Traffic State Estimation with Bayesian Networks at Extremely Low V2X Penetration Rates

Marek Junghans  
German Aerospace Center (DLR)  
Institute of Transportation Systems  
12489 Berlin, Germany  
Emails: marek.junghans@dlr.de

Andreas Leich  
German Aerospace Center (DLR)  
Institute of Transportation Systems  
12489 Berlin, Germany  
Email: andreas.leich@dlr.de

Abstract—In this paper the concept of Bayesian Networks (BN) is applied to the problem of traffic data acquisition by data fusion. Two wireless communication based sensors are used as data sources: IEEE 802.15.1 Bluetooth and IEEE 802.11p V2X (vehicle to vehicle and vehicle to infrastructure). Via V2X so called cooperative awareness messages (CAM) are received, which provide information on vehicle location and speed. For Bluetooth only the presence of a Bluetooth device can be detected. Currently and in the near future a low amount of road users is expected to be equipped with V2X. Therefore the rate of V2X vehicles is very low (≈1%). The penetration rate of Bluetooth devices is much higher. Approximately 5% to 50% of all road users can be detected and re-identified with a Bluetooth scanning device. Bluetooth detectors have notably been used for traffic management purposes for years, e.g. for obtaining journey times, but they have not been applied for speed estimation so far. The approach of this paper is to provide vehicle count data and vehicle speed by fusing Bluetooth data at moderate and V2X data at low penetration rates. The challenging task is to obtain accurate speed estimation data. By applying BNs for this purpose, we show the robustness of this stochastic fusion engine. It is capable of reaching speed RMSEs between 2 and 5 m/s and enhance the completeness of the state estimation by 35% by fusing 1% V2X with 30% Bluetooth. The investigations are made on the basis of simulation.

I. INTRODUCTION

Modern traffic light control (TLC) technologies require traffic state estimation, particularly speed and vehicle counts for switching traffic lights in a way that is optimal for traffic flow (actuated and adaptive TLC). Traditionally, this data is collected via inductive loop, video, radar, and other stationary detectors. The costs for purchasing, installation and maintenance of traditional detectors are high. Therefore many TLC algorithms in urban areas are still based on static historical data (fixed time control). Although urban areas with lower priority and less traffic demand are usually equipped with static TLC, ubiquitous actuated or adaptive TLC is desirable for improved traffic flow and reduced emissions. There is a growing research interest in alternative low cost traffic data acquisition solutions (cf. [3], [4]). One of the approaches is exploiting cooperative traffic data transmitted by V2X (vehicle to vehicle to infrastructure and vehicle to vehicle) communication in WiFi ad-hoc mode IEEE802.11p [8], which is now condensed to IEEE802.11-2012. Vehicles periodically send information about their GPS positions and velocities via Cooperative Awareness Messages (CAM). Those CAMs are received by Road Side Units (RSU). While this approach is attractive from the point of cost reduction, it is expected to take years to reach sufficient V2X penetration rates needed for actuated or adaptive TLC [3]. Therefore, it is a challenge to obtain speed data and vehicle counts of the vehicles approaching an intersection, particularly taking into account very low V2X penetration rates of around 1%.

In this paper, a novel approach is introduced that proposes to fill missing V2X data with data provided by a IEEE 802.15.1 Bluetooth device scanner. The penetration rate of Bluetooth devices is already high, because modern vehicles feature hands-free equipment, smartphones, navigation systems, etc. The classical Bluetooth occupancy detector has notably been tested for years for traffic data acquisition, e.g. for obtaining accurate journey times (e.g. in [12], [27]). Our approach adopts the Bluetooth occupancy detector as a speed and vehicle count detector. Depending on the road type and type of the traffic participant, the Bluetooth penetration rate differs from 3% to 50%, which depends on the application context, i.e. urban or suburban areas, motorways, the amount of trucks, etc. (cf. [2], [20], [27]). This results in higher detection rates in comparison to sparse V2X data. We applied Bayesian Networks (BN) to estimate the traffic state data at a very low V2X penetration rate of 1% and moderate Bluetooth penetration rate of 30%. Simulation based results show the robustness of this approach. It is capable of reaching speed RMSEs between 2 and 5 m/s and completing the traffic state estimation by 35% by fusing 1% V2X with 30% Bluetooth.

The paper is organized as follows: In section II the methodical approach and its requirements are described. In section III the concept of BN is applied to the problem of fusing the wireless communication technologies. In section IV the experimental setup and the obtained results are described. Finally, in section V conclusions and prospects of our future work are given.

II. METHODOICAL APPROACH

In this section the methodical approach and the requirements for speed and vehicle count estimation on the basis of Bluetooth occupancy and V2X data are described. They result in the implementation of the following sub-processes:

1. **Data Fusion Approach**: The goal is to fuse the data from Bluetooth and V2X sensors to estimate traffic state data at a very low V2X penetration rate of 1% and moderate Bluetooth penetration rate of 30%.

2. **Bayesian Networks (BN)**: BNs are applied to fuse the data from Bluetooth and V2X sensors. They are capable of reaching speed RMSEs between 2 and 5 m/s and enhancing the completeness of the state estimation by 35%.

3. **Implementation Setup**: The implementation setup includes the experimental setup and the obtained results. The setup is designed to test the robustness of the approach in various traffic scenarios.

4. **Future Work**: The conclusions and prospects of our future work are given, outlining the potential for further research and development in the field of traffic state estimation.
• Speed and vehicle count estimation
• Sensor data fusion of V2X and Bluetooth data
• Transport mode detection
• Direction filtering

The latter two sub-processes are beyond the scope of this paper and thus not considered.

A. Speed and Vehicle Count Estimation

RSUs for V2X communication receive the broadcasted speed information via CAMs. Due to the sparsity of V2X data, the number of V2X vehicles cannot reliably estimated. Bluetooth is an occupancy detector providing a device identification, e.g. MAC address, within a specific detection range. Because of the higher equipment rate, Bluetooth is a more appropriate means of vehicle count estimation and thus well complements V2X in this regard. Additionally, the speed of a Bluetooth device can be estimated indirectly. Both aspects are considered in this section.

1) Bluetooth Based Speed Estimation: The Bluetooth inquiry process (see [11]), which characterizes the handshaking procedure between Bluetooth sender and receiver, results in the exchange of the device IDs of the communication partners and therefore solely provides information on the presence of a device with given MAC address. Based on the known timestamps and the known detection range $r_{BT}$ of the sender (inquirer) it is possible to estimate the speed of a moving Bluetooth device $v_{BT}$ by equation (1) where $t_{First}$ is referred to as the first and $t_{Last}$ the last detection timestamp within $r_{BT}$.

$$v_{BT} \approx \frac{r_{BT}}{t_{Last} - t_{First}} \quad (1)$$

The number of Bluetooth detections of the same MAC depends on the speed of the Bluetooth device and the periodicity of the inquiry process, which is 2.56s. Therefore, $v_{BT}$ is more accurate the more frequent the same device is detected, i.e. the slower a Bluetooth device moves through $r_{BT}$. Vice-versa, the quicker it moves through $r_{BT}$, the larger is the estimated speed error. In general, the estimation of $v_{BT}$ is rather rough. Additional factors like signal propagation obstruction by the surrounding infrastructure and weather may influence the detection range, but are not considered here.

In this paper, we deal with the task of computing an optimal guess of $v_{BT}$ based on BN based fusion. We model different factors that influence $v_{BT}$. These factors in general include traffic parameters like traffic volume, desired, maximum ($v_{max}$) or current speed, speed difference ($\Delta v$) and time gap ($\Delta t$) to the preceding vehicle and physical weather conditions (heavy rain, ice and other phenomena). The initial guess of $v_{BT,1}$ can be written as a function of the mentioned factors:

$$v_{BT,1} = f(v_{max}, |\Delta v|, v_{BT}^{\prime}, \Delta t, \ldots) \quad (2)$$

In equation (2) $v_{BT}^{\prime}$ describes the speed of the preceding measured vehicle when it left the detection area $r_{BT}$. The speed difference of two successive Bluetooth devices $|\Delta v|$ is related to the influence of the traffic state on $v_{BT}$ since it is expected that $|\Delta v|$ and its variance will be different in heavily congested traffic compared to free flow. Further, the estimated speed $v_{BT}^{\prime}$ of a measured preceding Bluetooth device may be similar to the speed of the current Bluetooth device due to quasi-stationary conditions of traffic within a small time interval. Other influencing factors, e.g. intersection type, time of day and others, could be considered in (2), too. Some heuristics may help to obtain an acceptable estimation of $v_{BT,1}$, e.g.:

$$v_{BT,1}|_{t_{Last}=t_{First}} := v_{max} \quad (3)$$

$$v_{BT,1}|_{t_{Last}=t_{First}} := \nabla \quad (4)$$

If equation (3) is used, there will be big systematic speed errors for slow vehicles. Consequently, this equation is more appropriate for free flow at low traffic densities. Equation (4) may be suitable for synchronized and congested traffic, stop-and-go situations, and in case of traffic lights signalizing “red”. Whatever equation is chosen for the first guess, it is important to consider the surrounding traffic conditions adequately. Other influencing factors to be considered and handled throughout the computation of equation (2) are

- The transportation mode, e.g. pedestrians, cyclists, passengers of public transport. For instance a big amount of pedestrians “wearing” Bluetooth enabled smartphone devices could introduce a systematic error of $v_{BT}$. Therefore, it is recommended to implement a transport mode detection.
- It is necessary to distinguish between vehicles approaching and leaving the intersection to avoid systematic speed errors. Therefore, it is necessary to filter the direction of the detected Bluetooth devices.

In fig. 1 the idealized intersection setup with optimally placed V2X-RSU and 4 directed Bluetooth detectors is shown. According to [17] and [26] V2X-RSUs usually provide detection ranges from 100m up to 500m. In the Bluetooth case three different device classes exist providing different transmission/receiving power and communication ranges from 10m (Class 3 device) up to 100m (Class 1) [11]. We assume that all devices are of Class 2, which provide a detection range of 30 to 50m. The reason is that most traffic participants use Class 3 and Class 2 Bluetooth devices, e.g. smartphones and hands-free communication devices. In case the class of a Bluetooth device is known, equation (1) can be applied to obtain speed estimations for road users in each arm of the intersection.

2) Bluetooth Based Vehicle Count Estimation: As mentioned above, the detection of a moving Bluetooth device within a certain range of a Bluetooth receiver is influenced by the speed of the Bluetooth device and the functional principle of the Bluetooth inquiry procedure. Consequently, the longer the device is within $r_{BT}$, the higher is the probability of detection and vice-versa. We assume that every detection is assigned to exactly one traffic participant. Then, it is possible to estimate the number of vehicles $N$ passing through $r_{BT}$ on the basis of the number of detected Bluetooth devices $n_{BT}$ and a known Bluetooth penetration rate $p_{BT}$ in $[0; 1]$:

$$N(t) \approx \frac{n_{BT}(t)}{p_{BT}} \cdot c(t) \quad (5)$$
In case of low to moderate vehicle speeds and equally distributed Bluetooth devices in the road network the determination of \( N \) works satisfactory well if \( N \beta \geq 0.3 \). In equation (5) the parameter \( c \in R_+ \) is a correction factor reflecting the influence of different traffic conditions and the vehicle speed distribution at the intersection of interest.

**B. Sensor Data Fusion of V2X and Bluetooth Data**

As described in section II-A, the determination of vehicle counts and the Bluetooth based speed estimation are erroneous and incomplete. On the one hand there is accurate, but rather incomplete V2X data, at high time-resolution. On the other hand there is inaccurate Bluetooth data at low time-resolution. Therefore a fusion method is needed, which is capable of coping with uncertain data of V2X and Bluetooth. BNs are able to handle such data by the use of (conditional) probabilities. In this respect the traffic and measuring processes need to be modelled as cause-effect relationships and quantified by conditional probability density functions (CPDF). The application of BNs to the problem in question is presented in section III.

**III. BAYESIAN NETWORK BASED DATA FUSION**

Among others the concept of Bayesian Networks (BN) is capable of providing reliable and accurate data by inferring the results of the sensor measurements and combining them with a-priori knowledge [15]. To get a detailed view about BN, their computation and application areas, learning procedures, etc., the reader is referred to the literature [18], [22]. In the following BN for reliable and accurate speed estimation is developed. The creation of a BN for this purpose requires the consideration of both, the traffic and measuring process as random processes. They are modelled as nodes, whilst the cause-effect relationships among them are modelled as directed arcs, pointing from cause to effect. Afterwards, the nodes and their relationships need to be quantified by CPDFs representing the properties of the processes. Therefore it is necessary to know how the sensors determine the velocities of the Bluetooth devices and the V2X equipped vehicles taking into account the true velocities, which are called sensor likelihoods. Once having determined the sensor likelihoods for Bluetooth and V2X based detection and the a-priori probability density functions quantifying the statistical behaviour of the traffic process, the BN is capable of handling and combining the incomplete data of both sensors to estimate the speeds considering the properties of the sensors, the a-priori knowledge about the underlying traffic process and the measurement data. This simple BN is developed in section III-A. To consider other influencing factors, e.g. traffic conditions, the BN needs to be extended. An extended version of the BN is developed in III-B.

**A. Creating a Simple BN**

In fig. 2 a simple BN is shown that consists of the traffic process node \( V \) representing the instantaneous speed to be determined by the sensor nodes V2X (node \( V_{V2X} \)) and Bluetooth (node \( V_{BT} \)). The arcs point from \( V \) to \( V_{BT} \) and \( V_{V2X} \) modelling the causal relationships. The real values of the random processes are written in small letters, i.e. \( v_{BT} \in V_{BT} \) and \( v_{V2X} \in V_{V2X} \). The term \( P(v) \) quantifies the a-priori knowledge, i.e. the expected speed probability, and the terms \( P(v_{BT}|v) \) and \( P(v_{V2X}|v) \) are referred to as sensor likelihoods and quantify the sensor properties. In case there is additional knowledge about \( V \) we can estimate the sensors’ measurements probabilistically (causal evidences). This is important for learning the CPDFs. Usually, the learning process is done on the basis by expert knowledge or reference data of a more accurate sensor. In case there are data available at \( V_{BT} \) and/or \( V_{V2X} \) we are able to find out what happened at node \( V \) (diagnostic evidences) probabilistically. This procedure is called inference.

The joint probability distribution (JPD), which can be computed by the multiplication of the nodes CPDFs, characterizes the probabilities taking into account the measurements \( v_{BT} \) and \( v_{V2X} \):

\[
P(v, v_{BT}, v_{V2X}) = P(v) \cdot P(v_{BT}|v) \cdot P(v_{V2X}|v) \quad (6)
\]

Let \( \xi_{BT} \) be Bluetooth data within the Bluetooth detection area \( \Xi_{BT} \) and let \( \xi_{V2X} \) be V2X data within the V2X detection area \( \Xi_{V2X} \). Then, (6) is valid where \( \Xi_{BT} \) and \( \Xi_{V2X} \) overlap, i.e.:

\[
\xi_{BT}, \xi_{V2X} \in \Xi_{BT} \cap \Xi_{V2X} \quad (7)
\]

Evidences of the nodes \( V_{BT} \) and \( V_{V2X} \) are needed to compute the a-posteriori CPDF \( P(v|v_{BT}, v_{V2X}) \), which combines the measurements and considers the normalising constant \( \alpha^{-1} = \sum_i P(v_i|v_{BT}, v_{V2X}) = 1 \):

\[
P(v|v_{BT}, v_{V2X}) = \alpha \cdot P(v) \cdot P(v_{BT}|v) \cdot P(v_{V2X}|v) \quad (8)
\]
The most likely estimation of speed yields the speed for which maximum a-posteriori estimator (MAP) was applied, which estimators are described to get a detailed view. Here, the obtained by applying an adequate estimator. In [1], [10] different influences:

- due to the fact that speed estimation by a Bluetooth detector according to equation (1) is expected to be rather inaccurate, it is reasonable to consider vehicles that were detected several times within \( r_{BT} \) and vehicles that left \( r_{BT} \). It is assumed that a speed estimate at node \( V_{BT} \) can be improved by taking into account the mean instantaneous speed of the preceding Bluetooth device after leaving \( r_{BT} \) (node \( V_{BT}' \)). In a similar manner we determine the speed difference (node \( \Delta V \)) between the current and the measured preceding Bluetooth device. Therefore, arcs point from \( O, \Delta V \) and \( V_{BT}' \) to \( V_{BT} \).
- Different time gaps \( \Delta T \) between two vehicles are correlated with speed differences between the two considered vehicles. Therefore, an arc points from \( \Delta T \) to \( \Delta V \).

The BN in fig. 3 can be solved applying powerful inference mechanisms (cf. [18], [22]). However, due to its multi-connection the inference process can be time consuming, which might violate the requirements regarding real-time conditions of TLC. Therefore, the question is whether this BN can be simplified keeping it coherent without inducing systematic errors. The resulting BN should take into account different traffic conditions, e.g. traffic flow and density (macroscopic view), which can be modelled by the microscopic equivalent considering car following, i.e. by speed differences and time gaps between the leading and the following vehicles. Furthermore, we are interested how the Bluetooth based speed estimation is affected by the traffic conditions. Additionally, in the first step, we did not pay attention on the influence of TLC phases on \( V \), too. Node \( \Delta T \) could be neglected too, since considering the speed difference at node \( \Delta V \) is sufficient for the fusion process. Therefore, we neglected the nodes \( D, O \) and \( TLC \). In fig. 4 the result of simplification of the extended BN is shown. Analogous to section III-A the equations needed for processing and inferring the BN shown in fig. 4 taking into account two additional nodes to model the affection of node \( V_{BT} \). The resulting fusion equation for computing the a-posteriori CPDF considering the processes \( \Delta V \) and \( v_{BT}' \) with the normalising constant \( \alpha \) is:

\[
P(v|v_{BT}, v_{V2X}, \Delta v, v_{BT}') = \alpha \cdot P(v_{BT}'|v, \Delta v, v_{BT}') \cdot P(v_{V2X}|v) \\
\cdot P(v) \cdot P(\Delta v) \cdot P(v_{BT}')
\]
Applying the MAP-estimator on $P(v|v_{\text{BT}}, v_{\text{V2X}}, \Delta v, v'_{\text{BT}})$ we obtain the speed estimate $\hat{v}$:

$$\hat{v} = \arg \max_v P(v|v_{\text{BT}}, v_{\text{V2X}}, \Delta v, v'_{\text{BT}})$$ (13)

However, it may happen that the computation of equation (12) yields a zero vector, which occurs when the sensor likelihoods of the Bluetooth and the V2X sensor do not “overlap” and thus, $\hat{v}$ cannot be determined. This effect can be explained by learning node $\Delta v_{\text{BT}}$ and its affection by the $\Delta v'_{\text{BT}}$, which only considers the mean speeds of the preceding measured Bluetooth devices that had left $r_{\text{BT}}$. In case the Bluetooth device has not left $r_{\text{BT}}$ yet, no data might be available for the fusion process. If this happens, equations (14) and (10) need to be computed resulting in estimations $\hat{v}_{\text{BT}}$ and $\hat{v}_{\text{V2X}}$ separately, which need to be combined afterwards:

$$P(v|v_{\text{BT}}, v_{\text{V2X}}, \Delta v, v'_{\text{BT}}) = \alpha_1 \cdot P(v) \cdot P(\Delta v) \cdot P(v'_{\text{BT}})$$

$$\cdot P(v_{\text{BT}}|v, \Delta v, v'_{\text{BT}})$$ (14)

Here, a simple weighted mean value operator of both estimations (with the weight $w \in [0;1]$) was applied:

$$\hat{v} \approx w \cdot \hat{v}_{\text{BT}} + (1-w) \cdot \hat{v}_{\text{V2X}}$$ (15)

C. Learning the CPDFs

Reliable and accurate speed estimations according to the sections III-A and III-B are only possible if the sensor likelihoods $P(v_{\text{BT}}|v, \Delta v, v'_{\text{BT}})$ and $P(v_{\text{V2X}}|v)$ as well as the a-priori probability densities $P(v), P(\Delta v)$ and $P(v'_{\text{BT}})$ are quantified accurately. In this paper this was done by parameter learning algorithms, which are not discussed here (cf. [6], [18], [25]).

IV. EXPERIMENTAL RESULTS

The proposed method was implemented and tested using the microscopic traffic simulator SUMO (Simulation of Urban MOBility) [16]. In the following the SUMO traffic simulation setup (section IV-A), the traffic state estimation setup (section IV-B), the evaluation setup (section IV-C) and the simulation results (section IV-D) are described.

A. SUMO Traffic Simulation Setup

In fig. 5 the simulated scenario of a signalized intersection with 4 arms is shown, which relates to the idealized intersection in fig. 1. The boundaries of the 4 directed Bluetooth receivers are marked as green lines (entering $r_{\text{BT}}$) and red lines near the stopping line (leaving $r_{\text{BT}}$). For clarity, the detection range of the V2X-RSU is not shown here. A RSU capable of collecting V2X CAM messages of the vehicles was integrated into the RiLSA 1 scenario in SUMO [23]. The detection range was set to $r_{\text{V2X}} = 200m$. The Bluetooth detectors are standard class 2 receivers assuming a maximum detection range of $r_{\text{BT}} = 30m$. They implement the standard inquiry process [11] for the 4 arms of the intersection with directed antennas separately. This allows the RSU to distinguish between the intersection arms. In the simulation we retrieved vehicle counting and speed estimation in dependence on the different V2X penetration rates of $[1; 2; 5; 10; 20; 50; 100] \%$, whereas the Bluetooth penetration rate remained constant at 30%. The following parameters, constraints and assumptions were used in the simulation:

- Simulation time: 3,695.2 s or approx. one hour (36,953 simulation steps with a step size of 0.1 s)
- Traffic volume: $\approx 2,000$ veh/h approaching, non-uniformly distributed over the different intersection arms.
- The vehicles’ target speed is 13.9 m/s (50 km/h).
- The V2X and Bluetooth equipped vehicles are uniformly distributed in the simulation.
- Simulation runs: 400 (learning the a-priori probabilities and sensor likelihoods), 10 (data fusion)
- Determining vehicle counts: In our investigations we found adequate correction factors for vehicle count estimation of $c \in [0.9; 2.0]$. Therefore, for the northern, eastern, southern and western arms correction factors of $c_{\text{North}} = 1.21$, $c_{\text{East}} = 1.72$, $c_{\text{South}} = 1.0$ and $c_{\text{West}} = 1.74$ were applied.
- Basic detection probabilities $P(d)$: In case of Bluetooth the inquiry process was modelled probabilistically, i.e. the detection rate reaches 70% after 2.2 s. Additionally we assumed that a Bluetooth device cannot be detected twice or more within a full inquiry period of 2.56 s. In case of V2X we assumed that 90% of the CAMs were received by the V2X-RSUs.
- In case of Bluetooth, equation (3) was applied, i.e. the initial speed estimation per vehicle was set to $v_{\text{max}}$.

B. Traffic State Estimation Setup

The BN according to fig. 4 was created and its a-priori probability density functions as well as the sensor likelihoods were learned for integer velocities from 0 to 19 m/s and speed difference values from -20 to 20 m/s. The fusion process was only applied in the 30 m area where the detection ranges of V2X and Bluetooth overlap (equation (7)). In the following the results for the western intersection arm are presented only, while the remaining arms yield similar results.

1) A-priori Probabilities: In fig. 6 the learned a-priori probability $P(v)$ is shown for the western intersection arm.
There are mostly two peaks, one at low (0 m/s) and one at high speeds (14 m/s), which is due to the two dominant motions at different traffic states in this intersection scenario: (i) waiting due to traffic light phase red and (ii) free flow with the nominal speed of 13.9 m/s. In fig. 7 the learned a-priori probability $P(\Delta v)$ of the western arm is shown. There is a well-defined peak for $\Delta v = 0$ m/s yielding an almost symmetric CPDF. In fig. 8 the a-priori probability density $P(v_{BT}')$ is shown. It can be seen that the mean speed of the detected Bluetooth device, which entered, passed through and left the detection area has its maximum at about 2 m/s.

2) Sensor Likelihoods: The sensor likelihood of the Bluetooth detector, i.e. $P(v_{BT}|v,\Delta v, v_{BT}')$, cannot be easily visualized due to its high dimensionality. Therefore, the likelihood is processed to obtain the marginalized version of it, which is $P(v_{BT}|v)$. Both, $P(v_{BT}|v)$ for Bluetooth and $P(v_{V2X}|v)$ for V2X are shown in figures 9 and 10, the sensor CPDFs (y-axis) are plotted for different detected speeds (x-axis) given the true instantaneous speeds discretized as integers from 0 to 19 m/s, which is shown by the legends, e.g. BT 15 means the true speed of 15 m/s for Bluetooth. It can be stated that in case of V2X the speed estimation is as expected, i.e. the estimated values are spread around the true physical speed. The reason is, that the position and speed errors were modelled with normal distributions (not shown here). In case of Bluetooth the situation is more complex. The plot reflects that a Bluetooth detection is more likely to happen at low speed than at high speed of the Bluetooth device. But, particularly, in case of high speeds at 14 m/s we see that a Bluetooth detector provides frequent speed overestimation. This is the result of of applying
equation (3) for the initial detection of a Bluetooth device entering the detection area. Moderate and low speeds are estimated as expected according to the a-priori probability of the speed in fig. 6. Consequently, we expect that the Bluetooth detector frequently produces erroneous high and low speed measurements. Thanks to modelling the Bluetooth sensor likelihood by two additional nodes (see fig. 4) we can improve the speed estimation results of the Bluetooth detector.

C. Evaluation Setup
To evaluate the fusion of Bluetooth and V2X data for different V2X and constant Bluetooth penetration rates by the use of the BN modelled in 4 the following quality indicators were determined:

- Mean RMSE \( v_{RMS} \) of the estimated speeds compared with the ground truth. We use the RMSE as an indicator of accuracy of the fusion method.
- Vehicle count error \( \Delta N \) and the mean maximum vehicle count error \( \Delta N_{max} \) of the estimated vehicle counts compared with ground truth.
- Mean completeness \( q_c \) to indicate how many data is missing in the speed measurements.

D. Simulation Results
The fusion results for different V2X penetration rates are determined according to the requirements raised in the preceding sections IV-A, IV-B, and IV-C. In Table I the mean RMSE \( v_{RMS} \) as well as the completeness \( q_c \) in brackets for the speed estimation for different V2X penetration rates are shown. In case of the eastern and western arms \( v_{RMS} \) decreases as expected from 4.7 to 2.1 m/s and from 5.3 to 2.1 m/s, respectively. Due to the higher accuracy of the more frequent V2X speed estimations the maximum of the a-posteriori probability distribution is shifted to more accurate values since the contribution of the less accurate Bluetooth detector decreases. It should be noted that \( v_{RMS} \) is low for all penetration rates in the northern and southern intersection arms. The error even slightly increases when the penetration rate goes up from 1% to 20%. This is plausible, because vehicles statistically stick to the lower speeds, which is modelled by the a-priori probability. Therefore, independently on the data provided by the Bluetooth sensor as well as the V2X detector, the maximum of the a-posteriori probability distribution \( P(v|v_{BT}, \Delta v, v'_{BT}) \) provides low speed values, which is statistically true.

There is a slight increase of \( v_{RMS} \) at V2X penetration rates >2%, although the V2X detector is more accurate than the Bluetooth sensor. This also a notable effect, however the reason is yet unclear and needs to be investigated. At very high V2X penetration rates the massive presence of accurate V2X data enables the displacement of the maximum of \( P(v|v_{BT}, \Delta v, v'_{BT}) \) which, as expected, leads to an increase of the overall accuracy again.

The vehicle count error \( \Delta N \) is independent of the V2X penetration rate, since it is determined by the Bluetooth occupancy detection only. In our experiments we obtained the maximum vehicle count error for the intersection arms in the following intervals:

- Northern arm: \( \Delta N_{max} = [24; 33] \) vehicles or \([8.7; 11.7]\%\)
- Eastern arm: \( \Delta N_{max} = [20; 37] \) vehicles or \([3.0; 5.5]\%\)
- Southern arm: \( \Delta N_{max} = [7; 13] \) vehicles or \([2.5; 4.4]\%\)
- Western arm: \( \Delta N_{max} = [24; 45] \) vehicles or \([2.7; 4.9]\%\)

It can be stated that the maximum vehicle count error is about 12% for the northern arm, whereas for all other arms it is less than 6% – a satisfactory accuracy for traffic and transportation management purposes. In fig. 11 an example plot for the \( \Delta N \) over time on the western arm is shown. It can be seen how \( \Delta N \) fluctuates over time. In case of \( \Delta N > 0 \), there is a vehicle count overestimation, while for \( \Delta N < 0 \) the vehicle counts are underestimated.

V. CONCLUSIONS & FUTURE PROSPECTS
In this paper we presented a method for determining vehicle counts and speeds bases on Bayesian Network (BN) based data fusion. Sparse V2X speed and moderately dense Bluetooth occupancy data are used as input. It was shown that the method improves the overall accuracy and completeness of Bluetooth occupancy data are used as input. It was shown that the method improves the overall accuracy and completeness of speed estimation data. Although a Bluetooth detector is originally unable to provide speed estimation information of detected Bluetooth devices, our method enables speed estimation. We modelled the traffic and measuring processes as a BN and applied it for accurate and reliable speed estimation. We evaluated the performance of the algorithm by computing some quality indicators and obtained an RMSE for speed
estimation at a V2X penetration rate of only 1% between approximately 2 and 5 m/s. Even the worst results achieved are more than twice as good as throwing dices. Since the investigations were made on the basis of traffic simulations and other assumptions (e.g. constant penetration rate for Bluetooth of 30%, uniformly distributed V2X and Bluetooth technology, just one analyzed urban intersection, etc.) we expect the real world results worse than simulation based results. However, even at V2X penetration rates of 1% the method allows to achieve accurate and reliable speed data for traffic and transportation management purposes. We see a high potential for further investigation and promises to contribute to new ways for traffic detection and management. Thus, our future work will include the following:

- Improving the BN for data fusion including further ideas for obtaining a better and more accurate speed estimation in case of Bluetooth and an extensively analyzed V2X scenario. This includes the extension of the causal dependencies of the Bluetooth likelihood node, the consideration of steady-state conditions for traffic in case of specific free flow and congested flow conditions, the application of different time windows for speed estimation as a prerequisite to a certain Bluetooth detection, the modelling and quantification of different V2X conditions, e.g. multi-path effects, the variation of the Bluetooth penetration rate, and modeling and consideration of the traffic light control, to speed estimation
- Application of the method on other simulation scenarios with more realistic Bluetooth equipment rates and for a longer period of time, e.g. a whole week instead of 1 hour only
- Investigation of the traffic state and velocity dependency of the correction factor for vehicle count estimation
- Take into account other sensor technologies with higher detection ranges, but also rare penetration rates, e.g. WiFi
- Model the sensor CPDFs taking into account the affection of the sensors by external (weather, illumination, multipath effects) and internal (wear and tear, signal transmission) influences
- Aspects of timely queuing, multiple detections, noisy and biased data, etc. are to be taken into account, as for instance done in [13], [14].

ACKNOWLEDGMENT

The research leading to these results has received partial funding from the of the European Union’s Seventh Framework Programme FP7, FP7-ICT-2011-8 under REA grant agreement no. 318622. The authors acknowledge the support of the COLOMBO project partners.

REFERENCES


