Low Complexity SNR-Based Packet Level Burst-Error Model for Vehicular Ad-Hoc Networks

Thomas Blazek
Institute of Telecommunications, TU Wien
Vienna, Austria
tblazek@nt.tuwien.ac.at

Christoph F. Mecklenbräuker
Institute of Telecommunications, TU Wien
Vienna, Austria
cfm@nt.tuwien.ac.at

***Abstract*—Network simulators are a crucial tool for evaluating the performance of Vehicular Ad-Hoc Network (VANET) protocols. They allow the assessment of the scalability and the influence of geometric topologies. However, typical network simulators such as NS-3 or OMNET++ often resort to overly simplified packet error models, limiting the validity of the results. Measurements have shown that burst-error patterns are characteristic to VANETs, which are not modeled using the common approaches. In this contribution, we develop a low-complexity burst error model that is parameterized by the Signal-to-Noise Ratio (SNR). We adapt an approach based on the Gilbert-Elliot Markov model, which allows to model first order burst properties while retaining low complexity. Furthermore, we define three performance indicators, overall error probability as well as burst error and burst success probability, and use a multi-feature information bottleneck to find the optimal SNR quantization in the mutual information sense. Based on this, we present Gilbert-Elliot model fits that are easily implementable and demonstrate that low-level SNR quantizations of 4-9 intervals are sufficient to capture the statistics in the MSE sense.**

***Index Terms*—Intelligent Transportation Systems, Network Simulations, VANETs**

I. Introduction

The preferred tool for assessing the performance of Vehicular Ad-Hoc Networks (VANETs) remains the simulation at network layer. These simulations allow to evaluate large vehicular networks, incorporating real life city maps, mobility models such as SUMO [1]. Among the most popular tools for this purpose are NS-3 [2], [3] and OMNET++ [4], [5]. These tools generally compute a Received Signal Strength Indicator (RSSI) after path loss, shadowing and small scale fading, and then employ a mapping of RSSI to packet loss. Often, overly simplistic mappings of RSSI to packet error statistics are used, e.g. the all-or-nothing approximation depending on a constant RSSI threshold [6]. Alternatively, the packet error statistics are modeled as Bernoulli trials with either a constant packet error probability or a packet error probability which is a function of the instantaneous RSSI at the receiver [7]. This static mapping however does not take into account the fact that the communication performance of Vehicle-to-Vehicle (V2V)-channels was shown to exhibit bursty packet error patterns in many scenarios [8], [9]. Meanwhile, only few attempts have been made to capture more complex channels in such simulations [10], [11].

In 1960, Gilbert introduced an extension that significantly improves capturing the burstiness of the packet errors, the so-called Gilbert Model [12], a Hidden Markov Model (HMM) that models bursts of high error rates alternating with error-free bursts. In order to capture the location dependent channel conditions, such a model is used in [13] and the parameters are allowed to change with the distance between a vehicle and a Roadside Unit (RSU) mounted at an elevated position. This worked for the authors as both pathloss and position of static scatterers are directly determined by the distance. This however is not the case in V2V, since low transmitters and interferers mean that the LOS condition dominates the pathloss, and there are no static scatterers. On the other hand, the V2V scenario is approached in [14]. There, instead of the distance, the Signal-to-Noise Ratio (SNR) is used as handle for the model parameters, and the Gilbert-Model is extended to a Gilbert-Elliot model [15]. Using the SNR as estimation basis allows for better modeling of effects like Line-of-Sight (LOS) – Non Line-of-Sight (NLOS) distinction, and easier integration into network simulators. However, the authors calculate new model parameters for every 1 dB step, resulting in more than 25 sets of 4 model parameters. Furthermore, the a priori binning into 1 dB steps may be suboptimal. In [16], the authors demonstrated that a much lower number of parameter sets may still lead to satisfactory modeling performance. The results, again, are only valid for V2I scenarios, and are based on a vector quantization technique that does not consider the stochastic properties of the modeled quantities.

In this paper, we refine the approach given in [14] on two accounts. We solve the problem without requiring prior parameter quantization, and find the minimal required number of SNR-intervals that obtains the optimal performance in the Mean Squared Error (MSE) sense. Furthermore, we present a method to find the optimal intervals while considering multiple optimization goals.

The structure of the paper is as follows. In Section II, we present the stochastic modeling approach, as well as the measurements we want to replicate. In Section III, we first present the first- and second-order statistics that our model has to replicate as close as possible. We furthermore introduce a

This work was partially funded by Austrian Development Cooperation and Ministry of Education, Science and Technology of Republic of Kosovo, in the framework of the project HERAS K-01-2017 as research cooperation and networking between Austria, Kosovo and the Western Balkan Region.

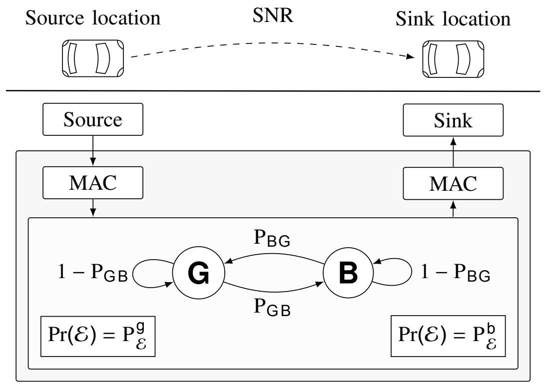


Fig. 1. Proposed system model, with the MAC transmission modeled as Gilbert-Elliot model.

multi-feature version of the agglomerative information bottleneck algorithm [17], which allows us to find a quantization of the SNR range that maximises the fidelity of the given first- and second-order statistics. This is done by maximising the mutual information between the original data and the quantized data. Finally, in Section IV, we use these intervals that minimize the information loss, and apply the Baum-Welch algorithm, to find Markov Model parameters which minimize the MSE of our given statistics. We use the MSE to evaluate the performance of the quantization.

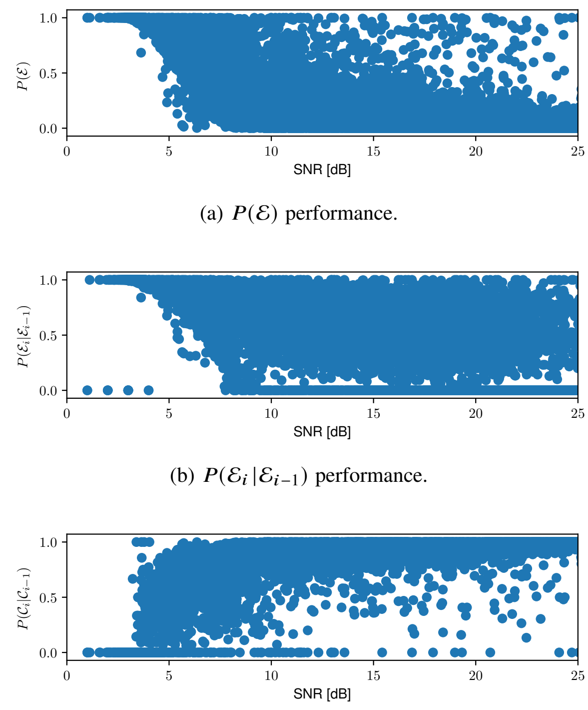
II. System Model

A. Packet-Level Modeling

Figure 1 illustrates the proposed approach of system level modeling. The communication link from one vehicle to the next is established over Medium Access Control (MAC) and Physical Layer (PHY), and the communication channel. At the system level, this is condensed into the probability of an error occurring in the communication $P(\mathcal{E})$. This probability may be time variant, and depend on prior events. Hence, we also introduce the probability of an error occurring given the prior event was an error $P(\mathcal{E}_i|\mathcal{E}_{i-1})$, as well as the probability of a correct transmission occurring given the prior transmission was correct $P(C_i|C_{i-1})$.

B. Gilbert-Elliot-Model

The Gilbert-Elliot model [15] uses a 2 state Markov chain and 4 parameters to model bursty packet traffic.The model is always either in *good* (G) or *bad* (B) state (Figure 1). One Bernoulli trial is executed on whether a state transition happens, with the transition probabilities P_{GB} from good to bad state and P_{BG} from bad to good state. After that trial, a second test evaluates whether the current packet is lost, according to $P^g_{\mathcal{E}}$ in good and $P^b_{\mathcal{E}}$ in bad state, with $P^b_{\mathcal{E}} > P^g_{\mathcal{E}}$. This allows to capture alternating bursts of low and high packet errors. Our goal is to find SNR intervals that we assign a set of model parameters. Then, we evaluate the current SNR, and choose the model parameters accordingly. This approach only requires 2 Bernoulli trials, and thus has negligible performance impact compared to the usual approach of static Packet Error Rates (PERs). However, it allows to additionally capture $P(\mathcal{E}_i|\mathcal{E}_{i-1})$ and $P(C_i|C_{i-1})$ accurately. In [14], a separate model quadruple was calculated for every 1 dB step of the SNR range. In the next section, we will instead use an algorithm to determine the optimal bounds on the SNR range to minimize the number of required parameters.



(a) $P(\mathcal{E})$ performance.

(b) $P(\mathcal{E}_i|\mathcal{E}_{i-1})$ performance.

(c) $P(C_i|C_{i-1})$ performance.

Fig. 2. Scatter plots of the performance indicators.

III. Choosing the Optimal SNR Intervals

We base our analysis on the measurements conducted in the ROADSAFE campaign [18], which recorded 10771 s of V2V highway traffic. Both LOS and NLOS scenarios were included in the measurements, and packet traffic was recorded along with SNR estimates, noise power estimates, time stamps and Global Positioning System (GPS) information and packet decoding successes using the IEEE 802.11p standard.

A. Packet-Level Measurements

TABLE I
Measurement Parameters

Bitrate	6 Mb/s
Modulation/Coding	Q-PSK, Rate 1/2
Packet length	500 Bytes
Mean packet rate	970 s^{-1}
Transmit power	10 dBm

The transmission parameters used in the measurement can be seen in Table I. The packet transmission rate was chosen at the given value to ensure a fine-grained analysis of the channel. This is subsequently the rate at which we allow state

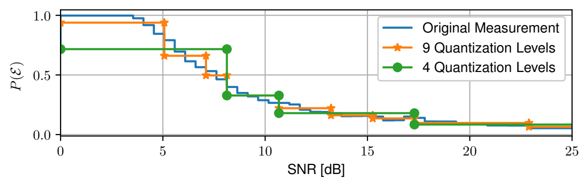


Fig. 3. $P(\mathcal{E})$ estimation for SNR quantization into 4 and 9 intervals.

transitions to happen, and hence has to be orders of magnitude higher than the standard-compliant transmissions.We group the measurement in 1 s intervals, to ensure that we capture the large scale nonstationarity of the channel, while the small-scale time-variance is expected to be captured in the Gilbert-Elliot model [14]. Figure 2 shows scatter plots for 1 second estimates of $P(\mathcal{E})$, $P(\mathcal{E}_i|\mathcal{E}_{i-1})$ and $P(C_i|C_{i-1})$ as a function of the SNR. These figures suggest that there exists a critical region between 5 and 15 dB SNR where the distribution changes rapidly with the SNR value, while outside of this region the behavior is relatively slow changing. In the next section we will introduce an algorithm that considers these three performance measures as *relevance* variables, and quantizes the SNR region in such a way that rapid changes are mapped with small intervals, and slow changes are assigned larger intervals.

B. Multi-Feature Information Bottleneck

Our goal is to classify the given SNR measurements $\gamma \in \Gamma$ into Q SNR intervals $\tilde{\gamma} \in \tilde{\Gamma}$ in such a way that our performance measures are optimally captured. In the case of the packet error rate this means that the empirical probability density functions (pdfs) of our estimated probabilities given the quantized SNR intervals $p(P(\mathcal{E})|\tilde{\gamma})$ should be as close as possible to the original conditionals $p(P(\mathcal{E})|\gamma)$, which is equivalent to demanding that the mutual information $I(P(\mathcal{E});\tilde{\Gamma})$ should be maximal for a given quantization $\tilde{\Gamma}$. Figure 3 illustrates this by showing the median $P(\mathcal{E})$ values as was used in [14], and comparing it to the same curve where the SNR has been quantized into 4 and 9 regions. The figure shows that the quantization will choose smaller intervals for regions with more change in $P(\mathcal{E})$-behavior. Finding the optimal quantization while maximizing this mutual information is called the information bottleneck [19] and is described via Lagrangian multipliers as

$$\tilde{\Gamma}_{\text{opt}} = \arg\min_{\tilde{\Gamma}} \mathcal{L}[p(\tilde{\gamma}|\gamma)] = \arg\min_{\tilde{\Gamma}} I(\Gamma;\tilde{\Gamma}) - \beta I(\tilde{\Gamma}; P(\mathcal{E})). \quad (1)$$

Here, $p(\tilde{\gamma}|\gamma)$ is the conditional pdf. Slonim *et al.* presented the *agglomerative bottleneck*, which is a greedy algorithm that starts with N partitions and sequentially merges two partitions until only Q partitions are left [17]. This is done in a locally optimal manner that ensures that at every merge step, the currently optimal 2 regions are merged. The key to this approach is the fact that the merge of the SNR intervals γ_i and γ_j reduces $I(\tilde{\Gamma}; P(\mathcal{E}))$ by

$$d_{i,j}(P(\mathcal{E})) = (p(\gamma_i) + p(\gamma_j)) JS_{\Pi_2}[p(P(\mathcal{E})|\gamma_i), p(P(\mathcal{E})|\gamma_j)], \quad (2)$$

where JS_{Π_2} is the Jensen-Shannon divergence resulting from merging the two pdfs [20]. The Jensen-Shannon divergence is a measure for the similarity of two distributions that extends the concepts of the Kullback-Leibler divergence [21].

However, our goal is to choose the intervals in such a way that we retain the optimal modeling of the $P(\mathcal{E})$ as well as the burst probabilities. Hence, we want to consider all three variables $P(\mathcal{E})$, $P(\mathcal{E}_i|\mathcal{E}_{i-1})$ and $P(C_i|C_{i-1})$ in the optimization. We do this analogous to [22] by introducing a new objective function, the multi-feature (m-f) objective function

$$\mathcal{L}_{m-f} = I(\Gamma;\tilde{\Gamma}) - \beta \underbrace{\sum_{k=1}^{K} \lambda_k I(Y^k;\tilde{\Gamma})}_{T(\mathbf{Y},\lambda,\Gamma)}. \quad (3)$$

Here, we consider K relevant variables, Y^k represents the kth *relevance variable*, and λ_k is a weighting factor for the kth variable with $\sum \lambda = 1$. In our case $K = 3$, $Y^1 = P(\mathcal{E})$, $Y^2 = P(\mathcal{E}_i|\mathcal{E}_{i-1})$ and $Y^3 = P(C_i|C_{i-1})$. Therefore, $T(\mathbf{Y},\lambda,\Gamma)$ represents a weighted average mutual information between the quantized SNR and the different relevance variables. For now we do not specify the weighting factors λ.

From the original agglomerative bottleneck, we know that $I(Y^k;\tilde{\Gamma})$ is reduced by $d_{i,j}(Y^k)$ if the clusters i and j are merged. As seen in Equation (3), the single mutual information terms do not depend on each other. Thus, when merging the ith and jth SNR interval, the function $T(\mathbf{Y},\lambda,\Gamma)$ is in total reduced by

$$D_{i,j}(\mathbf{Y},\lambda) = \sum_{k=1}^{K} \lambda_k d_{i,j}(Y^k) \quad (4)$$

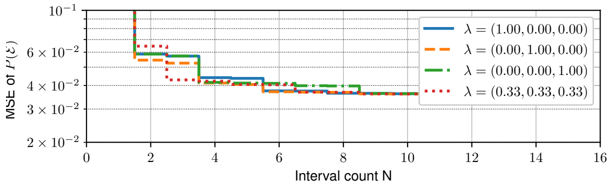
We apply the regular agglomerative information bottleneck algorithm to optimize (Equation (3)), by optimizing with respect to $D_{i,j}$ instead of $d_{i,j}$. Of course, the quantization will depend on the choice of the weighting coefficients λ_k.

C. Quantization Performance

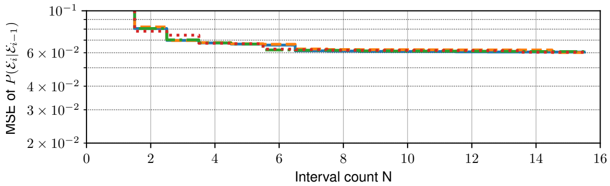
The results of the quantization are shown in terms of MSE in Figure 4. We compare 4 different quantization strategies. $\lambda = (1,0,0)$, which corresponds to only considering $P(\mathcal{E})$, and analogous approaches for $P(\mathcal{E}_i|\mathcal{E}_{i-1})$ ($\lambda = (0,1,0)$) and $P(C_i|C_{i-1})$ ($\lambda = (0,0,1)$), as well as an equally weighted approach ($\lambda = (0.33, 0.33, 0.33)$). The results show, that the MSE performance of all relevance parameters is almost equivalent, regardless of the optimization target. Furthermore, the MSE achieves very good values at $Q = 4$, and reaches it's saturation values around $Q = 9$, suggesting that more than 9 intervals result in over-fitting that do not improve the estimation quality anymore.

IV. Model Parameter Estimation

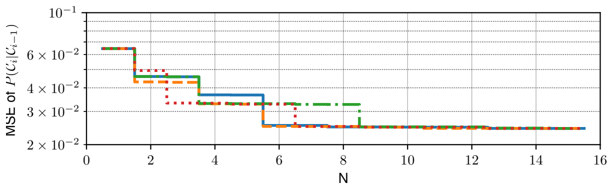
We now take the interval bounds found in the previous section, and use them to calculate Gilbert-Elliot model parameters. Analogous to [14], we split the measurements in 1 s intervals, and group all time steps whose mean SNR lies in the same SNR interval. We then use the Baum-Welch



(a) $P(\mathcal{E})$ performance.



(b) $P(\mathcal{E}_i|\mathcal{E}_{i-1})$ performance.



(c) $P(C_i|C_{i-1})$ performance.

Fig. 4. Multi-feature bottleneck estimation performance for different choices of λ.

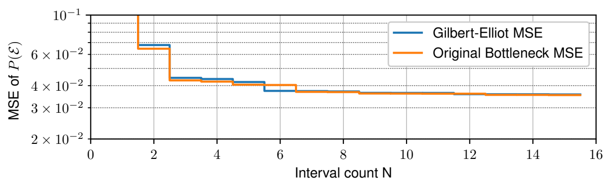
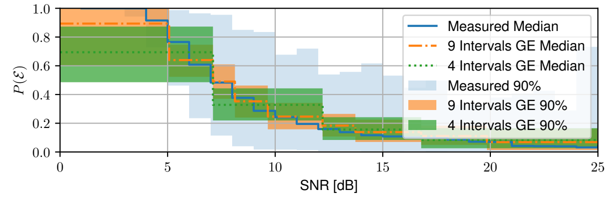


Fig. 5. MSE of $P(\mathcal{E})$ for the Gilbert-Elliot model compared with the original MSE of the bottleneck with N intervals.

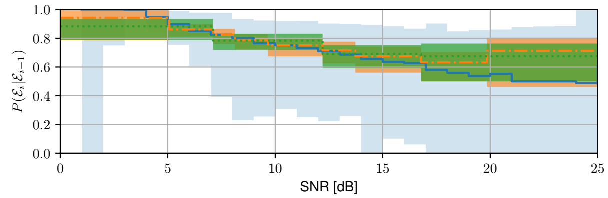
algorithm (see [23]) to derive the optimal parameter tuples $(P_{BG}, P_e^g, P_{GB}, P_e^b)$ from the packet measurements for each interval. We do this for 1 interval up to 15 intervals. Figure 5 depicts the resulting MSE of Gilbert-Elliot based estimated $P(\mathcal{E})$ for different numbers of intervals. As shown, the Gilbert-Elliot estimates behave almost identical to the direct results of the information bottleneck. We see that a flattening of the performance improvement 4 intervals, and 9 intervals marks the final error floor with almost no change beyond 9 intervals. Thus, we will consider those two selections as low-complexity and high-quality quantizations respectively.

A. Estimation Results for selected Interval Counts

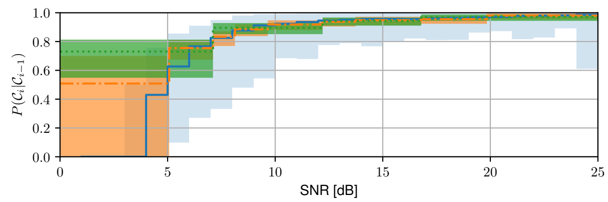
The exact Gilbert-Elliot model parameters for 4 and 9 intervals, as well as the accompanying SNR interval boundaries can be seen in Table II. The table illustrates that the transition probabilities never go above 0.1, meaning that both states are *sticky*, i.e. once the model enters the state, it will likely remain there for some iterations. This is combined with the fact that the error probability in bad state is uniformly high throughout the SNR range. Conversely, the error probability in



(a) $P(\mathcal{E})$



(b) $P(\mathcal{E}_i|\mathcal{E}_{i-1})$



(c) $P(C_i|C_{i-1})$

Fig. 6. Performance indicators for 9-interval Gilbert-Elliot model. The plots show the median value and the range where the central 90 % of events lie.

good state goes rapidly to 0, and is almost negligible for SNR above 10 dB. Figure 6 shows our key performance indicators as a function of the SNR. The plot shows the median and the range where 90 % of all values lie both for the original measurements, as well as the 9-interval Gilbert-Elliot model. In this case, the Gilbert-Elliot model has a much smaller spread than the original measurements. However, the median values are replicated very well. For 9 intervals, the results show that all measurements below 5 dB are grouped in one interval, and all measurements above 21 dB are also grouped in a single interval. In the transition region in between, the interval width is adapted to the observed parameter change. For 4 intervals, the same trend is seen with even larger border intervals.

Furthermore, the burst behavior ($P(\mathcal{E}_i|\mathcal{E}_{i-1})$ and $P(C_i|C_{i-1})$) is captured very well through the Gilbert-Elliot model. Deviations only occur at the border intervals, since the algorithm resulted in large interval sizes there. At the high SNR-end of the error burst curve, the interval stretches up to infinity. In the given measurement, the receiver would sometimes receive no packets while reporting abnormally high SNR values. However, these artefacts only influence the error burst statistic, which is based on a very small number of samples, due to errors almost never occurring at that SNR levels.

V. Conclusions

We propose and evaluate the modeling of statistical dependencies in time series of packet errors based on the Gilbert-

TABLE II
PARAMETERS FOR 4 AND 9 INTERVAL QUANTIZATION

Q	Nr.	Start SNR	End SNR	p_{BG}	P_e^g	p_{GB}	P_e^b
4	1	$-\infty$	7.108	0.026	0.151	0.055	0.957
	2	7.108	12.203	0.051	0.030	0.028	0.876
	3	12.203	16.786	0.065	0.011	0.015	0.783
	4	16.786	∞	0.054	0.005	0.006	0.759
9	1	$-\infty$	5.072	0.014	0.337	0.080	0.980
	2	5.072	7.108	0.033	0.129	0.054	0.945
	3	7.108	8.129	0.041	0.076	0.039	0.921
	4	8.129	9.657	0.049	0.034	0.030	0.875
	5	9.657	12.203	0.062	0.019	0.023	0.845
	6	12.203	13.728	0.065	0.013	0.017	0.810
	7	13.728	16.786	0.065	0.010	0.013	0.763
	8	16.786	19.841	0.070	0.008	0.012	0.730
	9	19.841	∞	0.043	0.004	0.004	0.793

Elliot model, whose parameters are piece-wise constants versus SNR. By, using the multi-feature agglomerative information bottleneck, we ensure that we do not use more parameters in the model than are required for accurate representations. Additionally, the agglomerative bottleneck ensures that the interval spread is well adjusted. That is, critical SNR regions are quantized finely, while SNR regions with little change are clustered in large intervals.

For safety critical applications, timely delivery is of high importance, and thus the burst characteristics of the channel have large impacts. Our results show clearly that the ad-hoc communication of our measurement indeed behaves bursty, and neglecting this burstiness in the simulations will introduce errors. We also demonstrated that the exact behavior can be split into a small number of categories. Depending on the requirements of the setting, 4 to 9 intervals are sufficient to incorporate the average as well as the burst behavior seen in the measurements. Extending network simulators to implement the given model is straight-forward, as it only relies on knowledge of the SNR, which is already available, and then introduces almost no additional computational overhead. Thus, the presented model strikes a good balance between ease of implementation, and exactness. This effect is amplified by the fact that using the agglomerative information bottleneck, the number of intervals becomes a design parameter. Furthermore, the presented approach can readily be applied to new protocols based on new measurements, allowing for packet-error models in network simulators to adapt more easily to new protocols without resorting to overly simplified models.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Bujar Krasniqi and Dipl.-Ing Taulant Berisha for their contributions.

REFERENCES

[1] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent Development and Applications of SUMO – Simulation of Urban MObility," *Int. J. Adv. Syst. Meas.*, vol. 4, 2012. [Online]. Available: http://www.iariajournals.org/systems_and_measurements/

[2] J. Mittag, S. Papanastasiou, H. Hartenstein, and E. Ström, "Enabling Accurate Cross-Layer PHY/MAC/NET Simulation Studies of Vehicular Communication Networks," *Proc. IEEE*, vol. 99, no. 7, pp. 1311–1326, Jul 2011.

[3] H. Boeglen, B. Hilt, P. Lorenz, J. Ledy, A.-M. Poussard, and R. Vauzelle, "A survey of V2V channel modeling for VANET simulations," in *Int. Conf. Wirel. On-Demand Netw. Syst. Serv.* IEEE, Jan 2011, pp. 117–123.

[4] M. Baguena, C. T. Calafate, J. Cano, and P. Manzoni, "Towards realistic vehicular network simulation models," in *Proc. IFIP Wirel. Days*, Nov 2012, pp. 1–3.

[5] S. Eichler, "Performance Evaluation of the IEEE 802.11p WAVE Communication Standard," in *Veh. Technol. Conf.* IEEE, Sep 2007, pp. 2199–2203.

[6] A. Rakhshan and H. Pishro-Nik, "Packet success probability derivation in a vehicular ad hoc network for a highway scenario," in *Annu. Conf. Inf. Sci. Syst.* Princeton, NJ: IEEE, Mar 2016, pp. 210–215.

[7] F. Abrate, A. Vesco, and R. Scopigno, "An Analytical Packet Error Rate Model for WAVE Receivers," in *Veh. Technol. Conf. (VTC Fall), 2011 IEEE*, Sep 2011, pp. 1–5.

[8] K. Karlsson, C. Bergenhem, and E. Hedin, "Field Measurements of IEEE 802.11p Communication in NLOS Environments for a Platooning Application," in *Veh. Technol. Conf. (VTC Fall).* IEEE, Sep 2012, pp. 1–5.

[9] T. Blazek, M. Ashury, C. Mecklenbräuker, D. Smely, and G. Ghiaasi, "Vehicular channel models: A system level performance analysis of tapped delay line models," in *Proc. Int. Conf. ITS Telecommun.*, 2017.

[10] C. Nelson, N. Lyamin, A. Vinel, C. Gustafson, and F. Tufvesson, "Geometry based channel models with cross- and autocorrelation for vehicular network simulations," in *In Proceedings VTC Spring*, 2018, pp. 1–5.

[11] T. Blazek, G. Ghiaasi, C. Backfrieder, G. Ostermayer, and C. Mecklenbräuker, "IEEE 802.11p Performance for Vehicle-to-Anything Connectivity in Urban Interference Channels," in *12th European Conference on Antennas and Propagation (EuCAP)*, 2018, pp. 1–5.

[12] E. N. Gilbert, "Capacity of a Burst-Noise Channel," *Bell Syst. Tech. J.*, vol. 39, no. 5, pp. 1253–1265, 1960. [Online]. Available: http://dx.doi.org/10.1002/j.1538-7305.1960.tb03959.x

[13] V. Shivaldova, A. Winkelbauer, and C. F. Mecklenbräuker, "Vehicular Link Performance: From Real-World Experiments to Reliability Models and Performance Analysis," *IEEE Veh. Technol. Mag.*, vol. 8, no. 4, pp. 35–44, Dec 2013.

[14] T. Blazek and C. Mecklenbräuker, "Measurement-based burst-error performance modeling for cooperative intelligent transport systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, pp. 1–10, Mar 2018. [Online]. Available: https://doi.org/10.1109/TITS.2018.2803266

[15] E. O. Elliott, "Estimates of Error Rates for Codes on Burst-Noise Channels," *Bell Syst. Tech. J.*, vol. 42, no. 5, pp. 1977–1997, Sep 1963.

[16] V. Shivaldova and C. F. Mecklenbräuker, "Quantization-based complexity reduction for range-dependent modified Gilbert model," in *Proc. 8th IEEE Sens. Array Multichannel Signal Process. Work.*, 2014, pp. 345–348.

[17] N. Slonim and N. Tishby, "Agglomerative Information Bottleneck," *Adv. Neural Inf. Process. Syst.*, pp. 617–623, 2000.

[18] "Project "Robust And Distributed Safety-Improved Traffic Telematics" (ROADSAFE)," Available online: https://portal.ftw.at/projects/roadsafe/.

[19] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *Proc. 37th Allert. Conf. Commun. Control Comput.*, Apr 1999, pp. 368–377. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.6.9199

[20] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 145–151, 1991.

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: John Wiley & Sons, Inc., Sep 2005. [Online]. Available: http://doi.wiley.com/10.1002/047174882X

[22] Z. Lou, Y. Ye, and X. Yan, "The Multi-Feature Information Bottleneck with Application to Unsupervised Image Categorization," in *Twenty-Third Int. Jt. Conf. Artif. Intell.*, 2013, pp. 1508–1515.

[23] L. R. Welch, "Hidden Markov models and the Baum-Welch algorithm," *IEEE Information Theory Society Newsletter*, vol. 53, no. 4, pp. 10–13, 2003.