A Novel PER Degradation Model for VANETs

P. Fazio, M. Tropea, F. De Rango

Abstract—In this work we present a new channel modeling approach for vehicular networks, where there are frequent topology changes due to a high grade of mobility. The Markov-based trace analysis is used to analyze packet error rate degradation dynamics in relation to both Doppler shift and signal to noise ratio. After the model definition, we applied an additional polynomial regression to remove dependence on boundary parameters. Differing from existing works, our idea takes into account the stochastic behavior of wireless links, in spite of most of the existing closed forms. We considered MATLAB for evaluation of the proposed model. We also propose numerical results as a possible application of our study.

Index Terms—VANET, Channel model, Markov, MTA.

I. INTRODUCTION

The contribution made in this work consists of a new channel model for Vehicular Ad-hoc NETworks (VANETs) which relates, in a closed form, the parameters of a Discrete Time Markov Chain (DTMC) to the performance of a VANET channel, as a function of Signal-to-Noise Ratio (SNR) and Doppler Shift (DS) [1], approaching the Markov-based Trace Analysis (MTA) algorithm. For vehicular environments, it is important to take into account channel conditions [2] as degradations are frequent due to the high grade of node mobility. These aspects play an important role especially when vehicular applications need to deal with channel conditions [3]. Many works in literature, such as [4] and [5], take into account the modeling of channel conditions in VANET by considering path-loss and shadowing effects, which are directly related to the distance between the pairs of nodes. In addition they evaluate the received signal strength by using classical Rayleigh or Nakagami fading/shadowing approaches, disregarding the relationship between the Bit Error Rate (BER) and the Packet Error Rate (PER); neither do they take into account boundary conditions, such as average SNR or Doppler Shift (DS), which are directly related to mobility. Starting from a Simulink model, we based our study on the works proposed in [6], [7], which introduced the Markov Theory, in order to face the non-deterministic nature of the channel. In particular, the authors conducted an in-depth analysis in terms of PER, consenting both the evaluation of channel performance at packet level and of tuning the model based on real observations. Our implementation of the IEEE802.11p PHY model in MATLAB Simulink, related to the transmitter, channel and receiver, derives from the IEEE802.11a model [8], since it has some very similar specifications, as stated in [9]. For example, the frequency bandwidth for each channel is 10MHz, instead of 20MHz and the number of OFDM subcarriers is 64, including 48 data subcarriers and 4 pilot subcarriers (used for tracking frequency offset and phase noise). The standard provides FEC operation, based on convolutional encoding (with generator polynomials \(g_0=133\) and \(g_1=171\)) and code rates of 1/2, 2/3 and 3/4. The MAC is a Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) implemented with the RTS/CTS mechanism. For more details about Simulink blocks and implementation please refer to [8], [9], [10]. The proposed idea consists of two different steps:

- a) Once the model was implemented, we collected many results in Simulink in terms of PER, varying SNR and DS values. The obtained Log-Files (LFs) were analyzed through the MTA algorithm [6], [7], obtaining a Markov model that can describe a stochastic channel behavior that is similar to that analyzed;

- b) We applied a polynomial regression was applied to the Markov chains parameters, to remove dependence on SNR and DS. Thus, we give a closed form for binding channel degradations to host mobility and environmental conditions as a contribution in this paper.

The result of this proposal is a Markov channel modeling approach, suitable for VANETs environments. The obtained parameters are expressed in a closed form, and are directly related to DS and SNR. This document is organized as follows: Section II provides a description of the considered environment, the MTA algorithm and the polynomial regression approach, that can bind the transition probabilities and state sojourn times with the degree of mobility (related to DS) and the environmental conditions (related to SNR). Section III presents a performance evaluation of the proposed model and Section IV summarizes the conclusions.

II. MARKOVIAN MODELING WITH TRACE ANALYSIS

In this section, we propose a DTMC as the result of the application of the MTA algorithm, starting from the LFs obtained by the Simulink model, with fixed parameters (DS and SNR). Then, we apply a polynomial regression to remove dependence on the mentioned parameters. Table I summarizes the main notations, symbols and abbreviations.

A. MTA modeling for VANET channel

As in [6], it is known that a DTMC is a stochastic process that assumes \(n\) discrete values from a finite set \(S=\{s_0,\ldots, s_n\}\), depending only on the current state (chain dependence property). By carrying out many transmissions of \(k\) packets with the obtained Simulink model, we can obtain many Log-Files (LFs). The \(l\)-th LF contains a sequence of \(k\) values: a value of \(\theta\) indicates that a packet was received correctly (no
bit errors), while a value of $I$ means that reception of a corrupted packet occurred, so $L_F \in \{0,1\}^I$. In this last case, for example, considering a convolutional coding rate of 1/2 and a related minimum Hamming distance $H_{	ext{min}}=10$, a received packet is corrupted if it contains at least $\lfloor (H_{	ext{min}}-1)/2 \rfloor +1 = 5$ wrong bits. Differently from [6], which considers only two states (Good or Bad), we do not focus on the correctness of a single packet transmission. Instead, we consider a set of packets transmitted into a Time Observation Window (TOW) and evaluate the badly received percentage, defined as PER Degradation Level (PDL).

**TABLE I. MAIN ABBREVIATIONS, SYMBOLS AND NOTATIONS**

<table>
<thead>
<tr>
<th>$k$</th>
<th>Length of a single Log-File (LF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>Size of an observation window TOW</td>
</tr>
<tr>
<td>$n$</td>
<td>Number of Markov chain states</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of needed TOWs for a LF</td>
</tr>
<tr>
<td>PDL(TOW$_i$)</td>
<td>PER Degradation Level related to $i$-th TOW</td>
</tr>
<tr>
<td>$\mu_{mijk}$</td>
<td>Element $i$ of the probability vector $\mu$, depending on DS</td>
</tr>
<tr>
<td>$\sigma_{mijk}$</td>
<td>$k$-th polynomial coefficient for 1st regression of $\mu_{mijk}$ depending on DS</td>
</tr>
<tr>
<td>$\lambda_{mijk}$</td>
<td>$k$-th polynomial coefficient for 1st regression of $\sigma_{mijk}$ depending on SNR</td>
</tr>
<tr>
<td>$\alpha_{mijk}$</td>
<td>Vector of coefficients $\alpha_{mijk}$ of the $p$-th order polyn.</td>
</tr>
<tr>
<td>$\beta_{mijk}$</td>
<td>Vector of coefficients $\beta_{mijk}$ of the $p$-th order polyn.</td>
</tr>
<tr>
<td>$\Lambda_{mijk}$</td>
<td>$k$-th polynomial coefficient for 2nd regression of $\sigma_{mijk}$ depending on SNR</td>
</tr>
<tr>
<td>$\Psi_{mijk}$</td>
<td>$k$-th polynomial coefficient for 2nd regression of $\Lambda_{mijk}$ depending on DS</td>
</tr>
<tr>
<td>$\Phi_{mijk}$</td>
<td>$k$-th polynomial coefficient for 2nd regression of $\Psi_{mijk}$ depending on SNR</td>
</tr>
<tr>
<td>$\Phi_{mijk}$</td>
<td>$k$-th polynomial coefficient for 2nd regression of $\Phi_{mijk}$ depending on DS</td>
</tr>
<tr>
<td>$W_{mijk}$</td>
<td>Vector of coefficients $W_{mijk}$ of the $p$-th order polyn.</td>
</tr>
<tr>
<td>$L_{mijk}$</td>
<td>Final regression matrix for $\Phi_{mijk}$</td>
</tr>
<tr>
<td>$H_{mijk}$</td>
<td>Final regression matrix for $\Psi_{mijk}$</td>
</tr>
</tbody>
</table>

The length of a Log-File can be covered by several TOWs of length $m$.

\[ PDL(TOW_i) = \sum_{j=1}^{TOW_i} TOW_{(j)} = PDL_i \]

where $TOW_i$ is the $i$-th window and $TOW_{(j)}$ is the $j$-th value of $TOW_i$. For the sake of simplicity, we use the notation $PDL=PD(TOW_i)$. It is clear that the summation takes into account only the total number of wrong packets (good packets are represented by 0, without contributing to the summation observed in the $i$-th window. Fig. 1 illustrates how the original log-file can be covered by several TOWs of length $m$ ($m=12$ in the example). Starting from Eq. 1, the Observed Wrong Packets Vector is defined as $OPWV=m$ PDL, because for each observation $i$, the number of wrong packets is $m$ PDL and $|OPWV|=N$. It is possible to notice that if $m=1$, then the elements of $OPWV$ will coincide with the original values of the LF ($N=k$). At this point, once we obtain the $OPWV$, the admissible values of PDL have to be discretized, in order to define a set of states for the DTMC, associated with the particular ranges of PDL. Following the MTA approach [6], [7] the number of states for the DTMC is set to $n=m+1$. We can state that if $OPWV(j)=q$, then the chain is in state $s_q$, with $0 \leq s < m$. The transition probability from $s_q$ to $s_r$ can be easily evaluated as:

\[ P_{n(q,r)} = tr(s_q, s_r) / tr(s_q) \]

where $tr(s_q, s_r)$ is the number of times that $OPWV[j-1]=q$ and $OPWV[j]=r$ and $tr(s_q)$ is the number of times that $OPWV[j-1]=q$ and $OPWV[j]=s_r, 0 \leq s < m$. $P_n$ is the matrix that resumes all the possible values. As stated in [11], [12], it is possible to increase the accuracy of the model if we consider a higher number of states, i.e. using a larger size of TOW (as shown later), until an unreachable state is added (for $m=1$ and $n=2$, the Gilbert-Elliot model [13] is obtained). In order to check if a Markov chain can model the PDL process, we conducted two tests. It is known that a DTMC $P$ with transition probability matrix $P_n$ is stationary (independent from time) if the equation $\pi \cdot P_n = \pi$ can be solved: if the solution exists, then the elements of $\pi$ represent stationary probability distributions. In order to verify the property for the considered process, the Run-Test (RT) theory, explained in [14] by Bendat and Piersol and only suitable for two-valued data sequences, was used. In particular, defining a run as a sequence of identical observations (followed or preceded by different observations), the $OWPV$ was divided into $L$ runs and, indicating with $L_0$ and $L_1$ the number of runs related to observations 0 and 1 respectively, we obtain $L=L_0+L_1$. Following Bendat and Piersol’s theory [14], if the observations are independent, the probability of having runs 0 and runs 1 does not change from one observation to the next and it is possible to consider the sampling distribution for the number of runs as a random variable $R$, with the following mean (expected number of runs) and variance:

\[ \mu_x = \frac{L_0L_1}{L} + 1, \ \sigma^2 = \frac{L_0L_1(L-1)}{L^2(L-1)} \]  

and for $L>10$ it is a Gaussian distribution $N(\mu_x, \sigma)$. An acceptance range of 95% was considered, so the hypothesis of stationarity can be accepted if $Le [\mu_x-1.96 \sigma]$, $Le [\mu_x+1.96 \sigma]$. Table II shows the obtained values for 5 experiments of the transmission of $k=5000$ packets. We can see how the number of runs respects lower and upper bounds.

**TABLE II. RUN-TEST RESULTS FOR SEVERAL EXPERIMENTS**

<table>
<thead>
<tr>
<th>$#\text{exper.}$</th>
<th>$#\text{runs}$</th>
<th>$\mu_x$</th>
<th>$\sigma$</th>
<th>$\text{Lower B}$</th>
<th>$\text{Upper B}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2476</td>
<td>2422.09</td>
<td>34.706</td>
<td>2353.909</td>
<td>2490.271</td>
</tr>
<tr>
<td>2</td>
<td>2203</td>
<td>2195.416</td>
<td>34.133</td>
<td>2131.47</td>
<td>2259.362</td>
</tr>
<tr>
<td>3</td>
<td>2386</td>
<td>2288.445</td>
<td>34.546</td>
<td>2232.734</td>
<td>2245.155</td>
</tr>
<tr>
<td>4</td>
<td>2231</td>
<td>2201.185</td>
<td>33.100</td>
<td>2136.11</td>
<td>2256.899</td>
</tr>
<tr>
<td>5</td>
<td>2467</td>
<td>2463.904</td>
<td>35.089</td>
<td>2399.219</td>
<td>2532.768</td>
</tr>
</tbody>
</table>

In addition, we performed the Kolmogorov-Smirnov Test (KST) [15] to evaluate the correctness of an exponential approximation for the distributions of the State Sojourn Time (SST) of a generic state $s_q$, characterized by the mean value of:

\[ SST_{n(q,r)} = P_{n(q,r)} \]

Fig. 2a presents the result of the KST for $SST_{n(q,r)}$, the distribution that minimizes the standard error is an exponential one with a mean value of 5.334s (standard error 0.13). The $n$
mean values can be resumed in the $SST_{n}$ vector, considering Eq. 4. Fig. 2b shows the correctness of the considered model for wrong packets cdf, for each suitable value of $m$; the experimental values obtained directly from MATLAB are compared with the ones obtained from the proposed model.

![Fig. 2. Approximations of $SST_{n}$ cdf (a) and Wrong Packets (WP) distribution for different values of $m$ (b).](image)

Fig. 3 illustrates different $P_{n}$ matrices for some values of $n$. Starting from $n=3$, we can see that considering $n=6$ is unnecessary, since the probability of reaching the last state is zero (or almost zero). For $n=6$, the near-to-zero values on the $n$-th column indicate that there is a near-to-zero probability of reaching the $n$-th state, while the value of 1 on the $n$-th row illustrates that the few times the $n$-th state is reached, then the system immediately leaves that state (reaching state 4).

![Fig. 3. Some $P_{n}$ matrices, for $n=3$...6. SNR=25dB and DS=250Hz.](image)

**B. Polynomial regression for SNR and DS**

The DTMC model is now defined, but its parameters ($P_{n}$ and $SST_{n}$) depend on boundary conditions, since the original log-files are obtained by fixing them. In order to remove this dependence we introduced a polynomial regression process, obtaining the elements of $P_{n}$ and $SST_{n}$ as a function of SNR and DS. One source of degradation in wireless communications with high mobility is the DS; the authors in [1] show we can express the relationship between relative nodes speed and DS as follows:

$$DS_{(k)} = \alpha_{[m-1]} \cdot V_{(m/s)}$$

(5)

where $\alpha=18.4237$ m$^{-1}$. Suitable values of DS belong to the range [100, 400] Hz. The average SNR in VANETs depends on the Received Signal Power (RSP) and the Noise Power (NP, thermal noise in particular for VANETs), as shown in [16], [17]. As stated in [18], suitable values of SNR belong to the interval [5, 35] dB. Choosing a fixed value of $m=n+1$ the main idea now consists of the regression of the $P_{n}$ and $SST_{n}$ parameters to obtain their values as functions of DS and SNR.

The MATLAB regression analysis (with a minimum value of determination coefficient $R^{2}$ over all obtained polynomial functions of 0.9853) was introduced in [12]. Fig. 4 and 5 show the relationship between $P_{5(3,2)}$ and $SST_{5(3)}$ with SNR and DS: $P_{5(3,2)}$ has a decreasing trend for increasing values of SNR and DS, while the trend of $SST_{5(3)}$ is the opposite. This type of analysis was performed for all the elements of $P_{n}$ and $SST_{n}$, with $n=3, 4$ and 5.

![Fig. 4. Trend of $P_{5(3,2)}$ for different values of SNR and DS.](image)

![Fig. 5. Trend of $SST_{5(0)}$ element for different values of SNR and DS.](image)

After choosing $n$, the approximation process begins with fixing SNR to a value within the considered range [5, 35] dB. Independently from the chosen SNR value, the MATLAB polyfit tool shows that a $2^{nd}$ order polynomial regression can express the relationship between the generic $P_{n(i,j)}$ and DS, as well as between the generic element $SST_{n(i,j)}$ and DS. Thus:

$$P_{n(i,j)}(DS) = \begin{bmatrix} \beta_{p}\alpha_{n(i,j)} & \alpha_{n(i,j)} & \alpha_{n(i,j)} \end{bmatrix} \begin{bmatrix} \alpha^{2}DS^{2} \end{bmatrix}$$

$$SST_{n(i)}(DS) = \begin{bmatrix} \beta_{p}\alpha_{n(i,j)} & \alpha_{n(i,j)} & \alpha_{n(i,j)} \end{bmatrix} \begin{bmatrix} \alpha^{2}DS^{2} \end{bmatrix}$$

(6)

(7)

where $\beta_{p}$ represents the transposition operator, $\alpha_{n(i,j)}$ and $\beta_{n(i,j)}$ are the coefficient vectors for $P_{n}$ and $SST_{n}$ respectively, $\alpha^{2}DS^{2}$, $\alpha^{2}DS$ are $(p \times 1)$ vectors containing the $DS$ and $SNR$ values respectively, with $q=0,...,p$. Since the SNR value was set for the first regression, the obtained coefficients will vary for different SNR values.

Considering the dependence of $P_{n}$ and $SST_{n}$ on SNR, the elements of $A^{2\alpha_{n(i,j)}}$ and $B^{2\alpha_{n(i,j)}}$ can be described by introducing a second polynomial regression of the $3$-rd order (also verified in MATLAB), as a function of SNR:

$$a_{n(i,j,k)}(SNR) = \begin{bmatrix} \lambda_{n(i,j,k)} & \lambda_{n(i,j,k)} & \lambda_{n(i,j,k)} & \lambda_{n(i,j,k)} \end{bmatrix} \begin{bmatrix} \alpha^{3}SNR^{3} \end{bmatrix}$$

$$b_{n(i,j)}(SNR) = \begin{bmatrix} \lambda_{n(i,j,k)} & \lambda_{n(i,j,k)} & \lambda_{n(i,j,k)} \end{bmatrix} \begin{bmatrix} \alpha^{3}SNR^{3} \end{bmatrix}$$

(8)

(9)

Replacing the expressions of Eq. 8 and 9 in Eq. 6 and 7, we obtain the following final relations:
we propose the regression (3x4) matrices
\[ P_{n(i,j)}(DS,SNR) = [\Lambda_{n(i,j)}^{DS} \cdot \frac{SNR}{DS}] \cdot DS \quad (10) \]
\[ SST_{n(i,j)}(DS,SNR) = [\Pi_{n(i,j)}^{SNR} \cdot \frac{DS}{DS}] \cdot DS \quad (11) \]
where \( \Lambda_{n(i,j)} \) and \( \Pi_{n(i,j)} \) are the following (3x4) matrices:
\[ \begin{bmatrix}
  \lambda_{0,0}^{(0)} & \lambda_{0,1}^{(0)} & \lambda_{0,2}^{(0)} & \lambda_{0,3}^{(0)} \\
  \lambda_{1,0}^{(0)} & \lambda_{1,1}^{(0)} & \lambda_{1,2}^{(0)} & \lambda_{1,3}^{(0)} \\
  \lambda_{2,0}^{(0)} & \lambda_{2,1}^{(0)} & \lambda_{2,2}^{(0)} & \lambda_{2,3}^{(0)} \\
\end{bmatrix} \quad (12) \]

In this way, all the elements of \( P_n \) and \( SST_n \) are expressed as functions of DS and SNR.

III. MODEL EVALUATION

In order to appreciate the accuracy of the proposed model, we conducted numerous simulation campaigns. In Fig. 6 we propose the regression (3x4) matrices \( \Lambda_{n(0),0} \) and \( \Pi_{n(0)} \) for different values of DS and SNR:
\[ \Lambda_{n(0),0} = \begin{bmatrix}
  -4.18 \times 10^{-6} & 2.88 \times 10^{-7} & -7.74 \times 10^{-7} & 1.527 \\
  7.98 \times 10^{-6} & -1.28 \times 10^{-6} & 3.62 \times 10^{-6} & 1.13 \\
  7.58 \times 10^{-6} & -6.16 \times 10^{-6} & 1.46 \times 10^{-6} & 0.001 \\
\end{bmatrix} \]

\[ \Pi_{n(0)} = \begin{bmatrix}
  8.18 \times 10^{-6} & -1.28 \times 10^{-6} & 3.62 \times 10^{-6} & 1.13 \\
  1.8 \times 10^{-6} & -1.8 \times 10^{-6} & 0.0047 & -0.314 \\
  -6.7 \times 10^{-6} & 7.1 \times 10^{-6} & -1.75 \times 10^{-6} & 0.018 \\
\end{bmatrix} \]

In Fig. 6, the regression matrices for \( P_{0(0),0} \) and \( SST_{0(0)} \) are plotted.

Fig. 7 shows the approximations obtained by Eq. 10 and 11 respectively, for \( P_{0(0),0} \) and \( SST_{0(0)} \). Markers represent the original values, while lines represent the obtained polynomial fitting. We can observe how the proposed polynomial approach fits the original values (as previously stated, all the obtained regression determination coefficients \( R^2 \) are greater than 0.9853).

In Fig. 8, the PDL trend is shown for different values of DS and SNR, and its approximation with the 4-states model (markers).

In conclusion, we found that the introduced approach can describe the PDL with a negligible amount of error.

IV. CONCLUSIONS

In this paper, we propose a new approach for modeling the wireless link in VANET environments. It is based on the MTA algorithm (for analyzing PER performance as a function of SNR and DS values) and on polynomial regression, in order to obtain a function of SNR and DS for each parameter of the Markovian model (transition probabilities and state sojourn times). In this way, we can obtain a complete dynamic of the wireless link degradation, deriving it from real environment conditions (distance, power, antennas, relative speed, etc.). In addition, we also show also that the number of states equal to 5 is an upper bound, since no greater accuracy can be obtained if n<6. The obtained curves show the goodness of the polynomial approximation.

References