Location Verification for Emerging Wireless Vehicular Networks

Ullah Ihsan, *Student Member, IEEE,* Shihao Yan, *Member, IEEE,* and Robert Malaney, *Senior Member, IEEE*

***Abstract*—The work reported here utilises the best aspects of information theory and deep-learning concepts so as to provide, for the first time, a solution for a real-world Location Verification System (LVS) in the context of vehicular networks. It is well established that Global Positioning System coordinates supplied by vehicles will be a vital component of such emerging networks. This supplied location information, if erroneous and not verified, can seriously degrade the overall system performance and lead to significant safety issues. A number of location verification protocols and systems have been developed to address this important problem but all have operational constraints and performance limitations due to their requirement for ideal static channel conditions and assumed threat models. In this work we remove such limitations by designing a Neural-Network based LVS (NN-LVS) that can accommodate *a priori* unknown channel conditions and unknown threat models. Under most channel conditions, the NN-LVS shows a performance improvement of 50%, or more, relative to other LVSs. We also derive a new information-theoretic bound on the Total Error for an LVS and show how this new bound allows for a useful trade-off in learning-time *vs*. verification-performance for the NN-LVS. We demonstrate an improved performance for the NN-LVS within the context of vehicular networks using Time of Arrival measurements of the vehicles' transmitted signals measured at multiple verifying base stations. The work reported here, we believe, paves the way to the actual deployment in real-world conditions of LVSs for emerging vehicular networks.**

***Index Terms*—ITS, VANET, Internet of Things, artificial intelligence, neural networks, location verification**

I. Introduction

Location information will play a pivotal role in many emerging networks - the Internet of Things, Trusted Autonomous Systems, and Vehicular Ad-hoc Networks (VANETs) being just a few. Indeed, it could be argued that location information on the physical entities that form such networks will form the most important knowledge attribute of the system. Because of this, the verification of such location information will play a crucial role in many aspects of network operations, with reliability and safety being of particular importance in the case of the emerging vehicular-network paradigm. The notion of an Intelligent Transportation System (ITS) has been a research topic in the wireless telecommunication industry for well over a decade now. VANETs are a specific type of ITS that enables vehicle-to-vehicle and vehicle-to-infrastructure communications [1]. Amongst other goals, VANETs aim to minimize traffic congestion, provide for greener cities, improve road-tolling infrastructure, advertise nearby facilities, and allow for seamless integration of new wireless entertainment services. However, much-improved safety outcomes are arguably the most hoped-for consequence of deployed VANETs. According to the World Health Organization, the number of fatalities worldwide due to traffic accidents were 1.25 million in 2013 [2]. Elimination (or drastic reduction) of such a catastrophic death toll is a challenge that is yet to be fully addressed. It is a challenge that motivates all future vehicular-network deployments.

Location information forms the basis of almost all network decisions in VANETs. Within most vehicular network architectures, a vehicle usually obtains its position from the Global Positioning System (GPS), perhaps with assistance from other components of the wider Global Navigation Satellite System (GNSS). The Wireless Access in Vehicular Environments (WAVE), specified in the IEEE 1609 family of standards [3], dictates that such location information is periodically reported to the wider network. A possibility exists whereby a malicious vehicle, in order to get an advantage over other vehicles, or to disrupt the system, or to trigger an unavoidable collision, can deceive the wider network by forging its location. Alternately, there is a possibility that a GPS receiver may wrongly calculate a vehicle's location due to poor reception or other non-malicious circumstances. Such events, if not carefully considered, can result in a host of undesirable outcomes.

Due to such considerations numerous algorithms for validating a vehicle's reported location in VANETs have been proposed over the past few years [4–10]. These algorithms, in general, make use of a network infrastructure that comprises Road Side Units (RSUs) and vehicles. RSUs are static Base Stations (BSs) placed at optimum locations to assist VANETs with a range of communication objectives. The inter-vehicle, RSU-to-vehicle and vehicle-to-RSU communication data is processed by an RSU in a specified coverage area so at best facilitate network operations. However, in the context of location verification, these same signals can be utilized to verify all reported GPS positions.

Researchers have formulated a number of information theoretic Location Verification Systems (LVSs) over the past few years [6] which utilize available signal metrics, such as the Received Signal Strength (RSS) of a transmitted signal from a vehicle, the Time of Arrival (ToA) of the transmitted signal from the vehicle and/or the Angle of Arrival (AoA) of the transmitted signal from the vehicle, to validate the vehicle's reported location. At the same time these LVSs have certain operation limitations in terms of the channel conditions and

U. Ihsan and R. Malaney are with the School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney, NSW 2052, Australia (e-mails: {ihsanullah, r. malaney}@unsw.edu.au).

S. Yan is with the School of Engineering, Macquarie University, Sydney, NSW 2109, Australia (e-mail: shihao.yan@mq.edu.au).

This work was partially supported by the Australian Government's Research Training Program. The corresponding author of this article is Shihao Yan.

can only work under the assumptions made at the time of their initial design. Recently, Artificial Intelligence (AI) has transformed many aspects of modern society through the use of its specialised algorithms. Enhanced clinical diagnosis [11], drug discovery [12], computer vision [13], speech recognition [14], efficient navigation [15], are a few of the areas impacted. AI has also laid the foundation for autonomous vehicles [16]. The integration of a neural-network (a sub-branch of AI) into LVS is crucial. Once trained to a certain limit, a Neural-Network based LVS (NN-LVS) can perform well in the channel conditions they are designed for. Due to the their adaptability, the NN-LVS can accommodate for changing channel and environment conditions, something which the ordinary LVSs lack. Further, the NN-LVS is believed to address multiple threat situations alone, thus eliminating the need for designing threat specific LVSs [17]. The NN-LVS also has the potential to update itself through continuous learning thus avoiding the risk of becoming obsolete.

The novel contributions in this work can be described thus.

- We introduce for the first the time the concept of location verification within vehicular networks using NN techniques.
- The NN schemes we develop deploy new techniques based on optimal-decision-theory frameworks that significantly reduce the training phase with only a minimal impact on performance.
- We study threat model under changing distance constraints for the malicious vehicle in an environment that closely relates to the real-world conditions and show how our newly designed NN-LVS beats the performance of an information theoretic LVS.
- We investigate in detail a specific example of these conceptual NN frameworks by determining a new theoretical lower bound on the optimal information-theoretic performance of location verification within the context of Non-Line-of-Sight (NLoS) effects as described by an exponential distribution of bias.
- By using this new lower bound we quantify the trade-off in NN training time *vs.* performance, showing how traditional training times (using traditional stopping criteria) can be cut substantially while impacting on the Total Error only marginally.
- We numerically investigate the performance of our new NN-LVS under a wide range of generalised conditions in which the incoming data is substantially different from the test data used to train the NN-LVS.

In this work we first review LVS deployment issues. We then briefly review techniques that have been proposed to address the location verification problem within VANETs; discussing a range of location verification techniques that are more of a heuristic nature, before bringing our attention to formal techniques based on optimal-decision theories. We then briefly introduce the use of AI in the guise of NN architectures and discuss their usefulness for location verification. We then study the performance of our newly designed NN-LVS against a state-of-the-art information theoretic LVS when the threat model for the malicious user is randomly changing. We then derive new information-theoretic bound on the performance of an LVS in the context of biased timings in realistic channel conditions before quantifying how the interplay between information theory and AI leads to improved location verification outcomes.

II. Related Works

A. Location Verification Deployment

In a configuration of a vehicular network where a LVS is deployed, several of the vehicles will have been already authenticated, but one vehicle (e.g., Vehicle A), has not. Vehicle A sends a request to join the network and reports its GPS position through its nearest RSU to the LVS server. All RSUs and nearby vehicles hearing this request report any relevant signal metrics (e.g. RSS) back to the server. The LVS server acts on this information, processing it to arrive at a verification decision. This decision is then reported back to all other vehicles in the network (or at least those in the vicinity of the requesting vehicle). In the case where the reported GPS position is not verified, the requesting vehicle has its security certificates revoked [3]. This communications system upon which this process is deployed over is quite generic and versatile. Dedicated Short Range Communications (DSRC), and the larger WAVE suite of protocols can easily accommodate the above configuration [3]. Vehicular network communications built on emerging 5G standards and beyond could also be easily accommodated.

B. Heuristic Algorithms

Most of the works prior to 2012 that have addressed the location verification issue within wireless networks have been extensively reviewed elsewhere [6]. The most well-known of these early proposals are perhaps those of [18–23].

The 'ECHO' protocol, based on the use of both radio and ultrasound frequencies, was developed in [18]. The main advantage of this protocol is that it is largely independent of the tight synchronization (timing) requirements demanded by other schemes. In this protocol, the verifier sends a packet with an unknown random value to the claimant. The claimant then instantly responds, via ultrasound, upon receiving the packet. The verifier then compares the round-trip time against an ideal delay time. If the time-of-flight is less (more) than the ideal time, the claimant is marked to be inside (outside) the region.

A 'verifiable multilateration' scheme was proposed in [19]. The scheme is based on the notion of 'distance bounding' (a malicious node can only claim he is further from the verifier not closer) and makes use of a minimum of three verifiers. The scheme can be shown to produce verifiable regions of space under a series of different threat models.

In [20], a series of autonomous sensors were adopted which looked at a range of metrics, such as the anticipated range of communication, the anticipated maximum speed of vehicles, and the anticipated density of vehicles. Using techniques akin to malicious behavior detection in intrusion detection algorithms, a location verification outcome was derived.

A time stamp embedded within the packet that contains the claimed location was used to detect malicious vehicles in [21]. The time stamp check ensures that the received packet is neither in the appropriate time window.

In [22], the notion of covert BSs (BS whose locations are assumed to be unknown to the attacker) were introduced into the location verification problem. In conjunction with the use of secret keys, the use of covert BSs was shown to significantly improve the verification performance.

Finally, in [23] an onboard radar system was used to verify the vehicle's claimed location. Taking noise into account, the scheme detailed in [23] separately determines the GPS position tolerance shadow and radar position tolerance shadow - with the algorithm accepting (rejecting) the prover's claimed location if there is (is not) an intersection between the GPS and radar position shadows.

More recent works on the issue of location verification in VANETs have appeared post 2012 [4, 5, 24–29]. In [4], a tile-based system was used to construct aggregated 'belief' type algorithms that include the use of nearby vehicles as a means to influence the belief value. To reduce the impact of erroneous location information on future location prediction for an optimum geographical routing, the authors in [5] proposed a protocol named Location Error Resilient Geographical Routing (LER-GR). The protocol of [5] utilizes the errors in location information on nearby vehicles in choosing the next forwarding vehicle to further improve the geographical routing and overall network performance. In [24], the location verification problem was extended to a larger cooperative framework by using other vehicles. In [25], the location of the vehicle was verified conditioned on a guarantee of privacy to the user. In [26], the integration of location verification with a popular alert-message algorithm was investigated. The use of token-based registration models and the use of multiple (and independent) authentication rounds were discussed in [27]. Power-optimization issues, in the context of location verification, were examined in [28], while in [29] the use of vehicle trajectories in aiding the location verification was studied.

C. Optimal-Decision Theory for Location Verification

The major difference of an LVS compared to a positioning system is that an LVS provides a binary output (either a 'True' or a 'False' value) as an output decision regarding the location of a user, while a positioning system outputs an estimated position. The generic operational procedures of an LVS involve inputs that include the 'claimed location' of a user along with independent signal inputs such as RSS, ToA, and/or AoA. After processing these inputs, the LVS labels the user's reported (claimed) position to be either true or false. We note that an LVS solves a different problem than the determination of the location of a node in a secure fashion. In an LVS, we are focussed on authenticating the validity of a claimed location reported by a node, and not on a node self-locating itself in a secure manner.

We consider the following scenario for a general LVS. A claimed location for a vehicle (usually a GPS position reported by that vehicle) is provided as input to the LVS processor along with the channel metrics (e.g., RSS, ToA, or AoA). The assumed channel conditions between the verifier and claimant vehicles will form part of the integrated system model adopted by the LVS. In general, it is assumed that the verifiers' locations are known and trustworthy. It could be that the verifiers are RSUs or other vehicles that have been previously authenticated. The LVS processor acts on these inputs with some pre-determined algorithm so as to produce a binary decision regarding the validity of the claimed location reported by the vehicle. A 'true' means that the reported location is legitimate or trustworthy, while a 'false' means that the reported location has not passed the verification test and the corresponding vehicle is likely spoofing its location (or its on board GPS is unreliable). This information can then be distributed to the wider network and it can be left up to individual vehicles, or the network system as a whole, on how to act on that information. In most situations a vehicle failing the location verification test will have its security certificates revoked [3]. The nature of the specific algorithm deployed within the LVS can be wide and varied, ranging from simple heuristic algorithms to those using optimal information-theoretic techniques, which will be discussed in the following.

Several location verification frameworks for VANETs, using optimal-decision theories, were formulated in a series of papers [7–10]. In the first of these works, the mutual information between the input and output LVS data was used as the objective optimization criterion. It was proven how a Likelihood Ratio Test (LRT), constructed from the probability of receiving a specific RSS value under different binary hypothesis (the claimant is truthful or untruthful), leads to an optimal decision [7]. The work was extended in [8] to correlated-shadowing environments where a more general optimization criterion based on the Total Error was utilized. The Total Error is a combination of the fraction of false positives (friendly vehicle stating a true location marked as malicious) and the fraction of missed detections (malicious user spoofing a location marked as friendly). The use of optimal decision theory based on Total Error in the context of Rician channels (the most likely form of channel in many vehicular environments) was reported in [9]. In [10], the use of directional antennas within the context of optimal decision theory and location verification for vehicular networks was reported, showing how the use of such antennas can increase verification performance.

D. Inclusion of Artificial Intelligence

The LVS processing discussed above will provide for the optimal decision conditioned on one important assumption - that the wider system model adopted is reality. However, in practice the adopted system model can only be considered an approximation to reality. This is because inferred channel conditions, noise parameters, receiver characteristics, position-error assumptions, and other system descriptors are very difficult to *a priori* determine. The issue is further complicated by the fact that the system parameters are likely to possess both spatial and temporal characteristics. To accommodate such real-world issues, a more sophisticated solution beyond optimal-decision frameworks will be desired. This additional ingredient, we believe, will be AI or machine-learning.

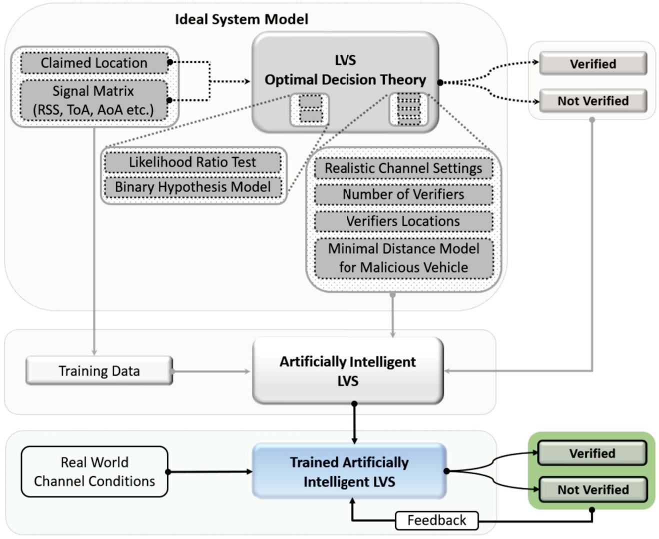
Ideal System Model
Claimed Location
Signal Matrix (RSS, ToA, AoA etc.)
LVS Optimal Decision Theory
Verified
Not Verified
Likelihood Ratio Test
Binary Hypothesis Model
Realistic Channel Settings
Number of Verifiers
Verifiers Locations
Minimal Distance Model for Malicious Vehicle
Training Data
Artificially Intelligent LVS
Real World Channel Conditions
Trained Artificially Intelligent LVS
Verified
Not Verified
Feedback

Fig. 1. Inclusion of Artificial Intelligence (AI) into the Location Verification System (LVS).

AI within the context of vehicular networks has been discussed previously. It is considered that AI will play an important role in many characteristics of the vehicular networks, all the way through from vehicle classification problems, vehicle vision systems, autonomous driving, big-data analytics, routing, vehicle security, and driver behaviour, to name just a few (see [30] and references therein for an overview). AI has recently been used for the theoretical simulation of an in-region LVS [31]. We believe the performance for the AI based LVSs in the context of IoT can be improved further by considering faster, extra reliable, and secure modern-communication frameworks [32–37]. These new communication frameworks will secure and strengthen the transmission backbone for the AI based LVSs.

It is part of our own ongoing work in this area to seamlessly integrate optimal decision LVS frameworks within AI architecture for location verification. We believe the combination of these techniques will provide for a 'best of both worlds' approaches to this important problem. The optimal decision theory will be used as part of the initial training set for the machine learning, and will be utilised in circumstances when the AI algorithm has not been trained for the specific locale being investigated. Once trained under ideal conditions, the deep-learning algorithms within the architecture will then be adjusted under real-world conditions when enough authenticated vehicles can be utilised as part of the training set.

III. System Model

A. The New LVS

A schematic of our combined AI architecture is given in Fig. 1. Here we see how the optimal decision theory outputs can be used as part of the initial training phase, as well as the ongoing training phase (as more vehicles are authenticated or marked malicious). The mutual information between the outputs of the optimal-decision theory and the AI can be used as an additional input to the AI deep-learning algorithms, as well as providing a measure of 'agreement' between the two processes. Confidence levels associated with this measure could further assist the LVS.

B. Adopted System Model Assumptions

Building on the work of [38] we assume the following.

1) A *single* vehicle to be identified (henceforth referred to as the user-vehicle) reports its claimed location, $\boldsymbol{\theta}_c = [x_c, y_c]$, to a network with N verifiers. Verifiers are fixed BSs with known locations. The location of the i-th BS is $\boldsymbol{\theta}_i = [x_i, y_i]$ $(i = 1, 2, \ldots, N)$. We select one of the N BSs as the Processing Center (PC) - which collects all measurements.
2) A user-vehicle (legitimate or malicious) obtains its true position, $\boldsymbol{\theta}_t = [x_t, y_t]$, from GPS with zero localization error. A legitimate user-vehicle's claimed (reported) position, $\boldsymbol{\theta}_c$, is taken to be identical to its true position $\boldsymbol{\theta}_t$. A malicious user-vehicle will falsify (spoof) its claimed position in an attempt to fool the LVS. The malicious user-vehicle's true location is unknown and its (spoofed) claimed location is taken to be $\boldsymbol{\theta}_c$.
3) We denote the null hypothesis where the user-vehicle is legitimate as $\mathcal{H}_0$, and denote the alternative hypothesis where the user-vehicle is malicious as $\mathcal{H}_1$. This can be summarized as,

$$\begin{cases} \mathcal{H}_0 : & \boldsymbol{\theta_c} = \boldsymbol{\theta_t} \\ \mathcal{H}_1 : & \boldsymbol{\theta_c} \neq \boldsymbol{\theta_t}. \end{cases} \tag{1}$$

C. Observation Model under $\mathcal{H}_0$

The ToA measured at the i-th BS from a legitimate user-vehicle, Y_i, is given by

$$Y_i = U_i + X_i + \phi_i, \quad i = 1, 2, \ldots, N, \tag{2}$$

where

- $U_i = d_i^c / c$, with d_i^c as the Euclidean distance from the i-th BS to a user-vehicle's true location (also its claimed location) given by

$$d_i^c = \sqrt{(x_c - x_i)^2 + (y_c - y_i)^2},$$

  and c as the speed of light,
- X_i is a zero-mean normal random variable with variance σ_T^2, which represents the additive noise term in the ToA measurements, and
- ϕ_i is a random variable (in ns) that represents the bias caused by NLoS channel conditions, that follows an exponential distribution with ρ_i as the scale parameter, *i.e.*,

$$f(\phi_i) = \rho_i e^{-\rho_i \phi_i}. \tag{3}$$

In the first instance we adopt the exponential distribution to model the bias term in the ToA measurements, since this bias should be always positive due to the NLoS channel conditions.

We assume that the observations collected by different BSs are independent and thus the likelihood function of the N-dimensional observation vector $\mathbf{Y}$ under $\mathcal{H}_0$ is given by

$$f\left(\mathbf{Y}|\mathcal{H}_0\right) = \prod_{i=1}^{N} f\left(Y_i|\mathcal{H}_0\right), \tag{4}$$

where $f\left(Y_i|\mathcal{H}_0\right)$ is the likelihood function of each Y_i under $\mathcal{H}_0$. We will derive the explicit expression of $f\left(Y_i|\mathcal{H}_0\right)$ later to facilitate the determination of the likelihood function $f\left(\mathbf{Y}|\mathcal{H}_0\right)$.

D. Observation Model under $\mathcal{H}_1$

It is reasonable to assume that the malicious user-vehicle's true location is not close to its claimed location when an attack occurs. Note also, the malicious user-vehicle can alter some system parameters to interfere with the observations collected by all BSs. As such, the ToA measured at the i-th BS, Y_i, is given by

$$Y_i = T_x + W_i + X_i + \phi_i, \quad i = 1, 2, \ldots, N, \tag{5}$$

where

- T_x is the time bias utilised by the malicious user-vehicle, and
- $W_i = d_i^t/c$, with d_i^t as the Euclidean distance from i-th BS to a user-vehicle's true location given by

$$d_i^t = \sqrt{(x_t - x_i)^2 + (y_t - y_i)^2}.$$

Since all BSs can communicate with each other, T_x is a constant value as seen by all BSs. Again, assuming that the observations collected from different BSs are independent, the likelihood function of the N-dimensional observation vector $\mathbf{Y}$ under $\mathcal{H}_1$ is given by

$$f\left(\mathbf{Y}|\mathcal{H}_1\right) = \prod_{i=1}^{N} f\left(Y_i|\mathcal{H}_1\right), \tag{6}$$

where $f\left(Y_i|\mathcal{H}_1\right)$ is the likelihood function of each Y_i under $\mathcal{H}_1$. In general, the malicious user-vehicle will optimally set the value of T_x in some sense in order to minimize the probability to be detected. We denote the optimal value of T_x as T_x^*. We will assume in the calculations to follow that the malicious user-vehicle optimizes T_x through minimizing the KL-divergence between $p(\mathbf{Y}|\mathcal{H}_0)$ and $p(\mathbf{Y}|\mathcal{H}_1)$, where $p(\mathbf{Y}|\mathcal{H}_0)$ and $p(\mathbf{Y}|\mathcal{H}_1)$ are the likelihood functions of observations under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. This KL-divergence is defined as

$$\begin{aligned} D_{KL}(p(\mathbf{Y}|\mathcal{H}_0)||p(\mathbf{Y}|\mathcal{H}_1)) &= \int p(\mathbf{Y}|\mathcal{H}_0) \ln \frac{p(\mathbf{Y}|\mathcal{H}_0)}{p(\mathbf{Y}|\mathcal{H}_1)} d\mathbf{Y} \\ &= \sum_{i=1}^{K} \frac{\left(U_i - W_i - T_x\right)^2}{2\sigma_T^2}. \end{aligned} \tag{7}$$

Then, the optimal value of T_x can be obtained through

$$\begin{aligned} T_x^* &= \arg\min_{P_x} D_{KL}\left(p(\boldsymbol{m}|\mathcal{H}_0||p(\boldsymbol{m}|\theta_t, \mathcal{H}_1)\right) \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(U_i - W_i\right), \end{aligned} \tag{8}$$

where $\boldsymbol{m}$ is the mean measurement vector for ToA of the signal under the relevant hypothesis. The value of T_x adopted by the malicious user-vehicle is not known to an LVS, and therefore the examination based on T_x^* is applicable to the worst scenario for an LVS (best scenario for the malicious user-vehicle).

E. Likelihood Functions and Performance Limits

In this section, we first formalize the decision rule embedded in a ToA-based LVS. In order to analyze the performance of this LVS, we also derive the likelihood functions, based on which we then discuss the strategy of the malicious user-vehicle to optimally set T_x. Finally, the performance limit of the LVS is examined.

F. Binary Decision Rule

In a localization system, the output is normally the estimated location of a user-vehicle. This is quite different from an LVS where the output is usually a binary yes/no decision. In this work we adopt the LRT as the decision rule embedded in our LVS. It has previously been proved that the LRT can achieve the minimum Total Error, which is the sum of the false positive rate and detection rate (we will assume in this work that the *a priori* probability of a user-vehicle acting maliciously is 0.5) [39]. The decision rule based on the likelihood ratio is

$$\Lambda\left(\mathbf{Y}\right) \triangleq \frac{f\left(\mathbf{Y}|\mathcal{H}_1\right)}{f\left(\mathbf{Y}|\mathcal{H}_0\right)} \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \lambda, \tag{9}$$

where $\Lambda\left(\mathbf{Y}\right)$ is the likelihood ratio, $\mathcal{D}_0$ and $\mathcal{D}_1$ are the binary decisions that infer whether the prover is legitimate or malicious, respectively, and λ is the LRT decision threshold. We define a simplified Bayes average cost function to quantify the performance of the LVS in terms of Total Error as below

$$\xi = p(\mathcal{H}_0)p(\mathcal{D}_1|\mathcal{H}_0) + p(\mathcal{H}_1)p(\mathcal{D}_0|\mathcal{H}_1), \tag{10}$$

where $p(\mathcal{H}_0)$ and $p(\mathcal{H}_1)$ are the *a priori* probabilities of occurrences of $\mathcal{H}_0$ (genuine user-vehicle), and $\mathcal{H}_1$ (malicious user-vehicle), respectively. We denote $\alpha = p(\mathcal{D}_1|\mathcal{H}_0)$ as the false positive rate and $\beta = p(\mathcal{D}_1|\mathcal{H}_1)$ as the detection rate and consider equal *a priori* probabilities for $p(\mathcal{H}_0)$ and $p(\mathcal{H}_1)$. Eq. (10) takes the form

$$\xi = 0.5\alpha + 0.5(1 - \beta). \tag{11}$$

We note that the decision rule given in (9) can achieve the minimum value of ξ, which is denoted by ξ^*.

IV. Performance Examination on Location Verification Systems

A. NN-LVS System Architecture

Based on a series of trials with changing architectures for the NN-LVS, we have designed a NN-LVS that has an input layer, a hidden layer with 10 neurons, and an output layer with a binary output (true or false) that decides on the integrity of a user-vehicle's reported location. This NN architecture provides better accuracy and robustness. The inputs to the

NN-LVS are the same as those considered for the information theoretic LVS, *i.e.*, ToA of the transmitted signal and the claimed location from a user-vehicle.

The activation function for a neuron in the hidden layer is given by

$$a_{(h,n)} = b + w_{(h-1,1)}a_{(h-1,1)} + \cdots + w_{(h-1,j)}a_{(h-1,j)}, \qquad n, j = 1, 2, \ldots, N,$$

where b is a constant, a is the activation for a neuron in a layer ranging from 1 to n, h represents the hidden layer and w is the weight connecting a neuron in the input layer to the neuron in the hidden layer.

Based on numerous rounds of simulation with changing transfer functions and backpropagation algorithms for the NN-LVS, we decided to utilize the hyperbolic tangent sigmoid transfer function in the hidden layer and the Levenberg Marquardt as the backpropagation algorithm.

We supply the NN-LVS with training data[1] at a speed of one random user-vehicle ToA data per second. Once the ToA data is supplied, the NN-LVS (via the backpropagation algorithm) optimises its weights and biases through a process called $learning$. The most frequent condition that terminates the learning process is the backpropagation algorithm's gradient descent. The gradient descent refers to minimizing the cost function for the NN-LVS through optimization of the weights as given below

$$w_i \leftarrow w_i - \gamma \frac{\partial}{\partial w_i} J(w_0, w_1 \ldots w_j), \quad i = 1 \ldots j,$$

where γ (a dimensionless constant) is the learning rate and J is the cost function for the NN-LVS. 'J' is the mean square difference between the calculated output for the training data, and the ground truth available with the training data.

$$J(w_0, w_1 \ldots w_j) = \frac{1}{2n} \sum_{i=1}^{n} (\hat{y} - y)^2,$$

where n is the number of training examples in the training data, $\hat{y}$ is the calculated output for the training data, and y is the ground truth available with the training data. The learning is terminated when $J_k \geq J_{k-1}$ over a set number of multiple iterations in a row (In our case this value is set to 6). 'k' refers to the iteration number. Once the learning has concluded, the weights and biases for the NN-LVS are considered as tuned. The NN-LVS thereafter can be applied to the test data[2] to calculate binary outputs for classifying the user-vehicles.

B. Effects of Bias and Assumed Threat Model

We now compare the NN-LVS's performance with the performance of the information theoretic framework formulated in [38]. Our modelling of the channel for the malicious user-vehicle attempts to take into account the uncertainty implicit in any real-world channel for a malicious user-vehicle a distance r from a claimed location. In reality, the BS-user-vehicle channel for any user-vehicle (a malicious or otherwise) is complex, dynamic (principally due to user-vehicle motions) and unknown. No accurate analytical theoretical model exists for such a real-world scenario. However, the key advantage of a NN-LVS over any purely information-theoretic LVS, is that the former can always 'learn' the channel models for both the legitimate and the malicious user-vehicles. This is true for all real-world channels irrespective of their nature (other than a purely random channel).

To make progress, and to test our NN-LVS numerically, we will model a specific real-world channel (roughly considered as a stochastic combination of bias and Rayleigh fading) for the malicious user-vehicle as follows. It is implemented in the following manner. Each measurement by the BS is probabilistically determined as either a pure bias term or a pure scattering term (*i.e.*, no LoS component) with the weighting of each term exponentially weighted with the distance r through e^{-ar} where a is set to $\frac{1}{x}$, x being the distance we expect the malicious channel to be pure bias with probability $1/e$. Random Gaussian noise is then added to the measurement. This channel model mimics the fact that as the malicious user-vehicle approaches the claimed location its channel model should approach that of a legitimate user-vehicle, and as it moves further from the claimed location the model provides for equal timing measurements, moduli noise. The legitimate user-vehicle channel is always modelled as a pure bias channel. As we will see in our results the NN-LVS does indeed learn these 'real-world channels' - an outcome that persists independent of the details of the actual malicious/legitimate channel model adopted.

We now compare the performance of the NN-LVS with a state-of-the-art information-theoretic LVS [38]. This information-theoretic LVS calculates its binary decision by taking into account the ToA of the transmitted signal from the user-vehicle via the decision rule

$$\Lambda\left(\boldsymbol{Y}\right) = \frac{e^{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{V})^{\mathrm{T}}\ \boldsymbol{R}^{-1}\ (\boldsymbol{Y}-\boldsymbol{V})}}{e^{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{U})^{\mathrm{T}}\ \boldsymbol{R}^{-1}\ (\boldsymbol{Y}-\boldsymbol{U})}} \quad \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \quad \lambda. \qquad (12)$$

In our comparison study shown in Fig. 2, we have four BSs in a 1000m X 500m area. The standard deviation for X_i and NLoS bias have been fixed at 100ns, and 300ns, respectively. The resultant Total Error calculated for the information-theoretic LVS based on the LRT in equation (12) is 0.14 for r = 150m, 0.10 for r = 200m, 0.07 for r = 300m, 0.06 for r = 1000m and 0.07 for r = 10000m. In comparison, the Total Error for the NN-LVS is 0.19, 0.14, 0.10, 0.04 and 0.01 for r equal to 150m, 200m, 300m, 1000m, and 10000m, respectively. Our comparison study shows that contrary to the information theoretic LVS of [38], the NN-LVS performs better and its Total Error improves with increasing r. Our results support our claim that the NN-LVS is efficient at learning channel conditions, has a steady performance, and is more dependable in real-world situations.

[1] By training data we mean ToA data received from user-vehicles who we know *a priori* to be legitimate or malicious. Use of such data in order to set the NN parameters, prior to its use on 'unlabeled' data (*i.e.*, data from user-vehicles who we do not know *a priori* to be legitimate or malicious), is known as the training phase. Knowing when to end this training phase is one of the key questions we answer in this work.

[2] The test data is simulated under a different realization with same settings as training data. Further, test data has no labels

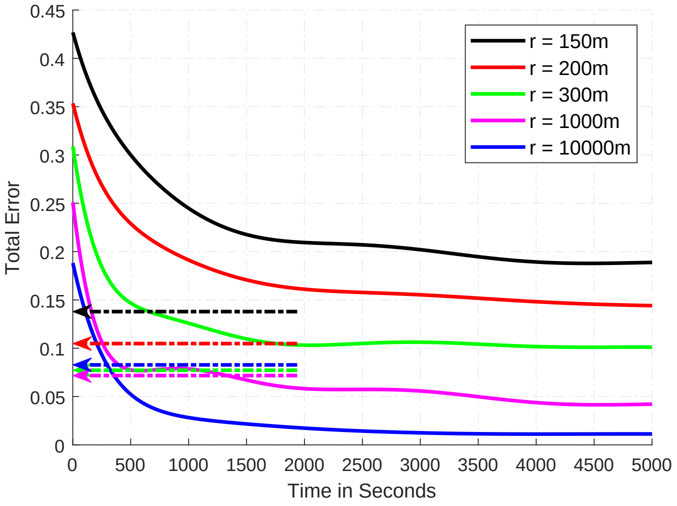

Fig. 2. Total Error performance for NN-LVS with four BSs and changing values of r. The scenario presented here closely relates to the real-world conditions where X_i is extracted from a random Gaussian distribution and the NLoS bias is extracted from a random exponential distribution. X_i and the NLoS bias have a fixed standard deviation of 100ns and 300ns respectively. The solid lines shows the Total Error performance for the NN-LVS under different values of r while the dashed arrow lines point to the respective Total Error calculated on the basis of the LRT method presented in [38]. The NN-LVS reports a continuous improvement in its performance with an increasing r.

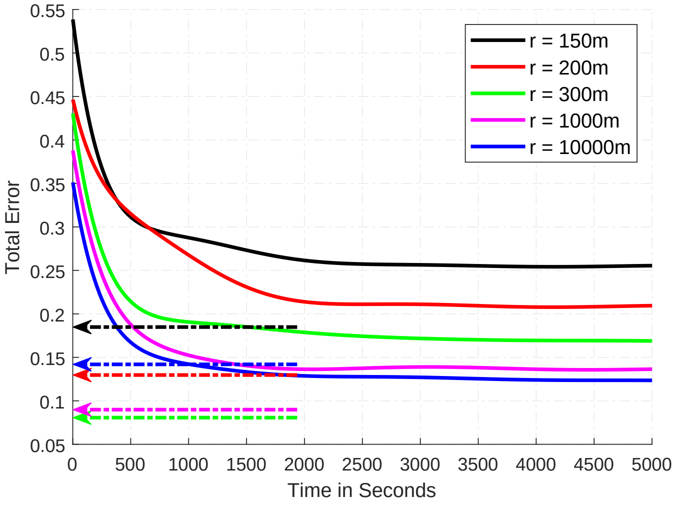

Fig. 3. Total Error performance for the NN-LVS as in Fig. 2 except the standard deviations for X_i and NLoS bias have been modified to 300ns and 100ns respectively.

We repeat our comparison study in Fig. 3. All other simulation parameters are kept the same as in Fig. 2 except the standard deviation for X_i and NLoS bias which are modified to 300ns, and 100ns, respectively. We can see again that compared to the information theoretic LVS, the NN-LVS has a better performance.

We evaluate and compare the performance of the NN-LVS with the information-theoretic LVS using a different metric, namely the Bayes Risk [40], given below

$$\mathcal{R} = p_0\, C_{00}\,(1-\alpha) + p_1\, C_{01}\,(1-\beta) + p_0\, C_{10}\,\alpha + p_1\, C_{11}\,\beta, \tag{13}$$

where p_0 and p_1 represents the proportion of the genuine and malicious user-vehicles. C_{00}, C_{01}, C_{10}, and C_{11} represent the different costs affecting the Bayes Risk; C_{00} is the cost associated with correctly identifying a genuine user-vehicle, C_{01} is the cost associated with falsely identifying a malicious user-vehicle as genuine (a missed detection), C_{10} is the cost associated with falsely identifying a genuine user-vehicle as malicious (a false positive), and C_{11} is the cost associated with correctly detecting a malicious user-vehicle. Here, C_{00} and C_{11} are considered 0 and therefore, have no impact on the Bayes Risk. After simplification Eq. (13) takes the form

$$\mathcal{R} = p_1\, C_{01}\,(1-\beta) + p_0\, C_{10}\,\alpha. \tag{14}$$

In the real-world scenario, a missed detection, in general, is considered a more serious issue as compared to a false positive. Henceforth, a higher cost is associated with a missed detection as compared to a false positive towards Bayes Risk calculation. With the unequal costs, the LRT decision threshold, λ, is also modified [40] and is given by

$$\frac{f\left(\mathbf{Y}|\mathcal{H}_1\right)}{f\left(\mathbf{Y}|\mathcal{H}_0\right)} \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \frac{p_0\left(C_{10} - C_{00}\right)}{p_1\left(C_{01} - C_{11}\right)}.$$

In this work we assume p_0, and p_1 are equal, *i.e.*, 0.5.

The classification threshold for the NN-LVS is also modified and is set equal to λ. This means if the output of the NN-LVS for a user-vehicle (from the test set) is greater than or equal to λ, the user-vehicle is classified as malicious, else it is classified as legitimate. For a wide range of changing costs for the missed detections (always greater than the cost for false positives) the same main result was found, namely the NN-LVS outperformed the information-theoretic LVS. As an example, we show the performance of the two LVSs by plotting the Bayes Risk with changing costs in Fig. 4.

C. *Other Network Issues*

We note our calculations in this work assume that network congestion issues do not play a role. That is, we assume the messages containing all the information required for the NN-based LVS calculations can be received (and processed) in a timely manner. If the number of user-vehicles becomes too large such that communications with the nearby BSs are interfered with the location verification processes will be curtailed. However, eventually we believe an LVS will be embedded directly within the WAVE architecture as described by the IEEE 1609 suite of standards and the IEEE 8011.p standard (see [3] for discussion of all these standards). WAVE mandates the position information of all user-vehicles is broadcast every 100ms as part of the wider vehicular network safety messages, and moreover, that priority is given to these messages. The bandwidth available to such high priority messages within WAVE is designed so as to accommodate all anticipated network conditions. In this work we assume all LVS messages are treated the same as the position information messages.

V. A Theoretical Lower Bound on Total Error

A quicker NN-LVS training stopping condition could be useful in that it can save training time, training data costs,

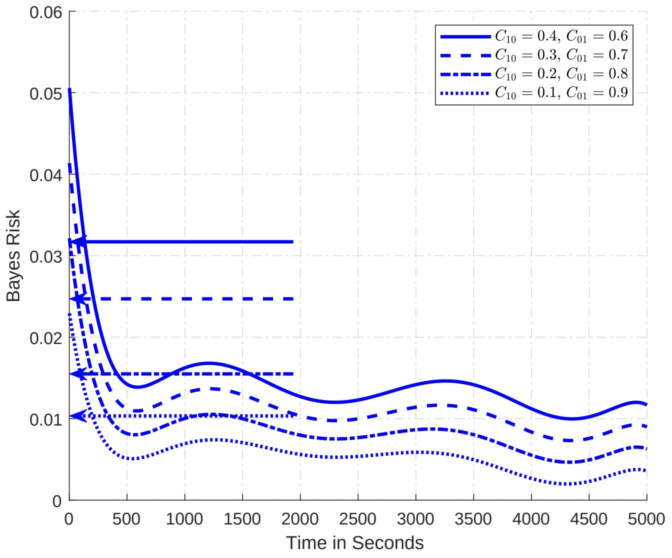

Fig. 4. Performance evaluation for the NN-LVS and information-theoretic LVS using Bayes Risk with changing costs C_{10} and C_{01}. Here the standard deviation for X_i, and NLoS bias are 100ns, and 300ns, respectively. The value of r is set to 10,000m. The different curves represent the Bayes Risk for the NN-LVS while the dashed arrows indicate the Bayes Risk for the information-theoretic LVS. It is evident from this figure that the NN-LVS outperforms the information-theoretic LVS in terms of Bayes Risk.

and make for a faster NN-LVS. To achieve this we introduce a new bound, referred to as the Total Error Lower Bound (TLB). This bound, as we will show, in conjunction with the usual learning process can indeed result in a stopping of the NN-LVS training at an earlier time - a time when the NN-LVS is performing adequately. The TLB is calculated using information-theoretic analysis as we now discuss assuming *a priori* knowledge on the distribution of the bias term at each BS.

A. Likelihood Functions under $\mathcal{H}_0$ and $\mathcal{H}_1$

The performance of the decision rule given in (9) is normally determined by the likelihood functions under $\mathcal{H}_0$ and $\mathcal{H}_1$. As such, in this subsection we derive the expressions of $f\left(Y_i|\mathcal{H}_0\right)$ and $f\left(Y_i|\mathcal{H}_1\right)$ explicitly.

Proposition 1: Following (2), the likelihood function of Y_i under $\mathcal{H}_0$ is derived as

$$f(Y_i|\mathcal{H}_0) = \frac{\rho_i}{2}\text{Erfc}\left(\frac{U_i+\rho_i\sigma_T^2-Y_i}{\sigma_T\sqrt{2}}\right)e^{\frac{\rho_i}{2}\left(2U_i+\rho_i\sigma_T^2-2Y_i\right)}, \tag{15}$$

where Erfc$(\cdot)$ is the complementary error function. Following (5), the likelihood function of Y_i under $\mathcal{H}_1$ is derived as

$$f(Y_i|\mathcal{H}_1) = \frac{\rho_i}{2}\text{Erfc}\left(\frac{V_i+\rho_i\sigma_T^2-Y_i}{\sigma_T\sqrt{2}}\right)e^{\frac{\rho_i}{2}\left(2V_i+\rho_i\sigma_T^2-2Y_i\right)}, \tag{16}$$

where $V_i = T_x + W_i$.

Proof: We detail the proof of (15) in the following and the proof of (16) follows a similar procedure. Since U_i and ϕ_i are assumed to be independent to each other, following (2) the expression of $f(Y_i|\mathcal{H}_0)$ is given by

$$f(Y_i|\mathcal{H}_0) = \frac{\rho_i}{\sqrt{2\pi}\sigma_T}\int_0^{\infty} e^{-\rho_i t-\frac{(Y_i-U_i-t)^2}{2\sigma_T^2}}dt. \tag{17}$$

With the aid of the following identity [41, Eq. (3.322.2)]

$$\int_0^{\infty} e^{-\frac{x^2}{4\beta}-\gamma x}dx = \sqrt{\pi\beta}e^{\beta\gamma^2}\text{Erfc}(\gamma\sqrt{\beta}),\ \text{for}\ \beta>0, \tag{18}$$

we solve the integral in (17), which leads to the desired result in (15). This completes the proof of this proposition. ■

Based on Proposition 1, the likelihood functions of $\mathbf{Y}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ can be explicitly determined as per (4) and (6), respectively. Due to the complex expressions of these likelihood functions, the false positive rate and the detection rate (*i.e.*, α and β, respectively) cannot be derived in closed-form expressions. This leads to the fact that the Total Error ξ cannot be derived or numerically calculated, and can only be evaluated through lengthy Monte Carlo simulations. As such, in the following subsection we determine a lower bound on the Total Error ξ, which can significantly facilitate our evaluation of the performance of the considered LVS.

B. A Lower Bound on Total Error

For the optimal binary decision rule, we have [42, 43]

$$\xi^* = \frac{1}{2}\left(1-\mathcal{V}_T(f(\mathbf{Y}|\mathcal{H}_0), f(\mathbf{Y}|\mathcal{H}_1))\right), \tag{19}$$

where $\mathcal{V}_T(f(\mathbf{Y}|\mathcal{H}_0), f(\mathbf{Y}|\mathcal{H}_1))$ is the total variation between $f(\mathbf{Y}|\mathcal{H}_0)$ and $f(\mathbf{Y}|\mathcal{H}_1)$. In general, computing $\mathcal{V}_T(f(\mathbf{Y}|\mathcal{H}_0), f(\mathbf{Y}|\mathcal{H}_1))$ occurs a high complexity and thus Pinsker's inequality is normally adopted to upper bound it. Based on Pinsker's inequality, we have

$$\mathcal{V}_T(f(\mathbf{Y}|\mathcal{H}_1), f(\mathbf{Y}|\mathcal{H}_0)) \leq \sqrt{\frac{1}{2}\mathcal{D}_{10}}, \tag{20}$$

or

$$\mathcal{V}_T(f(\mathbf{Y}|\mathcal{H}_0), f(\mathbf{Y}|\mathcal{H}_1)) \leq \sqrt{\frac{1}{2}\mathcal{D}_{01}}, \tag{21}$$

where $\mathcal{D}_{10}$ is the Kullback-Leibler (KL) divergence from $f(\mathbf{Y}|\mathcal{H}_1)$ to $f(\mathbf{Y}|\mathcal{H}_0)$, which is given by

$$\mathcal{D}_{10} = \int_{\mathcal{Y}} f(\mathbf{Y}|\mathcal{H}_1)\log\frac{f(\mathbf{Y}|\mathcal{H}_1)}{f(\mathbf{Y}|\mathcal{H}_0)}d\mathbf{Y}, \tag{22}$$

and $\mathcal{D}_{01}$ is the KL divergence from $f(\mathbf{Y}|\mathcal{H}_0)$ to $f(\mathbf{Y}|\mathcal{H}_1)$, which is given by

$$\mathcal{D}_{01} = \int_{\mathcal{Y}} f(\mathbf{Y}|\mathcal{H}_0)\log\frac{f(\mathbf{Y}|\mathcal{H}_0)}{f(\mathbf{Y}|\mathcal{H}_1)}d\mathbf{Y}. \tag{23}$$

We note that both (20) and (21) are valid, although they are different due to the asymmetry of the KL divergence, which can be seen from (22) and (23). Following (19), (20), and (21), a lower bound on the minimum Total Error ξ^* can be written as

$$\xi^* \geq \xi_l \triangleq \frac{1}{2}\left(1-\min\left[\sqrt{\frac{1}{2}\mathcal{D}_{10}}, \sqrt{\frac{1}{2}\mathcal{D}_{01}}\right]\right). \tag{24}$$

The lower bound ξ_l is of significant usefulness in the context of binary detection systems when the actual performance of a system cannot be directly examined. In the presence of bias the first outcome achieved from this lower bound is on the determination of the optimal T_x (*i.e.*, T_x^*). From the malicious user-vehicle's point of view, maximizing the Total Error is the goal. As such, when non-zero bias terms are present we consider that T_x^* is the value that maximizes this lower bound ξ_l. We note that this lower bound ξ_l is based on perfect knowledge of some system parameters (*i.e.*, the distribution of ϕ_i). However, this information may not be perfectly achieved in practical scenarios, which motivates us to adopt the machine-learning methodology in the considered LVS. The usefulness of this lower bound on the minimum detection error probability lies in the fact that it can provide practical guidelines on when to stop training the machine.

C. Truncated Gaussian Distributed Bias

If the bias term ϕ_i follows a truncated normal distribution, the pdf of ϕ given in (3) should be updated to

$$f(\phi_i) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma_i} e^{-\frac{\phi_i^2}{2\sigma_i^2}}, \tag{25}$$

where the mean and variance of the normal distribution before the truncation are zero and σ_i^2, respectively. Then, Proposition 1 should be updated to Proposition 2 as below:

Proposition 2: Following (2), the likelihood function of Y_i under $\mathcal{H}_0$ is derived as

$$f(Y_i|\mathcal{H}_0) = \frac{e^{-\frac{(U_i-Y_i)^2}{2(\sigma_i^2+\sigma_T^2)}}}{\sqrt{2\pi(\sigma_i^2+\sigma_T^2)}} \text{Erfc}\left(\frac{\sigma_i(U_i-Y_i)}{\sigma_T\sqrt{2(\sigma_i^2+\sigma_T^2)}}\right). \tag{26}$$

Following (5), the likelihood function of Y_i under $\mathcal{H}_1$ is derived as

$$f(Y_i|\mathcal{H}_1) = \frac{e^{-\frac{(V_i-Y_i)^2}{2(\sigma_i^2+\sigma_T^2)}}}{\sqrt{2\pi(\sigma_i^2+\sigma_T^2)}} \text{Erfc}\left(\frac{\sigma_i(V_i-Y_i)}{\sigma_T\sqrt{2(\sigma_i^2+\sigma_T^2)}}\right), \tag{27}$$

where $V_i = T_x + W_i$.

Proof: We detail the proof of (26) in the following and the proof of (27) follows a similar procedure. Since U_i and ϕ_i are assumed to be independent of each other, following (2) the expression of $f(Y_i|\mathcal{H}_0)$ is given by

$$f(Y_i|\mathcal{H}_0) = \frac{1}{2\pi\sigma_T\sigma_i}\int_0^\infty e^{-\frac{t^2}{2\sigma_i^2}-\frac{(Y_i-U_i-t)^2}{2\sigma_T^2}} dt. \tag{28}$$

With the aid of the following identify [41, Eq. (3.322.2)]

$$\int_0^\infty e^{-\frac{x^2}{4\beta}-\gamma x} dx = \sqrt{\pi\beta} e^{\beta\gamma^2} \text{Erfc}(\gamma\sqrt{\beta}), \text{ for } \beta > 0, \tag{29}$$

we solve the integral in (28), which leads to the desired result in (26). This completes the proof of this proposition. ∎

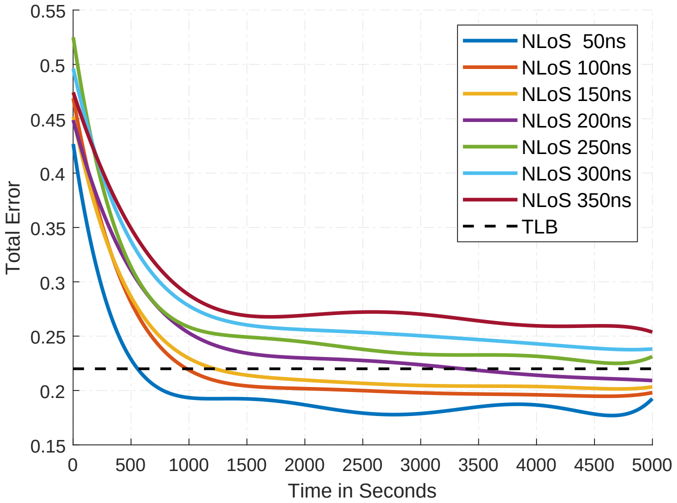

Fig. 5. Training of NN-LVS. Training and test data is simulated with four BSs. X_i has a fixed standard deviation of 300ns. r is 200m. Standard deviation for NLoS is changing in training and test data set at the same time which is represented by different colour of curves in the figure. TLB is the information theoretic Total Error lower bound. The Total Error for the NN-LVS increases with an increase in the NLoS bias.

VI. Numerical Results with the Total Error Lower Bound

We now present our results for both the information theoretic LVS and the NN-LVS. Four static BSs; two each at the two ends, with known true locations are present in a 1000m X 500m area. The area in focus resembles to a small district. The claimed locations for the to-be-verified user-vehicles are within the same area. For simulating the attacking scenario, we consider two user-vehicles, *i.e.*, a legitimate user-vehicle which is reporting its true location and a malicious user-vehicle which is falsifying its true location. We assume the locations of all BSs are known to any user-vehicle. This last assumption can be taken to mean any malicious user-vehicle can intercept all GPS information between the BSs.

As stated earlier, the malicious user-vehicle optimizes its attack location by setting T_x to T_x^* in an attempt to minimize its chances of being detected. In all the simulations, the malicious user-vehicle's forged location is at a minimum distance constraint, r, from its true location. If the malicious user-vehicle violates the minimum distance constraint, it will be easily caught by the BSs. The constraint r is an *a priori* known distance and is set to 200m in our simulations unless otherwise specified.

Simulated ToA data is used in this work. The *claimed* locations for genuine as well as malicious user-vehicles (in equal proportions) are generated randomly in the specified area and their respective ToAs is calculated at the four BSs. The malicious user-vehicle optimizes its true location at a distance r from its claimed location. We use equation (8) to calculate T_x^*. The receivers in the BSs are under the influence of independent thermal noise X_i and thus the ToA measurements they make have a certain degree of variation. We extract this variation (in nanoseconds) from a Gaussian random function that has a fixed standard deviation. As described in equation (2) and (5), NLoS bias, ϕ_i, is added to the ToAs of the respective user-vehicles. To mimic reality, we extract the NLoS bias from an exponential distribution with a fixed standard deviation as highlighted in equation (3).

We now take into account the TLB into our study in Fig. 5. The TLB is calculated using equation (24). In the TLB calculation, r is equal to 200m while the standard deviation for X_i and NLoS bias is 300ns, and 900ns respectively. We assume this value of standard deviation for NLoS bias to closely relate to the global average for NLoS[3]. The calculated TLB with four BSs and the above settings is 0.22.

For training the NN-LVS in Fig. 5, we use a training data set which is simulated with a NLoS bias standard deviation of 50ns. Other simulation parameters are the same as those considered for the TLB calculations earlier. The training data has genuine and malicious user-vehicles in equal proportions. Prior to training the NN-LVS, the training data is randomized to closely mimic the reality. The NN-LVS is trained with the random user-vehicle data in each second and is further used to calculate a Total Error for the test data. Before re-training the NN-LVS and calculating a revised Total Error for the test data in each subsequent second, the previous training data is accumulated with a new random user-vehicle training data from that second. Total Error for the test data is plotted against time as shown in Fig. 5 (a polynomial fitting is applied). The figure also has Total Error curves for different values of NLoS bias standard deviation.

In Fig. 6, we carried out the same simulation as in Fig. 5 but revised the number of BSs to six. We did not change any of the other simulation parameters. One can see that the NN-LVS with six BSs performs better than the NN-LVS with four BSs. This highlights the fact that more number of BSs will result in better performance for the NN-LVS.

In Fig. 7, we study the impact of changing the BS's thermal noise, X_i, in terms of the Total Error for the NN-LVS. It is evident from this figure that NN-LVS shows an improved performance at lower values of X_i. The TLB in Fig. 7 has been calculated with X_i = 300ns, NLoS bias equal to 900ns and r = 200m. We see the case of X_i = 300ns represents a scenario where the TLB would not be used as the stopping condition.

We have carried out numerous numerical simulations of our trained NN-LVSs as applied to a wide range of data simulating real-world deployment. The performance in term of the Total Error of the NN-LVSs in these tests show similar results to the flat parts of the curves shown in Figures 5 to 7. That is, our NN-LVSs once trained perform as expected in our mimicked real-world deployments. In a nutshell, we see that when the NN-LVS training is stopped based on the TLB, we on average can cut the training time by a factor of $\frac{1}{5}$. That is, we can save training data costs and ensure quick deployment of the NN-LVS with minimal impact on the overall performance.

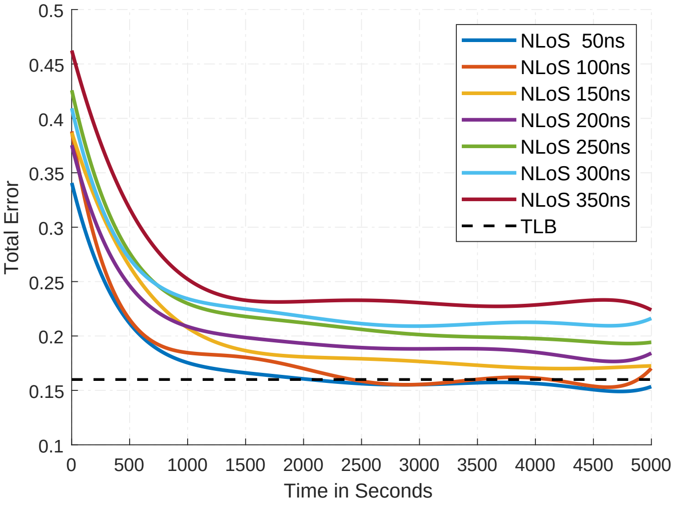

Fig. 6. Training of the NN-LVS as in Fig. 5 except the number of BSs in this simulation has been modified to six.

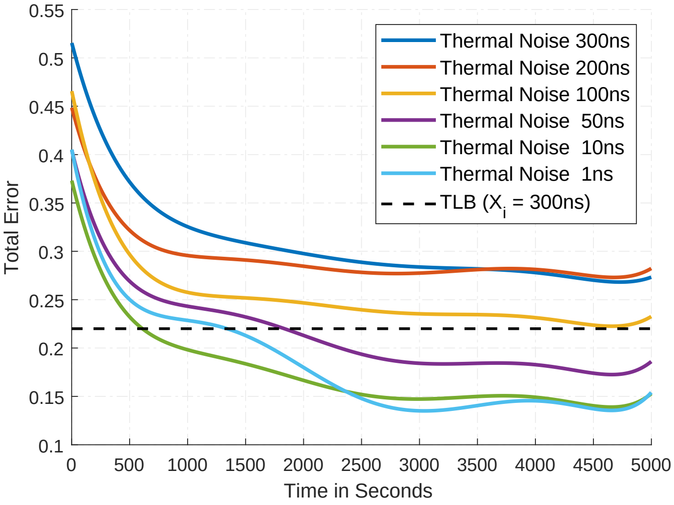

Fig. 7. NN-LVS performance with changing values of receivers thermal noise in the BSs. Other parameter settings in this simulation are: Standard deviation for NLoS is 400ns and the number of BSs are four. It can be seen that NN-LVS performs better with low receiver thermal noise.

VII. Conclusion

We have presented a new system for solving the important problem of location verification in the context of vehicular networks. Encompassing information theory and machine-learning concepts, the solutions we have provided will lead to enhanced and pragmatic location verification outcomes. Of particular importance is the fact that we have shown how a NN based LVS is able to outperform a pure information-theoretic LVS when channel conditions are *a priori* unknown and/or the malicious user-vehicle is at an unspecified distance from its claimed location. We have also shown how optimal-information-theoretic concepts can be encapsulated with the machine learning framework so as to allow for a very useful trade-off in the training-time *vs.* performance - thus making our solution even more practical in real-world deployments.

Our work will prove vital for the performance of location-oriented applications within IoT (such as smart vehicular

[3] That is, we are assuming here that we have *a priori* knowledge on the channel conditions over some global average. We then use this TLB to provide a new stopping condition under a range of differing real-world test data. As we will see, there will be some test data where the TLB has no effect (usual stopping condition applies), and other test data where the stopping condition is reached well before the usual stopping condition. Even though the new training time is only about 1/5 of the usual training time, the performance 'hit' is minimal.

networks, smart traffic lights, intelligent parking system, map services, smart supply chain and logistics, smart farming, etc.) and for enhancing the efficiency of numerous features within future wireless telecommunication systems (e.g., beamforming, Massive MIMO for enhanced throughput, and interference mitigation for quality and capacity improvements).

References

[1] H. Hartenstein and L. Laberteaux, "A tutorial survey on vehicular ad hoc networks," *IEEE Communications Magazine*, vol. 46, no. 6, pp. 164–171, Jun. 2008.

[2] Number of road traffic deaths. Accessed on: August 20, 2019. [Online]. Available: https://www.who.int/gho/road_safety/mortality/traffic_deaths_number/en/

[3] R. Meneguette, E. Robson, and A. Loureiro, *Intelligent Transport System in Smart Cities: Aspects and Challenges of Vehicular Networks and Cloud*, 1st ed. Springer, May. 2018.

[4] D. Sheet, O. Kaiwartya, A. Abdullah, Y. Cao, H. Ahmed, and K. Sushil, "Location information verification using transferable belief model for geographic routing in vehicular ad hoc networks," *IET Intelligent Transportation Systems*, vol. 11, no. 2, pp. 53–60, Mar. 2017.

[5] R. Kasana, K. Sushil, O. Kaiwartya, W. Yan, Y. Cao, and A. Abdullah, "Location error resilient geographical routing for vehicular ad-hoc networks," *IET Intelligent Transport Systems*, vol. 11, no. 8, pp. 450–458, Oct. 2017.

[6] S. Yan and R. Malaney, "Location verification systems in emerging wireless networks," *ZTE Communications*, vol. 11, no. 3, pp. 03–10, Jul. 2013.

[7] S. Yan, R. Malaney, I. Nevat, and G. Peters, "Optimal information-theoretic wireless location verification," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 7, pp. 3410–3422, Sep. 2014.

[8] S. Yan, I. Nevat, G. Peters, and R. Malaney, "Location verification systems under spatially correlated shadowing," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4132–4144, Jun. 2016.

[9] S. Yan, R. Malaney, I. Nevat, and G. Peters, "Location verification systems for VANETs in Rician fading channels," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5652–5664, Jul. 2016.

[10] M. Monteiro, J. Rebelatto, and R. Souza, "Information-theoretic location verification system with directional antennas for vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 93–103, Jan. 2016.

[11] E. Abdulhay, N. Arunkumar, K. Narasimhan, E. Vellaiappan, and V. Venkatraman, "Gait and tremor investigation using machine learning techniques for the diagnosis of parkinson disease," *Future Generation Computer Systems*, vol. 83, pp. 366–373, Jun. 2018.

[12] B. Ramsundar, S. Kearnes, P. Riley, D. Webster, D. Konerding, and V. Pande, "Massively Multitask Networks for Drug Discovery," *arXiv:1502.02072*, Feb. 2015.

[13] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proceedings of The European conference on Computer Vision*. Springer, Jul. 2006, pp. 430–443.

[14] G. Hinton, L. Deng *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[15] A. Giusti, J. Guzzi, D. Cireşan *et al.*, "A machine learning approach to visual perception of forest trails for mobile robots," *IEEE Robotics and Automation Letters*, vol. 1, no. 2, pp. 661–667, Jul. 2016.

[16] E. Guizzo, "How Google's self-driving car works," *IEEE Spectrum*, vol. 18, no. 7, pp. 1132–1141, Oct. 2011.

[17] B. Yu, C. Xu, and B. Xiao, "Detecting sybil attacks in VANETs," *Journal of Parallel and Distributed Computing*, vol. 73, no. 6, pp. 746–756, Jun. 2013.

[18] N. Sastry, U. Shankar, and D. Wagner, "Secure verification of location claims," in *Proceedings of the ACM Workshop on Wireless Security*, Sep. 2003, pp. 1–10.

[19] S. Capkun and J. Hubaux, "Secure positioning in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 2, pp. 221–232, Feb. 2006.

[20] T. Leinmuller, E. Schoch, and F. Kargl, "Position verification approaches for vehicular ad hoc networks," *IEEE Wireless Communications*, vol. 13, no. 5, pp. 16–21, Oct. 2006.

[21] C. Harsch, A. Festag, and P. Papadimitratos, "Secure position-based routing for VANETs," in *Proceedings of the IEEE Vehicular Technology Conference*, Oct. 2007, pp. 26–30.

[22] K. Rasmussen, M. Srivastava *et al.*, "Secure location verification with hidden and mobile base stations," *IEEE Transactions on Mobile Computing*, vol. 7, no. 4, pp. 470–483, Apr. 2008.

[23] G. Yan, S. Olariu, and M. Weigle, "Providing VANET security through active position detection," *Computer Communications*, vol. 31, no. 12, pp. 2883–2897, Jul. 2008.

[24] P. Zhang, Z. Zhang, and A. Boukerche, "Cooperative location verification for vehicular ad-hoc networks," in *Proceedings of the IEEE International Conference on Communications (ICC)*, Jun. 2012, pp. 37–41.

[25] F. Malandrino, C. Casetti, C. Chiasserini *et al.*, "A-VIP: Anonymous verification and inference of positions in vehicular networks," in *Proceedings of the IEEE INFOCOM*, Apr. 2013, pp. 105–109.

[26] W. Jaballah, M. Conti, M. Mosbah, and C. Palazzi, "Secure verification of location claims on a vehicular safety application," in *Proceedings of the International Conference on Computer Communication and Networks (ICCCN)*, Aug. 2013, pp. 1–7.

[27] I. Kim, B. Kim, and J. Song, "An efficient location verification scheme for static wireless sensor networks," *Sensors*, vol. 17, no. 2, p. 225, Jan. 2017.

[28] G. Caparra, M. Centenaro, N. Laurenti, and S. Tomasin, "Optimization of anchor nodes' usage for location verification systems," in *Proceedings of the International Conference on Localization and GNSS (ICL-GNSS)*, Jun. 2017, pp. 1–6.

[29] C. Vaas, M. Juuti, N. Asokan, and I. Martinovic, "Get in line: Ongoing co-presence verification of a vehicle formation based on driving trajectories," in *Proceedings of the European Symposium on Security and Privacy (Euro S&P)*, Apr. 2018, pp. 199–213.

[30] L. Fridman, D. Brown *et al.*, "MIT autonomous vehicle technology study: Large-scale deep learning based analysis of driver behavior and interaction with automation," *arXiv:1711.06976*, Nov. 2017.

[31] A. Brighente, F. Formaggio, M. Centenaro, G. M. Di Nunzio, and S. Tomasin, "Location-verification and network planning via machine learning approaches," *arXiv:1811.06729*, Nov. 2018.

[32] N. Kumar, S. Zeadally, and J. Rodrigues, "Vehicular delay-tolerant networks for smart grid data management using mobile edge computing," *IEEE Communications Magazine*, vol. 54, no. 10, pp. 60–66, Oct. 2016.

[33] A. Dua, N. Kumar, and S. Bawa, "Game theoretic approach for real-time data dissemination and offloading in vehicular ad hoc networks," *Journal of Real-Time Image Processing*, vol. 13, no. 3, pp. 627–644, Sep. 2017.

[34] A. Dua, N. Kumar, A. Kumar, and W. Susilo, "Secure message communication protocol among vehicles in smart city," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4359–4373, May 2018.

[35] O. Khan, M. Shah, I. Din, B. Kim, H. Khattak, J. Rodrigues, H. Farman, and B. Jan, "Leveraging named data networking for fragmented networks in smart metropolitan cities," *IEEE Access*, vol. 6, pp. 75 899–75 911, Nov. 2018.

[36] I. Din, B. Kim, S. Hassan, M. Guizani, M. Atiquzzaman, and J. Rodrigues, "Information-centric network-based vehicular communications: overview and research opportunities," *Sensors*, vol. 18, no. 11, Nov. 2018.

[37] I. Din, H. Asmat, and M. Guizani, "A review of information centric network-based internet of things: communication architectures, design issues, and research opportunities," *Multimedia Tools and Applications*, pp. 1–16, Dec. 2018.

[38] S. Yan, R. Malaney, I. Nevat, and G. Peters, "Timing information in wireless communications and optimal location verification frameworks," in *Proceedings of the Australian Communications Theory Workshop (AusCTW)*, Feb. 2014, pp. 144–149.

[39] J. Neyman and E. Pearson, "IX. On the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences*, vol. 231, no. 694-706, pp. 289–337, Feb. 1933.

[40] M. Barkat, *Signal detection and estimation*, 2nd ed. Boston: Artech House, Dec. 1991, ch. 5.

[41] I. Gradshteyn and I. Ryzhik, *Table of integrals, series, and products*. San Diego, CA: Academic Press Inc., Sep. 2014.

[42] E. L. Lehmann and J. P. Romano, *Testing statistical hypotheses*, 3rd ed. New York: Springer Science & Business Media, Mar. 2006.

[43] S. Yan, B. He, X. Zhou, Y. Cong, and A. L. Swindlehurst, "Delay-intolerant covert communications with either fixed or random transmit power," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 1, pp. 129–140, Jan. 2019.