\theta_1, ..., \theta_p$ and $\rho_1, ..., \rho_q$ are respectively assigned to m_1 and m_2.

The rule above can be extended to multiple m functions and Bel functions for integrated multiple expert advice. In view of this, we adopted Dempster-Shafer theory to address the traffic problem posed by strongly conflicting evidence from multiple sensors.

3. The data aggregation scheme.

3.1. network model.

Suppose the network deployment in an ITS is a multi-tiered and WSN based cluster structure as shown in Fig1. All sensor nodes installed in vehicles or on roadway infrastructure in this architecture are arranged in clusters Each cluster is constituted of a collector (the cluster head) installed on the roadside and one or more passing vehicles (the cluster members). The cluster head periodically collects traffic information from the cluster members, which will be sent to the nearest sink node. Traffic data are transferred from vehicles to the sink nodes, from sink nodes to vehicles and among the sink nodes themselves.

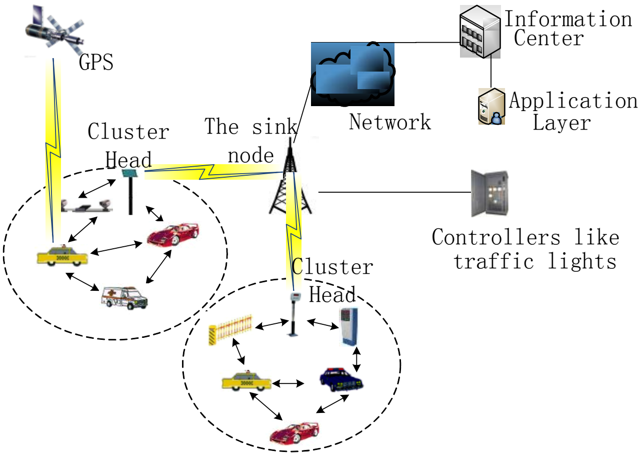

FIGURE 1. A WSN-based ITSs

The management of traffic information depends on data-based gridding urban traffic system. Assume that all the sensor nodes (vehicles and infrastructures) randomly distribute in a two-dimensional plane rectangular region: $Z^2 = \{(x, y), 0 \le x, 0 \le y\}$. This region can be a grid model where principles and methods [8, 9] are proposed to deploy the collection equipment for flow data Every square grid is defined as $\alpha \times \alpha$, the value of α depends on the accuracy demand of certain task application. Let the positions of P and Q be respectively $L_P(x_j, y_j)$ and $L_Q(x_j, y_j)$, then the distance of P and Q is:

$$d = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{2}$$

The region is divided into A_{head} partitions, the number of which is determined by the length of the region and the range of communication radius of the nodes in each partition. The sink node is deployed at the fix position outside the region. In order to balance energy consumption, the number of clusters in the partition closer to the sink node should be more than the partitions farther from the sink node. Due to the mobility of vehicles, each cluster head is designated by the sink node as an infrastructure in the corresponding cluster. There is only single hop communication between the cluster head and its cluster members, and the cluster head can adjust the communication radius to control the number of cluster members, which reduces the communication load. For the two adjacent cluster heads CH_i and CH_j, Eq. (3) is given:

$$D(CH_i, CH_j) \ge R_{CH_i} + R_{CH_j} \tag{3}$$

Where, $D(CH_i, CH_j)$ denotes the distance between CH_i and CH_j, R_{CH_i} and R_{CH_j} denote the corresponding radius respectively. Obviously, there may be some sensor nodes not in any cluster, which form the chain architecture [16] for multi hop communication.

Depending on the greedy algorithm, they select the intermediate nodes which have the strongest signal to forward the sensing data as shown in Fig 2. Then, a member node of one cluster as the chain head aggregates all the messages it receives, and sends the new messages to the cluster head.

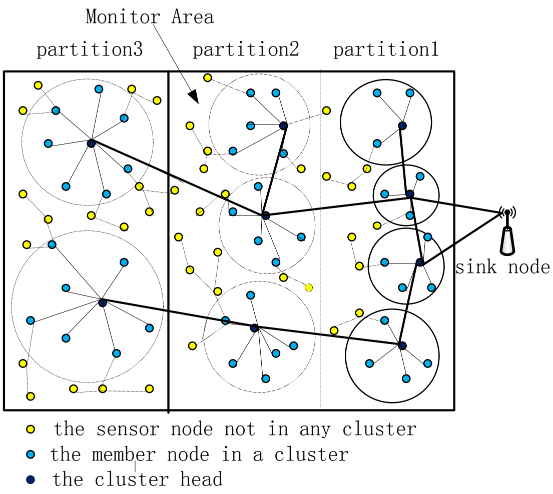

Figure 2. The proposed network architecture for ITSs

3.2. **data aggregation process in one cluster.** In a cluster region, the cluster head obtains other members' sampling data, which will be integrated into the non-redundant data set. Then, he generates a new message by GPS sent to the nearest sink node including the geographic location and timestamp. AS the sink node receives plenty of flow information from multiple clusters, it can perform classifications based on the accurate timestamps and geographic locations, and then separately upload them to the application layer. Application layer acts as the decision maker computes the aggregation result by using probability redistribution method for Dempster-Shafer theory, and gives timely decision for specific traffic application. It has been proved in section 5 and 6 that the proposed scheme ensures the accuracy for multi-source traffic data fusion, while prolonging the network lifetime.

Suppose a cluster head A_{head} periodically sends the inquiry message $MSGREQ$ to its members in order to synchronize time. In a sampling round t_i, each member node A_i (vehicle or roadside facility) separately sends a response message REP_i to the cluster head A_{head} as follows.

$$m_i = MSG\_REP||D_\tau||SN_A||Dat||Pos||Sensid||Token$$

Where:

D_τ is the leading code of message m_i, which represents the actual source data sensed by A_i. Such code can be used to distinguish between different data packet formats and data types.

SN_A is the identity of A_i.

Dat is the data field, and $Dat = E(\widetilde{d}_{sens}, K_{A_i,Sink})$. Symbol $K_{A_i,Sink}$ denotes the shared key between node A_i and the sink node. Symbol $\widetilde{d}_{sens}$ denotes the actual data sensed by A_i including the support degree for the evidence of making decision, and encrypted using $K_{A_i,Sink}$.

Pos is the couple (x, y)describing the geographical position of the node A_i, which can be used for target localization [10, 11].

$Sensid$ is the identifier of the message m_i which can be used to distinguish it from other messages transmitted by the same node A_i. We can get that $Sensid = f(Sensid') \bmod M$. In this formula, function f is monotonically increasing, and $Sensid'$ is the identifier of the previous message transmitted by A_i.

$Token$ is the authentication token of message m_i and $Token = SIG(D_\tau||SN_A||Dat||Pos ||Sensid, K_{A_i,TP})$. Symbol SIG denotes the digital signature of message m_i signed by node A_i, while symbol $K_{A_i,TP}$ denotes the private key of A_i distributed by the authentication centre on network layer.

In the time interval Δt , all the sensing data inside or outside a cluster are transmitted from the collection location or through chain heads to the cluster head node. Cluster head then selects the same attribute data to join together to generate a new message again. During this process, due to closely space between two members, their sensing range often overlaps. Furthermore, they might get the same data. In addition, sensor nodes can access a variety of data including traffic, speed, road share etc. which require classification according to different attributes. The value D_τ of leading code field in the packet format replaces the actual data for data aggregation ,which can reduce the redundant data in the network so as to save the nodes energy [12]. The Pos field also can avoid that the same data is aggregated into more than one group, which increases network congestion.

The data having the same attribute value are joined together as the same group, and therefore a new message is generated for each aggregation group. Let be G an aggregation group composed of l different messages, cluster head A_{head} aggregates the Dat fields of the l messages in G and computes Pos_{aggr} value. Then, new message m_{aggr} is generated and sent to the sink node which is structured as follows.

$m_{aggr} = MSG\_REP_{BS}||SN_{head}||data_{aggr}||Pos_{aggr}||Time_{stamp}||Token'||IdList$, where:

SN_{head} is the identity of A_{head}.

$$Dat_{aggr} = \sum_{i=1}^{l} m_i.Dat$$

$$Pos_{aggr} = \left(\frac{1}{w}\sum_{i=1}^{l} w_i m_i.Pos.x, \frac{1}{w}\sum_{i=1}^{l} w_i m_i.Pos.y\right).$$

Notice that w_i equals the size of the field D_τ, so w is equal to the sum of w_i, that is $w = \sum_{i=1}^{l} w_i$.

$Time_{stamp}$ is the timestamp, and the sink node can perform stateful classifications based on the accurate timestamps.

$Token'$ is the authentication token of message m_{aggr}, that is $Token' = SIG(SN_{head} ||data_{aggr}||Pos_{aggr}||time_{stamp}||IdList, K_{Ahead,TP})$.

Symbol $K_{Ahead,TP}$ denotes the private key of A_{head}.

$IdList$ is a list of nodes that aggregated the D_τ fields of N messages in G, that is $IdList = E(SN_{head}||SN_{A1}...|| \quad SN_{Ai}...||SN_{AN}, K_{Ahead,BS})$. Symbol $K_{Ahead,Sink}$ denotes the shared key between node A_{head} and the sink node.

Upon receiving the message m_{aggr} and checking that it is valid, the sink node accepts the data and forwards to the application layer of Information Centre. Otherwise, abandons m_{aggr} and updates the relevant nodes' credibility. The whole data aggregation process is described in Figure 3.

3.3. calculation and evaluation of the aggregation result.

After receiving the data which the sink node sends, the application layer calculates the function values for evidential reasoning, and then gives final decision through the combination rule. For this purpose, it is necessary to list all basic attribute values of the sensing object according to

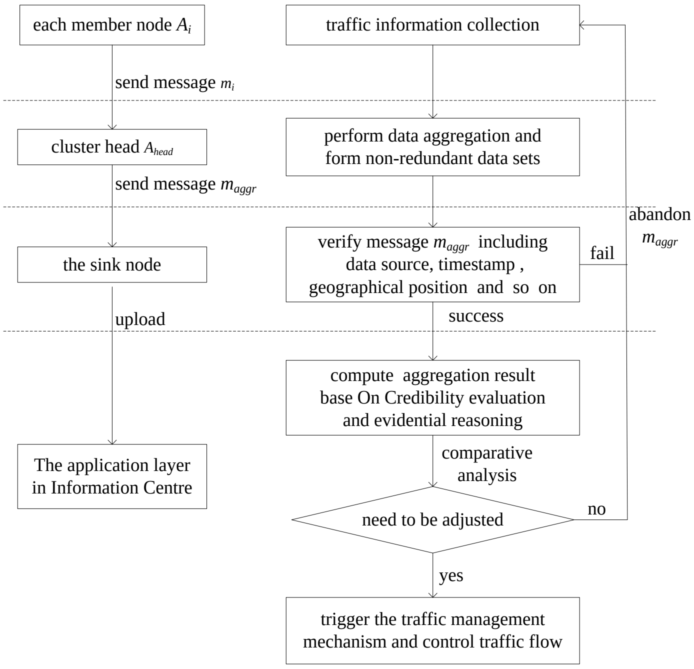
each member node A_i
send message m_i
cluster head A_{head}
send message m_{aggr}
the sink node
upload
The application layer in Information Centre
traffic information collection
perform data aggregation and form non-redundant data sets
verify message m_{aggr} including data source, timestamp , geographical position and so on
fail
abandon m_{aggr}
success
compute aggregation result base On Credibility evaluation and evidential reasoning
comparative analysis
need to be adjusted
no
yes
trigger the traffic management mechanism and control traffic flow

Figure 3. The data aggregation process

existing recognition framework Θ , so as to obtain the Decision Classification Set Uthat is a collection of the subsets of all the attributes in Θ [13, 14]. For instance, to detect whether a vehicle or more are running on this driveway, the recognition framework is defined as $\{0, 1\}$, where 0 denotes no, and 1 denotes yes. So, the corresponding set U is equal to the $\{\{0\}, \{1\}, \{0 \text{ or } 1\}, \varphi\}$, where symbol φ denotes zero elements.

The basic probability assignment function (*bpa*) is the basis of evidence reasoning theory. Therefore, to judge an event whether to be true or false, each of evidences used to determine this event has to be assigned a *bpa* function, and the *bpa* function makes that the set U is mapped to range $[0, 1]$, i.e. $m : 2^{\Theta} \longrightarrow [0, 1]$. Given a measurement θ, and if θ is a nonzero value, $m(\theta)$ is defined as the degree of support for A in the function m. By Definition 1, we can get that $m(\varphi) = 0, \sum\limits_{\theta \in 2^{\Theta}} m(\theta) = 1$. In the case of multi-sensor and multi-data source, each data source A_i, in terms of own *bpa* function m$_i$, respectively makes the judgment under the same recognition framework Θ. Finally, the application layer effectively merges the bodies of evidence of m$_i$ using the combination rule, and generates the decisions for traffic control [15].

In this paper we introduce a probability allocation method for function m$_i$ based on node credibility. Suppose the credibility of A_i is defined as r_i, whose value is initialized to 5 (the 5 is just a hypothetical value, in practice, this value may be other figure). Notice that, during every round, we compare the source node A_i's judgement with the application layer's decision, if the result is the same, r_i plus 1; otherwise r_i minus 1. Next, do the functions transform for r_i and define the weight coefficient w_i as follows.

$$w_i = \begin{cases} 0, \ r_i \leq -g \text{ or } \alpha_i \leq \lambda_0 \\ \dfrac{r_i + g}{\max(r_i) + g}, \ r_i > -g \text{ and } \alpha_i > \lambda_0 \\ \alpha_i = \dfrac{N_{success}}{N_{total}} \end{cases} \tag{4}$$

where $\max(r_i)$ denotes the maximum value of historical record of $r_i$in the region, $g$denotes the constant coefficient and $g = 5$ which corresponding to the initial value of r_i. Symbol α_i denotes the detection accuracy for the source node A_i, i.e. the proportion of the $N_{success}$ rounds (the judgement of A_i and the decision of the application layer remain consistent) in the total N_{total} rounds (A_i participates in sensing data).

If there are n data sources, considering Eq. (4), the Bel function of node $A_i$for the measurement θ:

$$\begin{cases} Bel_i(\theta) = m_i'(\theta) = w_i m_i(\theta), \theta \in \Theta \\ Bel_i(\overline{\theta}) = 1 - Bel_i(\theta), \overline{\theta} = \Theta - \theta \end{cases}, 1 \le i \le n \tag{5}$$

The PI function of node A_i for the measurement θ:

$$PI(\theta) = \sum_{\rho i | \rho i \cap \theta \neq \varphi} m(\rho_i) \tag{6}$$

The Bel function and PI function meets the following Eq. (5).

$$\begin{cases} PI(\theta) = 1 - Bel(\overline{\theta}) \ , \overline{\theta} = \Theta - \theta \\ PI(\theta) \ge Bel(\theta) \end{cases} \tag{7}$$

Suppose I is the largest subset of U for recognizing the event. According to Eq. (1), (4) and (5), the Bel functions $Bel_i(\theta_j)$ $(1 \le i \le n, 1 \le j \le p)$ from n data sources are merged together, and the new Bel functions $Bel(\rho_l)$ $(\rho_l \subseteq \overline{I}$ and $1 \le l \le m)$are generated as follows [12].

$$Bel(\rho_l) = m_1'(\theta_j) \oplus ... \oplus m_n'(\theta_j) = k \sum_{\theta_1 \cap ... \theta_p = \rho_l} \prod_{i=1}^{n} m_i'(\theta_j) \tag{8}$$

where $k = [\sum_{\theta_1 \cap ... \theta_p = \varphi} \prod_{i=1}^{n} m_i'(\theta_j)]^{-1}$ and $k < 1$.

As known that, between $\rho_1, ..., \rho_m$, any two are mutual exclusive. According to Eq. (6), (7) and (8), the value of the plausibility function $PI(I)$ can then be obtained by:

$$PI(I) = 1 - \sum_{\rho_1 \cup \rho_2 \cup ... \rho_m = \overline{I}} Bel(\rho_j) \tag{9}$$

Finally, the application layer can make the decision depending on the comparison of the values of $PI(I)$ and η_0 as described in Eq. (10).

$$\begin{cases} PI(I) \ge \eta_1, \text{executing decision } H_1 \\ PI(I) \le \eta_0, \text{executing decision } H_0 \\ \eta_0 < PI(I) < \eta_1, \text{waiting for the next decision} \end{cases} \tag{10}$$

Notice that, the value ofη_0 and η_1 are given on the analysis of a large amount of figures. As can be seen from Eq. (4) and (5), once the weight coefficient $w_i < 1$, the value of function $Bel_i(\theta)$ actually decreases, while for the set $\overline{\theta}$, the value of function $Bel_i(\overline{\theta})$ is on the increase. After several rounds of sampling, the higher the credibility of A_i becomes, the more close to 1 the weight coefficient w_i is going to be, on the contrary, close to 0 the w_i is. Hence, the function $Bel_i(\theta)$ tends to 0 which implicates the belief degree of A_i on the evidence θ decreases, while the possibility that θ can be neglected increases. In this way, depending on Eq. (4), (5), (7), (8) and (9), the influence of low reliability of the aggregation results caused by the dishonest nodes will be weakened.

4. **Security analysis of the proposed scheme.** In this section, we provide a security analysis of the proposed scheme above. Our scheme not only provides a privacy policy for the actions performed by nodes in the data aggregation process, but also greatly improves the authenticity and reliability of the detection results in multi-source traffic environment

1. The privacy policy in data aggregating ensures the traffic information collection. At first, the source node A_i, through the encryption of Dat field of the data packet makes the actual sensing data not leaked. Furthermore, A_i digitally signs the whole data packet to generate $Token$ using the private key $K_{A_i,TP}$, which ensures the integrity of the transmitted message. Then, the cluster head A_{head} verifies the received messages and generates a new aggregated message m_{aggr}. If the field $Token'$ of a received message m_i does not match with other corresponding fields of m_i, this message should be considered as corrupted and therefore be dropped. In the same way, the sink node which receives flow information from multiple clusters can verify the aggregated messages and perform stateful classifications based on the accurate timestamps and the geographical positions.

2. The leading code can replace the actual data for data aggregationwhich forms non-redundant data sets so as to save node energy. Therefore, through the D_τ field of the packet format, the cluster head node can directly aggregate the messages from the cluster members, without knowing the actual data values By the complexity of the D_τ generation algorithm, we believe that it is impossible for the attacker to derive the actual data from the D_τ field, which can prevent from some routing attacks like wormhole [18] and routing loops [19] for "inner network" data processing, so that improves the overall performance of the security mechanism.

3. The application layer provides the corresponding weight coefficient w_i based on the reputation evaluation for the nodeA_i, so that the weighted *bpa* function $m_i'(\theta)$ can be calculated. Then, the weighted *bpa* functions of n source nodes i.e. $m_1'(\theta)...m_i'(\theta)...m_n'(\theta)$ are merged by means of reliability allocation, and thus the *Pl* function $PI(I)$ is computed to make the decision for traffic management. This makes up for the weakness of frequent evidence conflicting problem of classic evidence theory, and the results of data fusion are more accurate than that of traditional monitoring sensors both in day and night time, especially at night time.

5. **Reliability test.** Classical Dempster-Shafer problem is on basis of the assumption that all the situations of the data sources are considered as equal importance. So we can get the correct result when the assumption is correct, on the contrary, when this assumption does not hold, it will cause erroneous results. Especially, the situation between the evidence is conflicting.

When the recognition framework is $\Theta = \{0, 1\}$,and the corresponding set $U = \{\{0\}, \{1\}, \{0 \text{ or } 1\}, \varphi\}$, two conflicting evidence and are defined as Table 1. (both m_P and m_Q are hypothetical values).

According to Eq.(4) ,as the reliability coefficient r_Q in Table 1 is adjusted from 4 to -5, the values of the *bpa* function and the *Bel* function are followed as shown in Table 2, depending on Eq.(5),(6),(7),(8). If $r_Q = r_P$,the value of $m(1)$ is 0.486, only half of m_P. However, as r_Q is adjusted from 4 to -1, the $m(1)$ rises to 0.731, and the $Bel(1)$ also rises from 0.514 to 0.817 obviously. When r_Q drops to -5, the data source Q is equivalent to being ignored, hence, the final decision is done by the *bpa* function of P.

6. **Performance evaluation of the proposed scheme.** This section discusses the effectiveness of the proposed scheme with respect to the energy consumption and data aggregation precision by showing the results of several simulations.

Table 1. Conflict of evidence and fusion results

m U	0	1	0 or 1	φ
m_P	0.1	0.8	0.1	0
m_Q	0.8	0.1	0.1	0
m	0.486	0.486	0.028	0

Table 2. Reliability of the fusion results calculated by adjusted Q in the case of conflict of evidence

credibility coefficient	weight coefficient	*bpa* function			*bel* fuction		
			1	0 or 1	0	1	0 or 1
$r_P = 4$, $r_Q = 4$	$w_P = 1$, $w_Q = 1$	0.486	0.486	0.028	0.514	0. 514	0.028
$r_P = 4$, $r_Q = 3$	$w_P = 1$, $w_Q = 0.88$	0.33	0.64	0.03	0.373	0.687	0.062
$r_P = 4$, $r_Q = -1$	$w_P = 1$, $w_Q = 0.44$	0.183	0.731	0.086	0.268	0.817	0.085
$r_P = 4$, $r_Q = -4$	$w_P = 1$, $w_Q = 0.11$	0.116	0.787	0.097	0.215	0.880	0.099
$r_P = 4$, $r_Q = -5$	$w_P = 1$ $w_Q = 0$	0.1	0.8	0.1	0.2	0.9	0.1

The chosen example is a traffic volume monitoring problem that we used NS2 to analyze the performance of the proposed scheme, and compared it with VLEACH in [5] and ESDA algorithms in [2]. The simulations were carried out considering a sensors field having a 1000m×800m rectangular shape, and assuming that 1 central node and several clusters are deployed in the field, where each cluster has 1 collector and 20 cluster members. Notice that the central node and collectors are fixed positions, but other nodes are random waypoints. The values of the other parameters used for the simulations are reported as follows.

The experiments are conducted on 200 rounds in total. In the proposed scheme, the cluster head is the fixed facility designated by the sink node, and most cluster members only need to send a message to complete data fusion in every sampling round. In addition, the proposed scheme adopts hybrid network structure to balance the energy consumption of the network. Hence, the energy consumption of cluster members makes significant improvement compared with VLEACH and ESDA algorithms. The results in the first simulation are reported in Figure 4. Early in the 120th round of the experiments, the average energy consumption of the mobile nodes exclude the central node and collectors reached to 0.7 J in VLEACH, while nearly 0.5J in case of ESDA and 0.3J in case of our scheme. When to the 180th round, VLEACH network has consumed to the extent of near paralysis, ESDA shows the similar situation, while our scheme just about 0.6J. Because the cluster heads and the sink node are all roadside infrastructures, the influence of the energy consumption problem on these nodes will be much smaller compared to the ordinary nodes.

Figure 5 shows the aggregation precision of these schemes suffering from the same network attack mentioned in [17] Because the VLEACH and ESDA do not solve the "intra-network attacking" problems, facing the same aggressive behavior in internal network, the proposed scheme depending on the D_τ and $Token'$ fields has much higher aggregation precision than VLEACH and ESDA. When $N = 200$, the aggregation precision of VLEACH is reduce to 80%, in ESDA is about 88%,while our scheme is still above 95%.

Table 3. Simulation parameters

Number of the nodes	8 Fixed nodes 100 Mobile nodes
Total simulation time	200s
sampling time	500ms
Area size	1000m × 800m
Packet size	1024 bytes
MAC protocol	IEEE 802.11b
Mobility model	Random waypoint
node speed	10km/h-40km/h
Link bandwidth	128kbps
Radio range	20 m
initial energy of mobile nodes	1.5J
initial energy of fixed nodes	20J
protocols	the propose scheme, VLEACH, ESDA

Figure 6 shows the aggregation precision comparison of these schemes along with the incensement of the untrusted behavior probability of source nodes. Aggregation accuracy tends to be effected by bad data and sensor errors, and, the distrusted behavior means that the data value transmitted from the source node is inconsistent with the real value. We can see from Figure 6, VLEACH protocol only checks out the properties of the output data, without considering the reliability evaluation on the properties of the node itself, ESDA is the same. On the other hand, the proposed scheme carries out data aggregation through the node credibility evaluation, which can still maintain a high aggregation precision even if the number of errors in sensor networks increases. For instance, when the distrusted behavior probability is 40%, the aggregation precision of the proposed scheme still remains 83%, while ESDA is 63%, and VLEACH is only 41%. This shows that the robustness of the proposed scheme is stronger than VLEACH and ESDA in the data aggregation process.

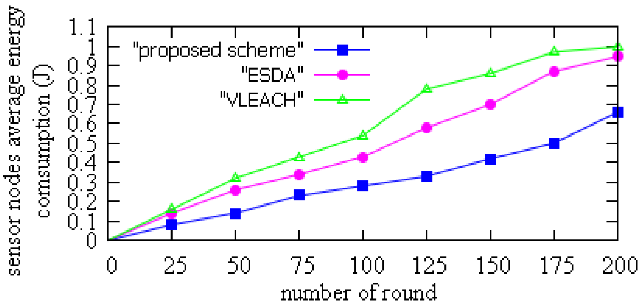

Figure 4. The energy consumption comparison of mobile nodes

7. **Conclusions.** Secure data aggregation in WSNs has been widely concerned by researchers at home and abroad. So far, how to balance safety and practicality in the data aggregation process has been one of the most critical issues of this research area.

This paper presents a dynamic data aggregation scheme based on the method of credibility estimation and reliability allocation, which is used for the data fusion and management of multi-source traffic information in ITSs. It adopts hybrid WSN based network structure, and combines the spatial and temporal characteristics of the traffic information. Furthermore the parameters i.e.D_τ, $Token'$ etc. joining in the message format ensures

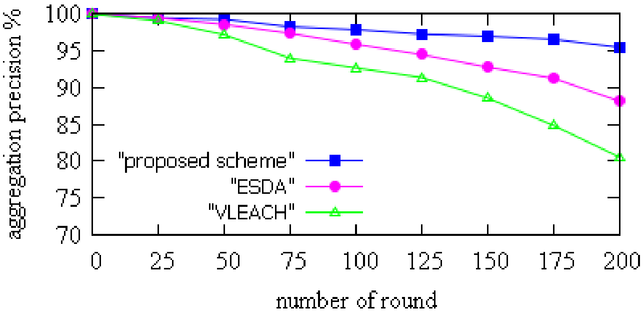

FIGURE 5. The aggregation precisions comparison (I)

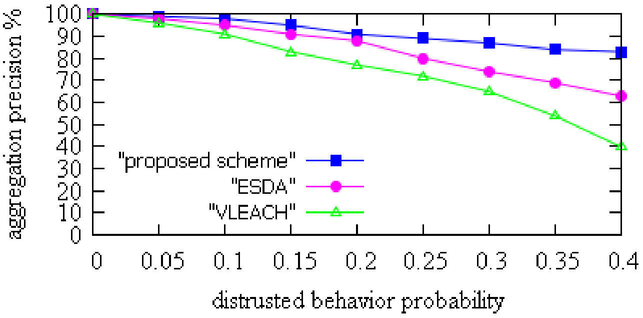

FIGURE 6. The aggregation precisions comparison (II)

that the data packets from the source nodes can be effectively integrated and transmitted. After receiving the aggregated messages from clusters, the application layer calculates and evaluates the aggregation results to generate decisions for traffic control. Simulation results show the proposed scheme can effectively reduce the energy consumption of the network compared with VLEACH and ESDA algorithms. In addition, it can still guarantee the authenticity of data aggregation in certain extent of distrusted behaviour attacks.

Therefore, this solution could be applied in the next generation of sensor technologies of ITSs In the near future, we will continue further study on the optimization of the routing based on the proposed scheme.

Acknowledgment. The work presented in this paper was partially financed by the National Natural Science Fund under grand 61462015 and in part by the Science and Technology Foundation of Guizhou Province No.J[2013]2222, and PHD Research Project Funding of Guizhou Normal University named “The control of network topology and secure information transmission in WSN based intelligent transportation system”.

REFERENCES

[1] C. C. Chung, L. C. Chao, S. J. Lou and Q. V. Nguyen, Benchmarking-based Analytic Network Process Model for Strategic Management, *Journal of Information Hiding and Multimedia Signal Processing,* vol. 6, no. 1, pp. 59-73, Jan. 2015.

[2] W. Luo, X. D. Hu, Efficient and secure data aggregation protocol for wireless sensor networks, *Journal of Chongqing University of Posts and Telecommunications*, Natural Science Edition, vol 21, no.1, pp.110-114, 2009.

[3] H. T. Mouftah, M. Khanafer, M. Guennoun, Wireless sensor network architectures for intelligent vehicular systems, *http://www.mcit.gov.sa/nr/rdonlyres/088 80e2f-f4c9-4029-988f-05bf1379516d/0/paper2.pdf*, accessed 4, Nov, 2011.

[4] M. B. Yassein, A. Al-zou'bi, Y. Khamayseh, W. Mardini. Improvement on LEACH Protocol of Wireless Sensor Network (VLEACH), *InternationalJournal of Digital Content Technology and its Applications*, vol 3,no.2,pp.132-136, 2009.

[5] Z. Y. You, X. L. Cao, Y. Wang, An Unequal clustering strategy for WSNs based Urban Intelligent Transportation System, *Journal of Information and Computational Science*, vol 12, no.10,pp.4001-4012,2015.

[6] T.C. Lin, H.C. Huang, B.Y. Liao, and J.S. Pan, An Optimized Approach on Applying Genetic Algorithm to Adaptive Cluster Validity Index, *International Journal of Computer Sciences and Engineering Systems*, vol. 1, no. 4, pp. 253-257, Oct. 2007.

[7] Y. M Wang, J. Li, Application of Dempster-Shafer theory for network selection in heterogeneous wireless networks, *The Journal of China Universities of Posts and Telecommunications*, vol. 19, no. 2, pp. 86–91, 2012.

[8] S. T. Jeng, Y.C.A. Tok, S. G.Ritchie, Freeway Corridor Performance Measurement Based on Vehicle Reidentification, *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 639 – 646, 2010.

[9] P. Ossenbruggen, E. Laflamme, Time Series Analysis and Models of Freeway Performance, *J. Transp. Eng.*, vol 138, no.8, pp.1030–1039, 2012.

[10] X. D. Jia, Y. F. Chang, C. C. Chang and L. -M. Wang, A Critique of a Lightweight Identity Authentication Protocol for Vehicular Networks, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 6, no. 2, pp. 183-188, Mar. 2015.

[11] C. Q. Wang, J. M. Chen, Y. X. Sun, Sensor network localization using kernel spectral regression, *Wireless Communications and Mobile Computing*,Vol 10,no.8, pp.1045-1054,2012.

[12] J. J. Gu, C. S. Can, Y. Zhuang, Wireless Sensor Network Based Topology Structures for the Internet of Things Localization", *CHINESE JOURNA L OF COMPUTERS*, vol. 33, no. 9, pp. 1548-1555,2012.

[13] H. Cam, S. Ozdemir, P. Nair, Muthuavina shiappan Energy-efficient secure pattern based data aggregation for wireless sensor networks, *Computer Communications*, vol 29, no.4, pp.446-455,2006.

[14] H. D. Wu, S. M. Stiefelhagen, R. J.Yang, Sensor Fusion Using Dempster-Shafer Theory, *IEEE Instrumentation and Measurement Technology Conference*, vol. 1, pp.7-12,2002.

[15] A. Ballenger, S. Gatepaille, Uncertainty in Ontologies: Dempster-Shafer Theory for Data Fusion Applications, *Workshop on Theory of Belief Functions*, Brest: France,2010.

[16] D. D.Xu, G. T. Chen, L. X. Tan, Y. J. Li,Routing Protocol Based on Non-uniform Clustering for Wireless Sensor Networks, *Journal of Hangzhou Dianzi University*, vol 32,no.3,pp.61-63,2012.

[17] X. Y. Zhang, X. T. Zhang, Y. Qi, L. W. He, Z. X. Xiahou, C. Y. Li, Security evaluation for wireless sensor networks based on attack test and fuzzy comprehensive judgement, *7 th CWSN*, pp.123-127,2013.

[18] S. Choi, D.Kim, D. Lee, J. Jung, WAP: wormhole attack prevention algorithm in mobile ad hoc networks, *IEEE International Conference on Sensor Networks*, Ubiquitous, and Trustworthy Computing, 2008.

[19] S. Khan, K. K.Loo, N. Mast, SRPM: Secure Routing Protocol for IEEE 802.11 Infrastructure Based Wireless Mesh Networks, *J Netw Syst Manage*, vol 18, pp. 190–209,2010.