Experience: Accurate Simulation of Dense Scenarios with Hundreds of Vehicular Transmitters

Bin Cheng, Ali Rostami, Marco Gruteser
WINLAB, Rutgers University
North Brunswick, NJ 08902, USA
{cb3974, rostami, gruteser}@winlab.rutgers.edu

ABSTRACT

This paper reports on our methodology and experience from a multi-year effort to cross-validate a vehicular network experiment with four hundred Dedicated Short Range Communications IEEE 802.11p transmitters through ns-3 simulations. With most of these transmitters in communication range, this represents an extremely dense wireless configuration that challenges radio and interference models. Field test and simulations were conducted in tandem and iteratively to facilitate model selection and configuration as well as to allow a detailed evaluation of simulation accuracy. We have learned that 1) results were most sensitive to parameter choices in the propagation and receiver models, with simulator default parameters not providing a good match; 2) results could, however, be significantly improved by adapting, implementing, and calibrating the propagation models and receiver models from the literature, yielding 88% accuracy (in terms of packet error rate compared to the field test) in such a complex large-scale setting; 3) the process was helpful in identifying errors both in the simulation models and in the experimental code and points to opportunities for further research.

CCS Concepts

•**Networks** → **Network simulations;** *Network performance analysis;*

Keywords

Large-scale Network Simulation, Vehicle-to-Vehicle Communication, Simulation Accuracy Analysis

1. INTRODUCTION

Given the increase of programmable wireless hardware, mobile networking research has shifted over the past decades from a heavy reliance on simulation towards a more experiment-driven methodology. Yet, network simulators remain an important tool, for example, for complementing studies of large-scale networks with hundreds or thousands of wireless nodes that require significant resources to fully evaluate experimentally. The validity of such simulation results is frequently questioned, and, however, little data exists for verifying the accuracy of modern network simulators in such settings.

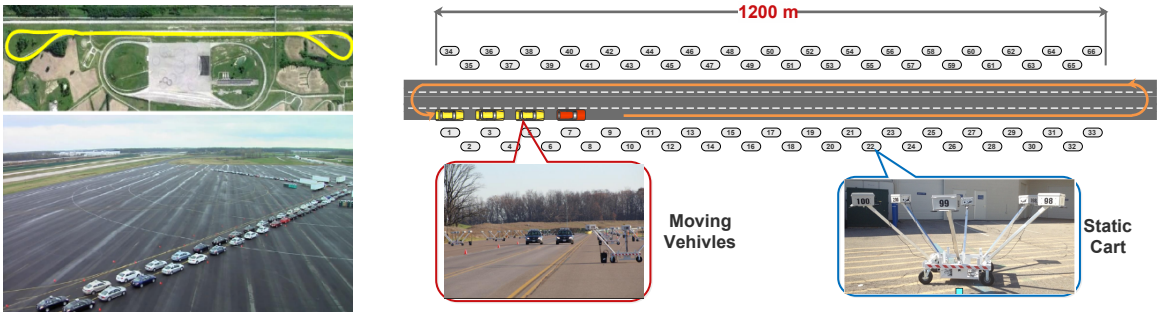
To fill this void, this paper reports on our methodology and experience from a multi-year effort to cross-validate a vehicular network experiment with four hundred Dedicated Short Range Communications (DSRC) IEEE 802.11p transmitters through ns-3 simulations. The primary use case for DSRC technology is for vehicles to periodically exchange their status information (including position, speed, heading, etc.) with other nearby traffic participants, thereby improving situational awareness among all traffic participants, whether human-driven or self-driving. This technology should also support safety applications with coverage and latency requirements that would be challenging to accommodate on cellular networks. Given the likely government mandate[1] and current deployment plans, it is possible that hundreds, even thousands, of periodically transmitting vehicles will congregate within communication range in future traffic situations. Such vehicle-to-vehicle communication (V2V) experiments can, therefore, serve as a particularly dense case study of wireless network simulations. In addition, the safety-related character of the system leads to heightened need for validation of results, which has made it possible to conduct field test with two hundred vehicles and up to four hundred prototype transmitters.

To date, simulation accuracy has only been studied with a relatively small number of nodes [4, 10–12]. Over the years, many model improvements have been proposed for network simulators. As to the receiver modeling, Chen et al. [8] proposed a novel receiver model for the ns-2 simulator. This model first introduced the frame capture effect into the network simulator. Based on [8], Bingmann et al. [6] developed a similar receiver model for the ns-3 simulator. Going beyond a packet-level simulator, Papanastasiou et al. [17] proposed models to mimic the device's behavior at a signal level, which promises more accuracy but significantly increases the computational load particularly in large-scale simulations. As to the propagation modeling, many chan-

[1]The United States National Highway Traffic Safety Administration (NHTSA) has issued an advance notice of proposed rulemaking with an intention to require V2V capability in new cars since 2020 [16]. The secretary of the United States Department of Transportation (USDOT) has also announced an aggressive regulatory plan for V2V deployment in the near future [2].

MobiCom'16, October 03-07, 2016, New York City, NY, USA

DOI: http://dx.doi.org/10.1145/2973750.2973779



(a) (b)

Figure 1: (a) The phase 1 testing environment illustration: the bird's eye view of the phase 1 testing facility (top) and an illustration of the vehicle distribution in one test (bottom); (b) the carts and moving vehicles layout in the phase 2 test

nel models have been proposed for characterizing the signal propagation in a V2V environment. Some of them were derived based on the gained experience in mobile cellular networks [22]. Many of them were derived based on the data from field measurements, e.g., [7, 13, 20, 21, 23]. However, the scale of these measurements was normally small, i.e., none of them involved hundreds of vehicles. Other large-scale testbeds, such as the Michigan Safety Pilot [1] also use large numbers of vehicles but they are distributed over a large geographic region, so that network load and interference effects usually remain negligible. We are not aware of any large-scale evaluation of simulation accuracy for any of these models or datasets.

In this effort, we, therefore, co-developed an experiment design and simulator model that would allow accuracy comparisons in V2V scenarios with hundreds of nodes. The key components that a V2V network simulator primarily relies on are: a network traffic generator, a node mobility manager, a packet reception controller (via receiver models) and a channel and signal propagation emulator (via propagation models). In a V2V network, the primary network workload is simply periodic broadcast messages, the basic safety messages, which simplifies the implementation of the network traffic generator. For node mobility, we use a trace-driven approach that relies on Global Position System (GPS) readings from the field experiment. The receiver and propagation models require more attention and therefore we focus in this paper on these models. In particular, we studied the accuracy that can be obtained with different model complexities, confirmed exact receiver parameters through hardware lab testing, and studied the sensitivity to the choice of receiver and propagation parameters.

In this work, we have learned

- The existing simulation models with default parameters may only achieve 15% simulation accuracy in term of packet error ratio. However, with calibrated parameters, the accuracy of these models can improve by $\sim 20\%$.

- Adjusting the complexity of propagation models and receiver models can result in 10-20% improvements in simulation accuracy with 8% overhead in simulation runtime and for well-calibrated models, it is possible to achieve 88% accuracy.

- The calibration process is helpful in identifying implementation errors of simulators. A calibrated simulator can assist researchers in validating, planning and predicting the field experiments results.

2. BACKGROUND

2.1 V2V Safety Communication

The basic concept of V2V safety communication is that each vehicle periodically broadcasts its most recent driving status to its nearby vehicles via safety messages; once a safety message is received, the receiving vehicle assesses the collision possibility with the sending vehicle; if the two vehicles are assessed to be in danger, an immediate action is taken by the vehicle to prevent the collision. The action can be to warn the driver or to reduce the vehicle speed automatedly. Thereby, an effective V2V safety communication system requires reliable and low-latency network support. However, once a large number of vehicles are present nearby, the shared communication channel can become highly congested. The Carrier Sense Multiple Access/Collision Avoidance Media Access Control (CSMA/CA MAC) protocol used by the DSRC technology is not sufficient to mitigate all the packet collisions and thus some such safety messages are lost. In this case, vehicles can lose the situational awareness of its nearby vehicles, which potentially results in a delayed reaction to a dangerous situation. This issue is known as the scalability issue, and it is one of the main challenges of the large-scale deployment of V2V systems.

2.2 Field Experiment Setting

To investigate the V2V scalability issue and study the congestion control, the Crash Avoidance Metrics Partnership (CAMP) Vehicle Safety Communications 3 (VSC3) Consortium, in partnership with USDOT, has conducted a series of V2V experiments in several testing facilities to evaluate seven driving conditions.

In these field experiments, each DSRC transceiver broadcasted safety messages several times per second (the default rate is 10 Hz) at data rate 6 Mbps on a channel with 10 MHz bandwidth and 5.9 GHz center frequency, using Atheros 802.11p chips. The packet size of safety messages varied from 310 to 390 Bytes due to different frame payload size.

Two major field test activities were conducted in this project. The phase 1 test involved up to 200 On-Board

Equipment devices (OBEs). Each OBE was mounted on the roof of a vehicle, acting as a DSRC transceiver. Each OBE acquired its position from an integrated GPS device at rate 10 Hz. In the phase 1 test, six primary driving configurations were tested: highway, intersection, V2V safety application, high dynamics winding road, hidden node and sudden loading effect. In each test trial, the vehicles were driving along a predefined route at a nearly constant speed with the intention of maintaining the separation distance between two adjacent moving vehicles constant. Fig 1a provides a bird's eye view of the testing facility and a snapshot of the vehicle distribution in one field experiment.

As an extension to the phase 1 test, the phase 2 test increased the number of OBEs to 400. The test was conducted in an open-space environment, and it primarily relied on the use of OBEs mounted on stationary carts, each of which held up to six OBEs, leading 400 OBEs to mount on 66 carts. These carts were placed in two rows on each side of a straight track and evenly spaced within each row. The length of the track was 1200 m and the spacing between two neighboring carts was 37.5 m. The test used a combination of stationary OBE carts and a small number of vehicles. In the majority of test trials, four moving vehicles were used. The speed of these moving vehicles was set to 40 km/h and the separation distance between two adjacent moving vehicles was 75 m. The layout of the carts and the moving vehicles is illustrated in Fig. 1b.

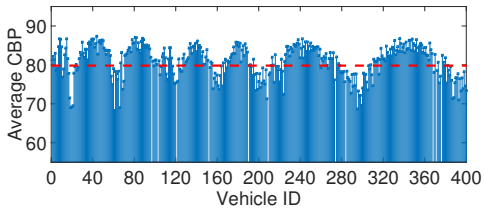


Figure 2: Average CBP measured by each OBE

The phase 2 test models a highway traffic jam where most vehicles are stationary and move forward only rarely. The moving vehicles mimic vehicles driving on less congested lanes, e.g., a carpool lane. Such a scenario of high density of stationary vehicles and a few fast moving vehicles is a particularly challenging scenario in V2V safety communications since reliable and low-latency communication needs to be maintained for the fast moving vehicles in this high-density situation. With such a large number of transmitters in the test, the channel was highly congested. The channel condition was indicated through channel busy percentage (CBP), which is defined as the percentage of the period during which the channel is measured as busy. As shown in Fig. 2, more than 88% OBEs observed that their average CBP was higher than 75%. To produce even higher channel load, higher transmission rates (up to 50 Hz) were used by each OBE to emulate up to 2000 transmitters on the channel. To avoid unnecessary redundancy while analyzing the performance of each OBE, a subset of OBEs, named loggers, were selected as representatives for data collection and performance evaluation. The selected 33 loggers were uniformly distributed in the experiment area. Due to various link distances between loggers, the link quality, indicated by the Received Signal Strength Indicator (RSSI) measurements, varied in a wide range, from -55 dBm to -95 dBm. Fig. 3 illustrates the connected links between several loggers and the average RSSI for each link.

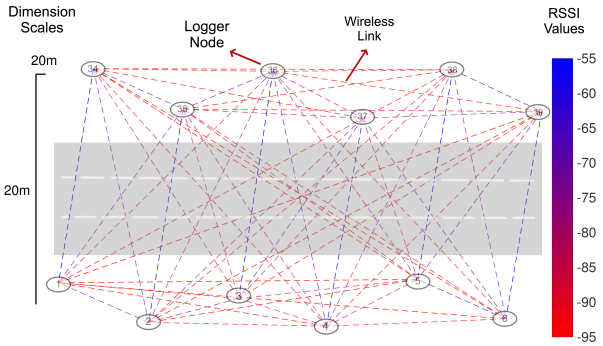


Figure 3: Link connections between a subset of loggers

It required a significant amount of time (nearly one day for the CAMP team) to deploy such a large-scale network over the experiment area. Thereby, for this multi-day test, the devices had to stay on the testing ground overnight. Due to the dropped temperature at night, icing occurred on the surface of the OBE boxes, which could cause temporary port connection issues (e.g., a port supplying the power to the OBE) while booting the OBEs in the next morning. However, note that all the data were collected when no device malfunctions occurred. Also to control this large-scale network and monitor the network operation in real time, a separate control network was built via a WiFi router integrated into each OBE and several access points. With this control network, the control center was able to notify OBEs to load configurations, to start/stop tests at approximately the same time. Meanwhile, the running status of each OBE was reported to the control center and displayed through visualization tools. By this way, any device failure or misbehavior could be detected and monitored in real time.

In this work, we primarily focused on simulating the phase 2 test because it provided a higher transmitter density and a more congested channel which challenges the congestion control study.

2.3 Simulating Signal Propagation

In a network simulator, the propagation model is responsible for reflecting the environmental effects on the channel, such as signal blockage and reflection, and then determining the power and delay of a transmitted signal at a receiver. The main factors of a propagation model are the distance-dependent path loss and the fading. The distance-dependent path loss captures how the average received power level varies with distance to the transmitter. The fading captures how the instantaneous signal level fluctuates over time, frequency, and space. Typically, a statistical modeling framework of received signal power (RSSI in dB) at a random time point and for a given distance of a transmitter-receiver pair is given by

$$RSSI = P_t + [10\log_{10} g_{med} + 10\log_{10} F_{sh}] + 10\log_{10} F_{mp}$$

In this equation, the bracketed term is the locally averaged path loss for a particular transmitter-receiver distance d; $g_{med}(d)$ is the median value of the path loss over all transmitter-receiver links of length d. This term is distance-

dependent. One simple model for this term is the **log-distance model**, where the path loss in dB changes linearly with the distance, $Gmed = 10\log_{10} gmed = A - 10Blog10(d)$ [19]. If there are no or only small obstacles between the transmitter and the receiver, the received signal could consist of a line of sight signal component and multiple reflected signal components dominated by a ground reflected signal. This effect can be modeled by the **two-ray model** [19]. To model a complex signal propagation environment, multiple models can be applied jointly, e.g., before a distance break point, the two-ray model is applied. After that point, the log-distance model is applied. For a particular link, F_{sh} represents the large-scale fading which varies slowly with physical location depending on the structures (e.g. buildings) in the environment. Several previous works [9,19] have suggested the large-scale fading follows the Gaussian distribution. F_{mp} represents the small-scale fading due to multipath and it can be modeled by the Nakagami or the Gaussian distributed random variable.

2.4 Simulating Receiving Behaviors

A wireless receiver starts the reception of a packet with looking for a known pattern of the preamble. If such pattern is found, the receiver then attempts to decode the Physical Layer Convergence Procedure (PLCP) header, which contains details of this transmission including data rate, frame length, etc. If the receiver decodes PLCP header successfully, it can next demodulate the frame body. Until the end of the frame body duration, all the incoming signals fluctuate the receiving packet's Signal-to-Interference-and-Noise-Ratio (SINR) and thus affect the current reception. At the end of the reception of a packet, a CRC code check is performed to determine if this packet can be received correctly.

In a network simulator, the receiver model typically determines whether a packet is received successfully based on the packet's SINR. As in [6, 8], the receiver's behaviors are normally modeled by identifying different states, e.g. TX, IDLE, CCA_BUSY, and RX.

A node is in TX state if it is transmitting a packet. If the receiver is in this state while a new packet arrives, the receiver will drop the newly arrived packet.

If a node is neither in transmission nor reception of a packet, it is in IDLE or CCA_BUSY state. If the receiver is in one of these states when a packet with high enough power arrives (e.g., higher than -94 dBm), an SINR check for PLCP header decoding is scheduled. If the SINR of the receiving packet passes the check, the reception continues. Otherwise, it aborts. At the end of the reception of this packet, another SINR check for frame body decoding is performed to determine whether the packet is received successfully.

If a node is receiving a packet, it is in RX state. If the receiver is in this state when a new packet arrives, the decision whether to switch to the new signal is made by the frame capture policy. Instead of directly ignoring newly arrived packet, the **frame capture effect** enables a receiver to switch to a stronger signal during the reception of a weaker one. If a newly arrived packet has high enough signal strength (e.g. 4 dB greater while receiving the preamble portion of the current packet and 10 dB greater while receiving the frame body portion), a receiver can then pick and lock to the new signal. The switch is allowed in both preamble duration and frame body duration. If the capture occurs, the earlier reception is terminated immediately.

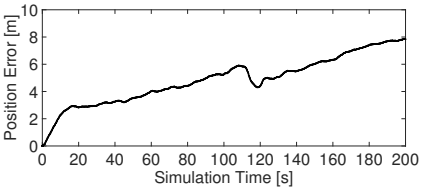


Figure 4: Cumulative position error of a moving vehicle between its position in the field experiment and in the simulation at a given time

3. SIMULATION APPROACHES

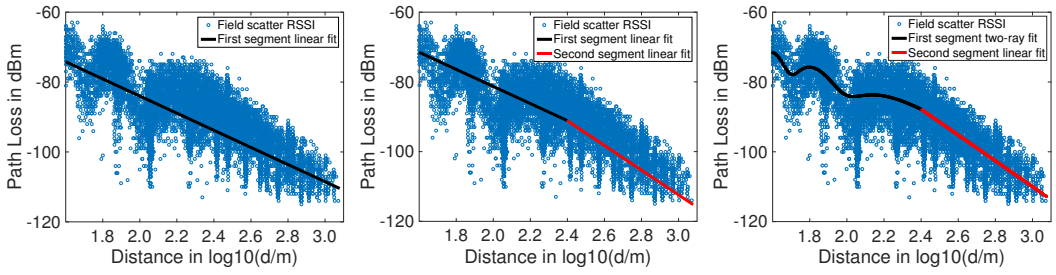
The main components of a V2V network simulator are the node mobility manager, the network traffic generator, the propagation model and the receiver model. Our calibration process involved all these four components but focused on the propagation and the receiver models. In this section, we describe our methods for calibrating these components in detail.

3.1 Mobility Manager Calibration

In the field experiments, each OBE relied on the frequent GPS readings to obtain its location and speed information. The information was further piggybacked in safety messages and shared with nearby OBEs. To accurately recreate the same mobility pattern in simulations, these GPS readings were collected and organized into a GPS trace file. This trace file was then converted to a format which can be directly processed by the ns-3 simulator. By using this approach, the movement of each node in the simulation is scheduled according to the input trace file. Thereby, the logged node position from simulations is expected to be same as that logged in the experiment. However, the default ns-3 mobility scheduler extrapolates nodes to their new positions by primarily using current speed vectors and previously extrapolated positions with an assumption that the speed and heading are constant. As shown in Fig. 4, this speed-dependent scheduling approach can introduce an accumulative error between a node's position in the simulation and its position in the input trace file. It is because the speeds in the trace file were instantaneous reading by the GPS and with instantaneous speeds as well as constant speed/heading extrapolation in the simulation, there is no guarantee that the node reaches the trace-reported position at each position update time. Such undesired error can further have a negative impact on calibrating the signal propagation. To eliminate such position error, we introduced the position-dependent scheduling along with the default speed-dependent scheduling, such that the nodes are relocated to the trace-reported position at each position update time. A slight discontinuity in the nodes trajectory can occur. However, we did not observe any adverse effect of these discontinuities on the simulation results. This two-level scheduling method improves the accuracy of the trace-driven approach and guarantees that the trace-reported positions are replayed in the simulations.

3.2 Traffic Generator Calibration

In the field experiment, the network traffic was simply periodic transmissions. However, to reproduce a similar network traffic in simulations, several simulation parameters



(a) **(b)** **(c)**

Figure 5: Different fitting models: (a) One-segment log-distance; (b) Two-segment log-distance; (c) Two-segment, two-ray for the first segment, log-distance for the second segment

including the transmission power, the transmission rate and jitter, the channel bandwidth and its center frequency, the packet length, the noise floor, the threshold for clear channel assessment (CCA) detection, the threshold for energy detection require correct configurations. A hardware lab test was conducted by the CAMP VSC3 team to clarify some of the radio characteristics of the DSRC transceivers and help to determine the correct values for these parameters. As to the packet size, it varied in the field experiment. We believe it would be highly inefficient to reproduce the exact packet size for every packet in the simulations, since the packet size is relatively small and it has a marginal impact on the system performance. Instead, we unified packet size to 316 Bytes which was the mode value of all the packet sizes.

3.3 Propagation Model Calibration

Since a propagation model consists of two main components, i.e. the distance-dependent path loss model and the fading (large-scale and small-scale), our calibration efforts primarily targeted these components. Given that the calibrated model was expected to be used in large-scale simulations, model complexity should be taken into account as well as model accuracy. Therefore, our model calibration started with commonly used low complexity models.

3.3.1 Choices for Distance-dependent Path Loss

One-segment log-distance model: Recall that the one-segment log-distance model is simple with only two fitting parameters in it, i.e. the parameter A and B, which can be obtained by minimizing the mean-square deviation of the calculated path loss to the field test data. This one-segment model was a conscious decision to avoid overfitting the RSSI measurement noise which was mostly present for longer distances since the signal strength was approaching to the noise floor. However, as the field test data included RSSI measurements for longer distances (up to 1200 m), the fitted path loss exponent (the parameter B) tended to be smaller for longer distances. It is because the RSSI samples were normally biased towards higher values at the longer distances, resulting in a gradual decrease in the path loss. Therefore, applying a single value for the path loss exponent can be problematic since it is difficult to accurately represent the path loss for the large distances.

Two-segment log-distance model: As an enhancement to the one-segment log-distance fit, a two-segment log-distance regression was considered, where a distance break point d_{br} was defined. This two-segment model allowed us to provide separate fits for shorter distances and longer distances. The shorter distances and the longer distances were identified via a distance threshold d_{br}. However, it is important to emphasize that this model does not capture local effects such as the null effect at ∼100m from the transmitter (see Fig. 5) while the two-ray model does so.

Two-ray two-segment model: From the field test data, we have observed that in some testing scenarios (e.g., highway scenario), the spacing distance between vehicles was comparatively small and thus the ray reflected by the ground was blocked by multiple metallic obstacles (i.e., car bodies) along its signal path. In these scenarios, the null effect was weak as the signal strength of the reflected ray was weak. The two-segment log-distance model would be a good fit to the data from these testing scenarios. However, in the phase 2 test, very few metal structures were present on the signal propagation path. As a result, the direct and the reflected ray can interact with opposite phase and thus the null effect was prominent. Therefore, we assumed that the two-ray model would be a good fit for the shorter distances range. For longer distances, the effect of two-ray superposition diminished and thus we kept using the log-distance model. In our calibrated model, the distance break point is 250 m.

Fig. 5 illustrates the fitting performance of the three models. The blue dots represent RSSI samples collected in the field experiments while the solid lines indicate fitted models. In general, the one-segment linear model performs worse than the two-segment model. It is mainly because that the path loss exponents are normally different for shorter and longer distances. The one-segment model is not able to capture this characteristic. Due to the null effect, the RSSI scatter points show a winding shape in the first 250 m. As mentioned above, the linear model is not sufficient to model the null effect, while the two-ray model is a better choice to follow the winding trend. After the break point, the RSSI scatter points form a dense, fairly uniform cloud whose trend is downward with increased distances. This indicates that the log-distance model is competent to model the median value of RSSI samples after the break point.

3.3.2 Large-scale Fading and Small-scale Fading

With respect to the fitted model of the distance dependent path loss, the variation of the field RSSI scatter points was captured by the fading model. In our models, it was represented as σu, where u was a zero-mean, unit variance Gaussian random variable, and σ was the standard deviation of the variation. Note that the fluctuating due to multipath

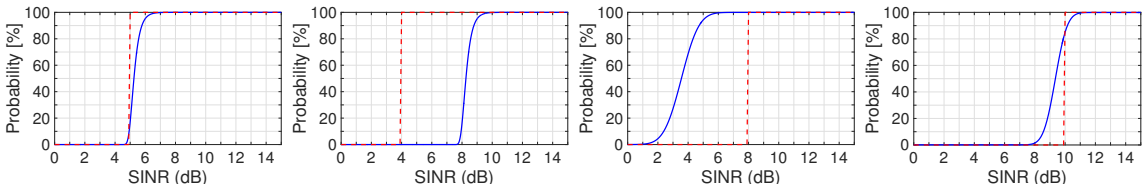


Figure 6: Fitted probability model v.s. the threshold-based model in [6, 8]: (a) preamble decoding; (b) preamble capture; (c) frame body decoding; (d) frame body capture - blue solid line for the fitted probability-based model, red dash line for the threshold-based model

was also included in each RSSI sample, and thus the variance computed from the field test data is actually $\sigma^2 = \sigma_{sh}^2 + \sigma_{mp}^2$, that is, the variances of the two kinds of fading were highly coupled. To decouple them, we first ascertained the distribution of the small-scale fading. The details of exacting the distribution have been presented in [15]. Based on the estimated σ_{mp}, the large-scale fading σ_{sh} can then be computed.

3.4 Receiver Model Calibration

CAMP VSC3 hardware lab test results have indicated that the transmitters used in the field test supported the frame capture effect. Unfortunately, This feature is not modeled in the default ns-3 simulator yet. Some previous work [6, 8] proposed a receiver modeling framework which considered the implementation of the frame capture effect. In that receiver model, a binary decision of whether a packet can be received successfully or whether the receiver can switch to a stronger signal is made by comparing the packet's SINR with specific thresholds. However, many existing works [5, 18] have identified the probabilistic nature of the packet reception. The CAMP lab test results also showed a similar probabilistic characteristic of the device's packet reception and capture. That is, between 0% and 100% successful reception, a comparatively large transition region exists. Within this region, packets are successfully received/captured with a certain probability for a given SINR value. Therefore, it is desired to model the packet reception and capture with a probability function of SINR and then create such a transition region between unsuccessful and successful receptions. We defined two probability models for the preamble reception and the frame body reception, respectively. The model for the preamble reception and capture was obtained by empirically adjusting the default ns-3 probability model for decoding the Orthogonal Frequency Division Multiplexing (OFDM) 6 Mbps signal for improving agreement of the simulation results with the field test results. The model for frame body reception and capture was derived by fitting the following equation to the CAMP hardware test results.

$$p = \frac{2a}{\sqrt{\pi}} \int_0^{(SINR-b)/c} e^{-z^2} dz + d$$

where a, b, c and d are the parameters to be fitted. A comparison between the threshold-based model and the fitted probability-based model is depicted in Fig. 6.

4. ACCURACY & LESSONS LEARNED

Next, we quantitatively analyze the simulation accuracy improvement introduced by the calibrated models and the increased model complexity. The Packet Error Ratio (PER) is selected as the primary evaluation metric because the effectiveness of V2V safety communication relies heavily on successful message exchanges. The Inter-Packet Gap (IPG) latency, which is defined as the elapsed time between two consecutive successful receptions from one particular transmitter, is introduced as the secondary metric supporting the observations from the PER evaluation. The simulation accuracy of a particular link is defined the match to the field experiment results in terms of these evaluation metrics, e.g., for PER, it is $(1-|\frac{PER_{sim}-PER_{exp}}{PER_{exp}}|)\times 100\%$, where PER_{sim} and PER_{exp} are the PER of a specific link in the simulation and in the field experiment, respectively.

4.1 Accuracy Gain of Model Calibration

To investigate the accuracy improvements introduced by the model calibration, three propagation models, i.e., one-segment log-distance, two-segment log-distance, and two-segment two-ray, were selected and tested with the default ns-3 receiver model. The default ns-3 simulator has provided the uncalibrated versions of the log-distance models. Although our calibrated two-segment two-ray model and the Nakagami model described in [8] differ somewhat in the way the median path loss is modeled, two approaches are essentially the same since they share the same distribution of the ratio of instantaneous receiver power to transmission power. Thereby, the Nakagami model in [8] was selected as the uncalibrated version of the two-segment two-ray model. Fig. 7 depicts the average simulation accuracy of all logged links in term of PER and average IPG. With the default model parameters, the three propagation models can only achieve 15%, 27%, 31% PER accuracy and 22%, 35%, 42% IPG accuracy. However, with the calibrated model parameters, the simulation accuracy improves by 25%, 21%, 28% regarding PER and 24%, 18%, 20% regarding average IPG, respectively.

The receiver model calibration follows a similar trend. The model described in [6] and the default ns-3 receiver model were selected as the uncalibrated version of the threshold-based and the probability-based approach, respectively. Note that the frame capture effect was not implemented in the default ns-3 model and to the best of our knowledge, we are not aware of any existing work describing a probability-based frame capture model which can be implemented jointly with the default ns-3 receiver model to serve as the uncalibrated version of the probability-based model. The selected receiver models were tested with our calibrated propagation model. As shown in the bottom plot of Fig. 7, with the calibrated thresholds, the threshold-based approach

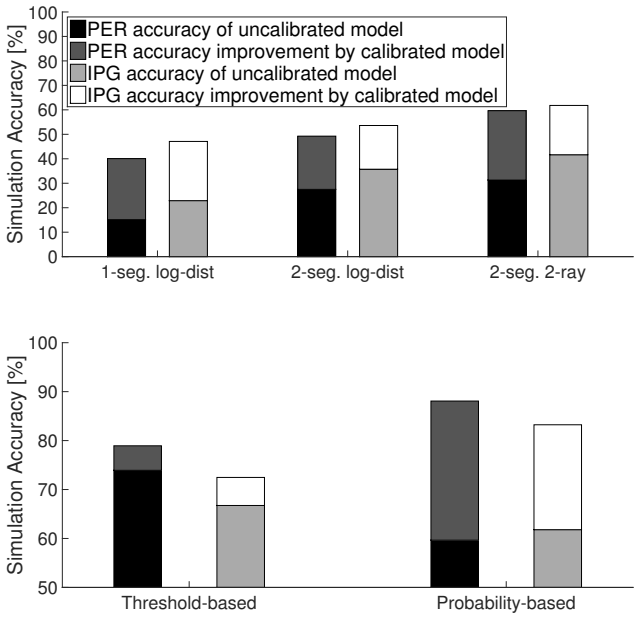


Figure 7: The improvement of simulation accuracy with different : (a) propagation models (top); (b) receiver models (bottom)

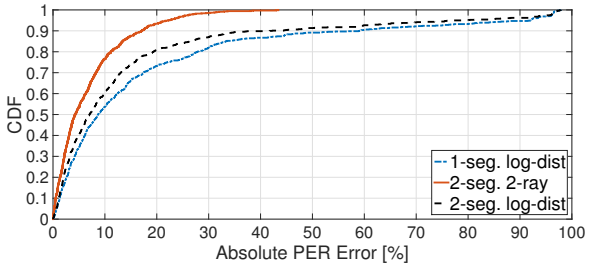


Figure 8: CDF of absolute PER errors of simulations applying different propagation models

gains 5% - 6% more accuracy in both terms of PER and IPG. As to the probability-based approach, the PER and the IPG accuracy improve by 29% and 21%, respectively, comparing to the default ns-3 receiver model.

These results show that the accuracy gain of switching to a more sophisticated model without parameter calibration is lower than the gain of calibrating the parameters of a simpler model. This underscores the need for careful model parameter choices in simulations.

4.2 Accuracy of Well-calibrated Models

Generally speaking, our calibrated propagation and receiver model were developed based on the models in existing literature. Then, according to the extracted features from the field test data, we selected the models which can largely capture these features and carefully adjust their model parameters. The results in Fig. 7 indicate that with well-calibrated models, it is possible to achieve 88% simulation accuracy regarding PER and 84% accuracy regarding IPG.

The calibration process was complicated by us having to rely primarily on the collected RSSI samples for the model calibration. Using RSSI samples in such a dense setting creates two challenges. First, the RSSI measures the total power in the full system bandwidth. Thereby, it is possible that the interference power is also included in the RSSI samples. This issue becomes much more severe when the interference power is strong. Second, the RSSI was reported only for packets sufficiently clean to be accurately received. Thus, many RSSI samples may have been lost due to weak received power or strong interference. To extract useful models from experiment RSSI data, we have developed methods for decoupling the interference from fading effects and for counteracting the effects of biased RSSI values [14].

Link-level accuracy. Fig. 8 depicts the per-link CDF of the absolute simulation errors in terms of PER obtained from the simulations where the three calibrated propagation models were tested with the calibrated probability-based receiver model. We observe that while the median errors are less than 10%, the largest error for the two-ray model is $\sim$ 43%, but over 98% for the log-distance models. This reflects the stochastic nature of the propagation model: links are assigned random fading values that match the distribution observed in the field-test but not necessarily the exact fading characteristics of a link. This primarily affects links with weak signals near the reception threshold. For example, for links with only a small subset of packets received successfully in the field experiment, the log-distance models fail to reproduce these successful receptions in the simulation and wrongly treat these weak links as broken links, then resulting in larger errors. We also notice that for the two-ray model, the simulation error of about 78% links is less than 10%, but only 54% and 60% links in the one-segment and the two-segment log-distance simulations can achieve the same level of accuracy.

Two aspects of complexity were discussed in the receiver modeling: 1) with v.s. without the frame capture implementation; 2) the threshold-based v.s. the probability-based approach. Since the frame capture effect is expected to primarily benefit these links with higher signal strength, the performance of links with length less than 100 m were investigated in three types of simulations: the calibrated threshold-based model without and with the frame capture effect, and the calibrated probability-based model. As shown in Fig. 9, the simulation accuracy are 2%, 76%, and 87%, respectively. These results indicate that the implementation of the frame capture significantly affects the accuracy given that this feature has already been supported by the field test devices. Completely without modeling this feature in the simulations can potentially lead to a low accuracy. Further, if the probabilistic nature of packet reception and capture is modeled, the simulation accuracy can be improved by 11%.

Runtime tradeoffs. More detailed models can lead to higher computational loads and increased simulation runtime. The runtimes for simulating 400 nodes for 260 seconds using these aforementioned models on a machine with a 2.4 GHz Intel Xeon E7 CPU is listed in Table 1. When using the two-ray model, the runtime increases by 8.3% comparing to the log-distance model, due to the more complex propagation equation. Further, in the two-ray simulation, more nodes may receive interference by a transmission due the changed propagation range. These additional nodes then start the receiving procedure which also increases the computational load. Second, for the calibrated probability-based

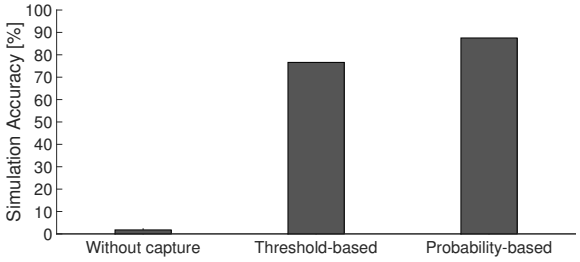


Figure 9: PER accuracy of links with link distance less than 100 m

model, the runtime increases by 8.7% comparing to the default ns-3 model. It is because both the threshold-based and the probability-based model introduce additional procedures for handling the frame capture and the probability-based model uses a more complex implementation. We find these tradeoffs acceptable for simulations with several hundreds even thousands of nodes given the improvements in simulation accuracy.

Table 1: Runtime comparison of simulation models

Models	Runtime
One-segment log-distance, default receiver	48 min
Two-segment log-distance, default receiver	48 min
Two-segment two-ray, default receiver	52 min
Threshold-based, two-ray propagation	54 min
Probability-based, two-ray propagation	57 min

4.3 Cross-validation of Simulation and Field Experiment

The calibration process is helpful in identifying errors in simulations. While examining the behaviors of the MAC layer in the default ns-3 simulator, we have identified that the implementation of the backoff timer countdown in the Enhanced Distributed Channel Access (EDCA) function does not follow the descriptions in the IEEE 802.11 standard [3]. This incorrect implementation can enlarge the queuing time of each packet in the MAC layer and increase the mismatch to the field experiments. We have shared these corrections with the ns-3 maintainers.

The process of joint experiments and simulations also helps in validating, planning, and predicting the field experiment results. Once a level of credibility of the calibrated simulator is established, we found the simulator useful in planning additional experimental tests. As an example, recall that in some field experiments, the OBEs were configured to transmit at higher rates (up to 50 Hz instead of 10 Hz) in order to emulate more nodes that were actually available (emulating 2000 transmitters on the channel). There was uncertainty whether this emulation method is valid, which was greatly reduced by simulating 2000 vehicles and comparing the results simulations of the emulation as well as with the field test. We found that up to 30 Hz the emulation method was accurate but 40 Hz and higher queuing issues compromised its accuracy. Moreover, with a calibrated simulator, one can predict field test results before the test is conducted and study possible issues that may be encountered. Simulations can also be used to explore a large number of mobility and propagation scenarios to identify candidates for future experimental study.

5. DISCUSSION

In this work, the simulation models were calibrated for the particular field experiments. However, we believe these calibrated models can be to some extent generalized to other scenarios.

The field experiments were conducted in open-space environments with few large obstacles. Such environments are commonly seen in many rural or highway communication scenarios. We expect that our calibrated propagation model for the field experiment is still suitable for other similar open-space environments. Moreover, as briefly mentioned in Section 3.3, while calibrating the propagation model, we have developed a set of techniques to process the collected RSSI data for model fitting, e.g., decoupling the interference from the fading effect in the collected RSSI samples, counteracting the effect of biased RSSI samples. We believe these techniques are also useful for fitting experiment RSSI data to statistical models in other propagation environments.

The packet recovery and capture parameters are not defined in the IEEE 802.11 standards, instead, they are radio chipset dependent. In theory, receivers should perform similarly if their chipsets are from the same vendor and of the same model. Our receiver model is calibrated for an Atheros AR5414 chipset. We believe such chipsets have been used in many other WiFi devices, which means that our model should match these devices' receiver behaviors as well.

6. CONCLUSIONS

In this paper, we share our experience in evaluating a dense V2V network through joint large-scale field tests with hundreds of nodes and simulations and then reflect on the state of network simulation. We have learned that: the default propagation and receiver models were not able to produce a good match to the field experiment results despite the relatively straightforward open-space simulation environment. However, with well-calibrated parameters and models from the literature, the simulation accuracy improves by about 20% and the stochastic simulation achieved 88% accuracy in term of PER in this scenario. Such accuracy can be achieved with existing models from the literature and at a relatively small additional runtime overhead 8.7%, although we did not find implementations of these models readily available in the ns-3 simulator. We have shared our models and corrections with the ns-3 maintainers to facilitate wider availability[2].

7. ACKNOWLEDGMENTS

We thank our shepherd and reviewers for their insightful comments. We would also like to acknowledge CAMP VSC3 Consortium for providing the field test data and supporting this work.

8. REFERENCES

[1] Safety Pilot websit, University of Michigan Transportation Research Center, Ann Arbor, Michigan. http://www.safetypilot.us/.

[2]The ns-3.16 implementation of our models are available in http://www.winlab.rutgers.edu/ gruteser/projects/patch/patch-list.html.

[2] U.S. Department of Transportation Announces Decision to Move Forward with Vehicle-to-Vehicle Communication Technology for Light Vehicles. http://www.nhtsa.gov/About+NHTSA/Press+Releases/2014/USDOT+to+Move+Forward+with+Vehicle-to-Vehicle+Communication+Technology+for+Light+Vehicles.

[3] IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *IEEE Std 802.11-2012 (Revision of IEEE Std 802.11-2007)*, pages 1–2793, March 2012.

[4] N. Baldo, M. Requena-Esteso, J. Núñez Martínez, M. Portolès-Comeras, J. Nin-Guerrero, P. Dini, and J. Mangues-Bafalluy. Validation of the ieee 802.11 mac model in the ns3 simulator using the extreme testbed. In *Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques*, SIMUTools '10, pages 64:1–64:9, 2010.

[5] E. Ben Hamida, G. Chelius, and J. M. Gorce. Impact of the physical layer modeling on the accuracy and scalability of wireless network simulation. *Simulation*, 85(9):574–588, Sept. 2009.

[6] T. Bingmann. Accuracy enhancements of the 802.11 model and edca qos extensions in ns-3. Master's thesis, University of Karlsruhe, 2009.

[7] M. Boban, J. Barros, and O. Tonguz. Geometry-based vehicle-to-vehicle channel modeling for large-scale simulation. *Vehicular Technology, IEEE Transactions on*, 63(9):4146–4164, Nov 2014.

[8] Q. Chen, F. Schmidt-Eisenlohr, D. Jiang, M. Torrent-Moreno, L. Delgrossi, and H. Hartenstein. Overhaul of ieee 802.11 modeling and simulation in ns-2. In *Proceedings of the 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, MSWiM '07, pages 159–168. ACM, 2007.

[9] V. Erceg, L. J. Greenstein, S. Y. Tjandra, S. R. Parkoff, A. Gupta, B. Kulic, A. A. Julius, and R. Bianchi. An empirically based path loss model for wireless channels in suburban environments. *IEEE Journal on Selected Areas in Communications*, 17(7):1205–1211, Jul 1999.

[10] S. Ivanov, A. Herms, and G. Lukas. Experimental validation of the ns-2 wireless model using simulation, emulation, and real network. In *Communication in Distributed Systems (KiVS), 2007 ITG-GI Conference*, pages 1–12, Feb 2007.

[11] S. Jansen and A. McGregor. Performance, validation and testing with the network simulation cradle. In *Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, 2006. MASCOTS 2006. 14th IEEE International Symposium on*, pages 355–362, Sept 2006.

[12] D. B. Johnson. Validation of wireless and mobile network models and simulation. In *DARPA/NIST Network Simulation Validation Workshop*, 1999.

[13] J. Karedal, N. Czink, A. Paier, F. Tufvesson, and A. Molisch. Path loss modeling for vehicle-to-vehicle communications. *Vehicular Technology, IEEE Transactions on*, 60(1):323–328, Jan 2011.

[14] S. Kokalj-Filipovic and L. Greenstein. EM-Based Channel Estimation from Crowd-Sourced RSSI Samples Corrupted by Noise and Interference. In *49th Conf. on Information Systems and Sciences (CISS)*, Mar. 2015.

[15] S. Kokalj-Filipovic, L. J. Greenstein, B. Cheng, and M. Gruteser. Methods for extracting v2v propagation models from imperfect rssi field data. In *Vehicular Technology Conference (VTC Fall), 2015 IEEE 82nd*, Sept. 2015.

[16] NHTSA. Federal motor vehicle safety standards: Vehicle-to-Vehicle (V2V) communications advance notice of proposed rulemaking, August 15, 2014.

[17] S. Papanastasiou, J. Mittag, E. Strom, and H. Hartenstein. Bridging the gap between physical layer emulation and network simulation. In *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*, pages 1–6, April 2010.

[18] G. Pei and T. Henderson. Validation of ofdm error rate model in ns-3 simulation. *Boeing Research Technology*, pages 1–15, 2010.

[19] T. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition, 2001.

[20] I. Sen and D. Matolak. Vehicle-vehicle channel models for the 5-ghz band. *Intelligent Transportation Systems, IEEE Transactions on*, 9(2):235–245, June 2008.

[21] C. Sommer, D. Eckhoff, R. German, and F. Dressler. A computationally inexpensive empirical model of ieee 802.11p radio shadowing in urban environments. In *Wireless On-Demand Network Systems and Services (WONS), 2011 Eighth International Conference on*, pages 84–90, Jan 2011.

[22] Q. Sun, S. Tan, and K. Teh. Analytical formulae for path loss prediction in urban street grid microcellular environments. *Vehicular Technology, IEEE Transactions on*, 54(4):1251–1258, July 2005.

[23] C.-X. Wang, X. Cheng, and D. Laurenson. Vehicle-to-vehicle channel modeling and measurements: recent advances and future challenges. *Communications Magazine, IEEE*, 47(11):96–103, 2009.