Analysis of alert message propagation on the highway in VANET assuming Markovian vehicle arrival process

Dhari Ali Mahmood1,2 | Gabor Horvath1,3

1Department of Networked Systems and Services, Budapest University of Technology and Economics, Budapest, Hungary
2Department of Computer Engineering, University of Technology, Baghdad, Iraq
3MTA-BME Information Systems Research Group, Budapest, Hungary

Correspondence
Dhari Ali Mahmood, Magyar Tudósok krt. 2, 1117 Budapest, Hungary.
Email: dhariali@hit.bme.hu

Funding information
OTKA K-123914; National Research, Development and Innovation Fund, Grant/Award Number: TUDFO/51757/2019-ITM

Summary
Vehicular ad hoc networks (VANETs) provide the infrastructure for the intelligent transportation system (ITS). The number of vehicles that are capable of communicating with each other through VANET is increasing. In this paper, we provide several analytical results regarding the message propagation on the highway in VANET. In our scenario, there is a stationary message source (i.e., alert messages are generated at an accident). Vehicles can receive the message and get informed when they reach the radio transmission range of the message source or another informed vehicle. Most papers published on the analysis of message propagation assume that the inter arrival times between vehicles follow a Poisson process; there are very few results available with more general traffic model. In this paper, we show that the Poisson process is not always suitable for modeling vehicle traffic. Instead of the Poisson process, we propose to use the more general Markovian arrival process (MAP) to model the vehicle headway times and derive the probability that the message propagates beyond a certain distance from the accident under this traffic assumption. Through several numerical examples, we demonstrate how much the statistics of the vehicle arrival process impacts the message propagation distance.

KEYWORDS
ITS, Markovian arrival process, message propagation, transient analysis, V2V, VANET

1 | INTRODUCTION

The intelligent transport systems (ITSs) are going to be widespread in the foreseeable future. One of the primary components of ITS is the communication platform connecting the vehicles and the roadside infrastructure, called vehicular ad hoc network (VANET1,2).

Three types of communication are distinguished in VANET, based on the type of the communicating parties: V2V refers to vehicle-to-vehicle, V2I refers to vehicle-to-infrastructure, while V2X refers to vehicle-to-everything communication.3 The V2V systems will be deployed first, as they do not require the installation of roadside units, so we focus on the V2V technology in this paper.

IETF has identified a number of use cases for V2V communication,4 including collision avoidance based on context-aware navigation, platooning on the highway, and cooperative environment sensing. All of these applications rely on exchanging messages among vehicles that can carry sensor data or can report road events (accident, traffic jam, etc.).
Hence, analyzing the behavior of message propagation is necessary to measure the quality of service in these use cases. Two factors make the analysis difficult: the lack of a suitable model for vehicular traffic, and the complexity of the system due to the underlying protocol stack, radio specific behavior, and the impact of the structure of the road network.

There are two main techniques to obtain performance measures in a VANET system. The typical approach is based on discrete-event simulation, with SUMO/Omnet++ being the most popular tool for this purpose. Analytical models have several advantages over simulations: they are typically faster to evaluate and provide a better, more direct insight into the behavior of the system; on the other hand, they are much more difficult to develop. Nevertheless, some results based on analytical models are also available. In both cases, an accurate traffic model is necessary to characterize the arrival process of the vehicles. In some papers, the motion of the vehicles is modeled in detail, taking not only the arrival time, the speed (velocity), but also the acceleration also into account. However, in the vast majority of the cases, researchers rely on the Poisson model and assume constant velocity for analytical simplicity. Among other papers, the failure of Poisson modeling for vehicle arrivals was recognized in previous studies.

More general traffic models are needed to describe the arrival process of the vehicles more accurately. As natural extensions of the Poisson process, Markov arrival processes (MAPs) have been successfully used in several areas to model complex traffic patterns. MAPs are not only more general than the Poisson process, in fact, MAPs include the Poisson process as a special case. When MAPs fail to model the realistic traffic, the Poisson process fails, too. However, if the Poisson process fails, MAPs can still approximate the real behavior reasonably well. MAPs have been used for modeling vehicular traffic previously, starting with Alfa et al. and including Farkas et al. This idealistic behavior is called deterministic message passing in the literature, and novel broadcasting schemes are getting closer and closer to it. In our model, the arrival process of the vehicles is time stationary (we do not consider seasonality), the velocity of the vehicles is constant, and the effect of the counter-flow traffic is ignored. While the constant velocity assumption does not hold in reality, we need this restriction in order to keep the analytical model tractable (relaxing this restriction is among our future plans). This scenario seems simple and idealistic, yet it is challenging to study with analytical tools, even assuming Poisson traffic. Our aim to study the message propagation distance, thus how far the alert message propagates (measured from the accident) after a certain amount of time.

A similar problem has been solved in Babu and Muhammed Ajeer, with a Poisson process and the presence of channel randomness. In Zhuang et al. and in Miorandi and Altman, the mean message propagation distance has been derived for the same scenario, with Poisson vehicle traffic. In Mahmood and Horvath, the detailed transient analysis of the message propagation is provided, still under the Poisson assumption. Not surprisingly, other studies and Mahmood and Horvath arrived to the same results, but with three different analysis approaches: they express the mean message propagation distance explicitly, as the function of the radius of the radio coverage, the vehicle speed, and the vehicle arrival intensity. In a more recent paper, the message delivery delay is investigated in case of nonconstant, uniformly distributed vehicle velocity, but it still assumes Poisson vehicle arrival process.

In this paper, we show that the Poisson process is not always suitable for modeling vehicle traffic. Instead of the Poisson process, we propose to use the more general Markovian arrival process to model the vehicle headway times, which still allows for tractable analysis. We provide analytical results on both the stationary and the transient properties of the message propagation distance assuming Markovian vehicle arrival process and show that the more accurate modeling of the vehicle arrival process implies a better approximation of the message propagation distance.

The rest of this paper is organized as follows. Section 2 introduces MAPs and shows how to obtain them based on empirical measurement data. The main results on the clusters of informed vehicles are presented in Section 3. The stationary and the transient properties of the message propagation distance are described by Section 4. Section 5 provides some numerical examples; finally, Section 6 concludes the paper.

2 STOCHASTIC MODELS FOR THE VEHICLE ARRIVAL PROCESS

The Poisson process has been used in the literature for a long time to model the interarrival times of the vehicles. The reason for this modeling choice is not its validated correctness, but its analytical simplicity; many useful performance measures can be expressed analytically when Poisson traffic is assumed.
2.1 The failure of Poisson process in modeling vehicular traffic

The Poisson process is, however, not a good model for the vehicular traffic in the vast majority of the cases. After examining a number of recent traces comprising of LIDAR measurements, it is easy to show the apparent differences between the statistics of the empirical measurements and the Poisson process.

The left plot in Figure 1 depicts the empirical probability density function (pdf) of the empirical measurements of vehicle interarrival times (also referred to as headway times in the literature) and the pdf of the Poisson model having the same traffic intensity. It can be seen clearly that the shapes of the pdfs are significantly different. A popular measure used to quantify “burstiness” of the traffic is the squared coefficient of variation (SCV), which is defined by the ratio of the variance and the square of the mean value. In the Poisson case (when the interarrival times are exponentially distributed), \( SCV = 1 \) holds. SCV values lower than 1 correspond to more “regular” (closer to deterministic) traffic, while \( SCV > 1 \) means that the traffic is highly varying. For this measurement trace, we got \( SCV = 46.913 \), which means that the vehicle interarrival times are extremely varying.

Another property of the Poisson process is that it is unable to capture the correlations between the vehicle interarrival times. In practice, however, the interarrival times are correlated, as shown by the right plot in Figure 1, which depicts the correlation of an interarrival time and the \( k \)th successive interarrival time.

In the subsequent sections, we show that the accurate modeling of vehicular traffic is essential to obtain accurate analytical results for the message propagation distance (and, we believe that for many other performance measures, too).

2.2 Markov arrival process

Having shown that the Poisson model is not always suitable for vehicular traffic, we propose an alternative solution in this section. The Markovian arrival process (MAP) is a useful modeling tool that is capable of characterizing non-Poisson traffic. It is widely applied in many areas, including telecommunication and logistic systems. MAPs, while being rather general and flexible traffic models, are still relatively simple to work with during analytical derivations.

In a Poisson process, the interarrival times are exponentially distributed. In the case of MAPs, the interarrival times are not exponentially distributed but are given by the composition of several exponentially distributed phases. More precisely, MAPs have a background process, which is an irreducible continuous-time Markov chain (CTMC), where two kinds of state transitions are distinguished: those that generate an arrival and those that do not. If the size \( N \times N \) generator matrix of the background process is denoted by \( D \), we have that \( D = D_0 + D_1 \), where matrix \( D_1 \) contains the rates of those state transitions which are accompanied by a vehicle arrival and matrix \( D_0 \) contains the rate of those transitions that do not generate arrival events, they are internal transitions only. Since \( D_0 \parallel = \parallel \) (\( \parallel \) is the column vector of ones and \( \parallel \) is the column vector of zeros of appropriate size), we have that \( D_0 \parallel = -D_1 \parallel \). Figure 2 presents an example with \( N = 5 \), where the dashed lines are the internal transitions and the solid ones are accompanied by vehicle arrivals. The states of the background process are often referred to as phases, too.

If the state process of the MAP is denoted by \( J(t), t \geq 0 \), and the \( k \)th interarrival time by \( T_k \), then the complementary cumulative distribution function (ccdf) and the density function (pdf) of the interarrival times are given by

\[
P(T_k > t, J(t) = j | J(0) = i) = \left[ e^{D_0 t} \right]_{ij},
\]

\[
\lim_{\Delta \to 0} \frac{1}{\Delta} P(T_k \in [t, t + \Delta), J(t + \Delta) = j | J(0) = i) = \left[ e^{D_0 t} D_1 \right]_{ij},
\]

**FIGURE 1** Comparison of the empirical pdfs and lag-\( k \) correlations of the vehicle interarrival times
if the initial state is \(i\) and the state at \(t+\Delta\) is \(j\). (Note that \(e^A\) denotes the matrix-exponential function and \([A]_{i,j}\) denotes the \(i, j\)th element of matrix \(A\). Matrices are denoted by bold letters throughout the paper). Since the pdf of the inter-arrival times in a Poisson process is \(e^{-\lambda t}\), MAPs can be considered as the matrix generalizations of the Poisson process.

The stochastic matrix \(P\) containing the state transition probabilities between two consecutive interarrival events is obtained by

\[
P = \int_0^\infty e^{D_0 t} D_1 dt = (-D_0)^{-1} D_1^n
\]

and will be used frequently in the sequel. The stationary row vector of matrix \(P\), denoted by \(\pi\), is the solution of \(\pi P = \pi\), \(\pi_1 = 1\). With this vector, we can express the scalar (state independent) pdf of the interarrival times from (2) as

\[
h(t) = \lim_{\Delta \to 0} \frac{1}{\Delta} P(T_k \in (t, t+\Delta)) = \pi e^{D_0 t} D_1 \mathbb{1}.
\]

Based on this result, it is easy to derive the following simple quantities characterizing the interarrival times:

\[
H(t) = P(T_k < t) = 1 - \pi e^{D_0 t} \mathbb{1},
\]

\[
E(T^n_k) = n! \cdot \pi (-D_0)^{-n} \mathbb{1}.
\]

The interarrival times generated by a MAP can be correlated, too. The lag-\(k\) correlation, thus the correlation between \(T_1\) and \(T_{k+1}\) can be calculated by

\[
\text{Corr}(T_1, T_{k+1}) = \frac{E(T_1 T_{k+1}) - E(T_1) E(T_{k+1})}{\sqrt{E(T_1^2) - E^2(T_1)} \sqrt{E(T_{k+1}^2) - E^2(T_{k+1})}} = \frac{\lambda_0 \mathbb{P}^k (-D_0)^{-1} \mathbb{1} - 1}{2 \lambda_0 (-D_0)^{-1} \mathbb{1} - 1},
\]

where \(\lambda = 1/E(T_k)\) is the mean arrival rate and \(\pi\) is the stationary probability vector of the CTMC given by matrix \(D\), satisfying \(\alpha D = 0, \alpha \mathbb{1} = 1\).

### 2.3 Obtaining the MAP from empirical measurement data

MAPs form a dense class in the set of point processes, which means that every point process can be approximated arbitrary well with a MAP having appropriately many phases. The critical question is how to obtain the matrix parameters of the MAP, \(D_0, D_1\), to approximate the real traffic behavior accurately.

One possibility is to build the background Markov chain of the MAP based on the intuitive understanding of the traffic. For instance, in Figure 3, the arrival process is Poisson, but the arrival rate is time-dependent, where the periods
determining the arrival rate are exponentially distributed and follow each other in a circular way. The solid lines represent transitions of $D_0$, and the dashed lines the ones of $D_1$ (here, vehicle arrivals are not accompanied by phase transitions).

When empirical measurement data is available, an automatic approach can also be used. The process of obtaining the $D_0,D_1$ matrix parameters of a MAP from empirical measurements is called MAP fitting. There are several MAP fitting algorithms available. Some of them compute statistical quantities (marginal moments, auto-correlations, joint moments, etc.) and create the $D_0,D_1$ matrices such that the resulting MAP exhibits the same statistics. Some other algorithms try to find the MAP that is most likely to generate the measurement data with the expectation-maximization algorithm. Nevertheless, MAP fitting is an inherently difficult problem, the corresponding algorithms are continuously improving, but there is no single perfect tool available yet.

For our VANET traffic trace, we have applied the EM algorithm in Horváth and Okamura and the KPC-toolbox. From the two, the first one provided better results, and it will be used for numerical investigations in the rest of the paper. As depicted by Figure 1, the pdf of the vehicle interarrival times is accurately captured by the MAP model, as opposed to the Poisson model. We got somewhat worse matching for the correlations, which can be a consequence of the too low number of data to fit.

3 | CLUSTERS OF INFORMED VEHICLES

In this paper, we consider a scenario where there are a single road and a message source. The message source is stationary (it does not move) and emits messages continuously. The message source, as well as all the vehicles, communicate through radio transmission, with the radio coverage assumed to be fixed with radius $R$ (Table 1). The transmission is considered to be perfect; the messages are delivered without any delay and error (also referred to as the Friis model). When a vehicle receives the message, it becomes informed, and, while moving, emits the message itself, too.

All vehicles are assumed to be identical and have the same constant speed $v$. Since all vehicles have the same speed, they can not overtake each other. The road is straight and has infinite length.

A group of vehicles, where the distance between subsequent vehicles is less than the radio transmission range $R$, is called a cluster. We assume that the vehicles in the same cluster can exchange information immediately. Whenever the first vehicle sends a message, all the vehicles within the cluster receive it immediately. The cluster length, represented by random variable $G$, is measured from the position of the first vehicle to the last position where the information is available, that is, the position of the last vehicle plus $R$ (see Figure 4).

In the rest of the section, we are going to study the stochastic behavior of $G$, as it is an essential ingredient in the analysis of the message propagation distance. Since we have a MAP vehicle arrival process, to describe the system completely, we have to consider the joint behavior of $G$ and the phase of the MAP. Namely, we have to keep track of the phase of the MAP at two specific time points: at the time when the first vehicle of the cluster was generated, and at $G/v$ later (at the end of the cluster).

If the phase of the MAP at time $t$ is denoted by $J(t)$, and the first vehicle of the cluster was generated at time $t=0$ (without loss of generality), the complementary cumulative distribution function (ccdf) of $G$ and the phase at the end of the cluster is defined by

$$G_{ij}(x) = P(G > x, J(G/v) = j | J(0) = i),$$

and the corresponding matrix $G(x)$ is defined by $G(x) = [G_{ij}(x)]$.

Before expressing $G(x)$, let us introduce matrix $Z = [Z_{ij}]$ with the transition probabilities of the MAP phases between the beginning and the end of the cluster as

**FIGURE 3** State-transition diagram for MAP based on VANET
Thus, $Z_{ij}$ is the probability that the phase of the MAP is $j$ at the end of the cluster, given that it was $i$ at the beginning of it. For matrix $Z$, an explicit solution is provided by the next theorem.

**Theorem 1** Matrix $Z$ can be expressed by

$$Z = (I - P + e^{D_1R/v}P)^{-1}e^{D_1R/v}. \tag{10}$$

**Proof** By conditioning on the time between two vehicles, there are two possibilities: either the headway time is greater than $R/v$, or it is less than or equal to $R/v$. In the former case, the two vehicles do not belong to the same cluster. In the latter case, they are in the same cluster, and the phase transition probabilities can be calculated recursively, based on the cluster initiated by the second vehicle. Using (1) and (2), we get

$$Z_{ij} = [e^{D_1R/v}]_{ij} + \sum_{k=1}^{N} \int_{y=0}^{R/v} [e^{D_1y}D_1]_{ik}Z_{kj} \, dy, \tag{11}$$

which, switching to matrix notation, translates to
Proof To obtain the mean value, the integral of the ccdf (12) is calculated as follows:

\[
Z = e^{D_0 R/v} + \int_{y=0}^{R/v} e^{D_0 y} D_1 Z \, dy,
\]

since at speed \( v \) the transmission range \( R \), in distance, translates to \( R/v \), in time. Moreover, the second term of \( Z \) can be simplified as

\[
\int_{y=0}^{R/v} e^{D_0 y} D_1 Z \, dy = \int_{y=0}^{R/v} e^{D_0 y} D_1 Z \, dy - e^{D_0 R/v} \int_{y=R/v}^{\infty} e^{D_0 (y-R/v)} dy D_1 Z
\]

\[
= (-D_0)^{-1} D_1 Z - e^{D_0 R/v} (-D_0)^{-1} D_1 Z,
\]

which, after simple transformations, establishes the theorem.

Similar to the results available for the Poisson case,\(^\text{18}\) by the stochastic interpretation of the system, \( G(x) \) can be expressed recursively as

\[
G(x) = \begin{cases} 
Z, & \text{if } x \leq R, \\
\int_{y=0}^{R/v} e^{D_0 y} D_1 G(x-yv) \, dy, & \text{if } x > R.
\end{cases}
\] (12)

The first term of (12) corresponds to the case when \( x \) falls into the transmission range of the first vehicle of the cluster. In this case \( G_i(x) = Z_i \), since \( G \geq R \) holds, we have to take care only about the phase transitions (see (8) and (9)).

In the second case of (12), \( x \) is out of range for the first vehicle, thus to cover \( x \), the cluster has to consist of more vehicles (Figure 4). Conditioning on the interarrival time between the first and the second vehicle of the cluster (having pdf \( e^{D_0 y} D_1 ) \) leads to (12).

In order to compute the mean cluster length \( E(\mathcal{G}) \), we define the phase-dependent mean cluster length that contains information on the MAP phase when the first car was generated \( J(0) \) and the MAP phase at the end of the cluster \( J(\mathcal{G}/v) \). The \( i,j \)th element of the phase-dependent mean cluster length matrix is defined by \( [E(\mathcal{G})]_{ij} = E(\mathcal{G} \cdot \mathcal{I}_{J(\mathcal{G}/v)=j} | J(0)=i) \), where \( \mathcal{I}_{\{\}} \) is the indicator variable.

**Theorem 2** The phase-dependent mean value of \( G \) can be calculated by

\[
E(\mathcal{G}) = (I - P + e^{D_0 R/v} P)^{-1} [(I - P) R Z - D_0^{-1} Z v] + Z D_0^{-1} v.
\] (13)

**Proof** To obtain the mean value, the integral of the ccdf (12) is calculated as follows:

\[
E(\mathcal{G}) = \int_{x=0}^{\infty} G(x) \, dx = \int_{x=0}^{R} G(x) \, dx + \int_{x=R}^{\infty} G(x) \, dx
\]

\[
= R Z + \int_{y=0}^{R/v} e^{D_0 y} D_1 \int_{x=0}^{\infty} G(x-yv) \, dx \, dy
\]

\[
= R Z + \int_{y=0}^{R/v} e^{D_0 y} D_1 \int_{x=yv}^{\infty} G(x-yv) \, dx \, dy - \int_{y=0}^{R/v} e^{D_0 y} D_1 \int_{x=yv}^{R/v} G(x-yv) \, dx \, dy
\]

\[
= R Z + \int_{y=0}^{R/v} e^{D_0 y} D_1 E(\mathcal{G}) \, dy - \int_{y=0}^{R/v} (R-yv) e^{D_0 y} D_1 Z \, dy,
\]

\[
= R Z + \frac{(1-e^{D_0 R/v}) P E(\mathcal{G})}{R P Z + v D_0^{-1} P Z - v e^{D_0 R/v} D_0^{-1} P Z}.
\]
where in cat \((I - e^{D_0 R/y})PZ = Z - e^{D_0 R/y}\) and that \((I - P + e^{D_0 R/y}P)^{-1}e^{D_0 R/y} = Z\) from (10). Collecting the \(E(G)\) terms together provides the theorem.

In the next step, we compute the phase-independent mean cluster length \(E(G)\) from matrix \(E(G)\). To do so, we have to obtain the stationary phase distribution of the MAP at the beginning of the clusters, denoted by row vector \(\gamma\).

**Lemma 1** The stationary phase distribution vector of the MAP at the beginning of the cluster, vector \(\gamma\), satisfies the linear set of equations \(\gamma Z P = \gamma, \gamma Z I = 1\).

**Proof** Let us define a discrete-time Markov chain characterizing the evolution of the phases at the beginning of the clusters. The phase transitions between the beginning and the end of the clusters are governed by stochastic matrix \(Z\). The phase transitions between the beginning and the end of the intra-cluster periods are governed by stochastic \(P = (-D_0)^{-1}D_1\). Hence, the phase transitions over the beginning of two consecutive clusters are given by matrix \(ZP\) (that is also stochastic), whose stationary distribution provides \(\gamma\).

**Corollary 1** The phase-independent mean cluster length, \(E(G)\), is

\[
E(G) = \gamma E(G) I = \gamma \left( (I - P + e^{D_0 R/y}P)^{-1} (-D_0)^{-1} Z Z (-D_0)^{-1} \right) I. \tag{14}
\]

After the description of the first moment by (14), the next theorem and the corresponding corollary express the second moment. Based on the second moment, it is possible to study the variance of the cluster length, but it is an important ingredient necessary to compute the mean information distance in the next section, too.

**Theorem 3** The phase-dependent second moment of the cluster length can be expressed by

\[
E(G^2) = Z (vD_0^{-1} - RI) (2PE(G) - 2vD_0^{-1}PZ) + (I - P + e^{D_0 R/y}P)^{-1} \left[ 2v(-D_0)^{-1}PE(G) + 2v^2(-D_0)^{-2}PZ \right]. \tag{15}
\]

**Proof** To obtain \(E(G^2)\), the integral of the ccdf (12) is calculated by

\[
E(G^2) = 2 \int_{x=0}^{\infty} xG(x) dx = 2 \int_{x=0}^{R} xG(x) dx + 2 \int_{x=R}^{\infty} xG(x) dx = 2 \int_{x=0}^{R} xZ dx + 2 \int_{y=0}^{R/y} e^{D_0 y} P \int_{x=R}^{\infty} xG(x-ny) dx dy, \tag{16}
\]

where the term marked with (*) can be simplified as

\[
\int_{x=0}^{\infty} (x-ny + ny)G(x-ny) dx - \int_{x=0}^{R} (x-ny + ny)G(x-ny) dx. \]

Putting this result back to (16), for all terms of \(E(G^2)\), we get the following:
\[ E(G^2) = R^2 Z + \frac{R}{\nu} \int_{y=0}^{\frac{R}{\nu}} e^{D_0 y} D_1 E(G^2) dy + 2 \int_{y=0}^{\frac{R}{\nu}} e^{D_0 y} D_1 yv E(G) dy \]

\[ - \int_{y=0}^{\frac{R}{\nu}} e^{D_0 y} D_1 (R^2 - (yv)^2) Z dy. \] (17)

To simplify the third term of (17), we apply integration by parts, leading to

\[ 2 \int_{y=0}^{\frac{R}{\nu}} e^{D_0 y} D_1 yv E(G) dy = \left[ -2Re^{D_0 R/\nu} - 2vD_0^{-1} + 2vD_0^{-1} e^{D_0 R/\nu} \right] PE(G). \]

In the same way, we can also simplify the the last term of (17) as

\[ \int_{y=0}^{\frac{R}{\nu}} e^{D_0 y} D_1 (yv)^2 Z dy = \left[ -R^2 + 2vRD_0^{-1} e^{D_0 R/\nu} + 2v^2 (D_0)^{-2} - 2v^2 e^{D_0 R/\nu} (D_0)^{-2} \right] PZ. \]

Finally, putting all terms together gives

\[ E(G^2) = (I - P + e^{D_0 R/\nu} P)^{-1} \left( R^2 (I - P) Z - 2Re^{D_0 R/\nu} PE(G) \right) - 2vD_0^{-1} PE(G) + 2vD_0^{-1} e^{D_0 R/\nu} PE(G) + 2vRD_0^{-1} e^{D_0 R/\nu} PZ \]

\[ + 2v^2 D_0^{-2} PZ - 2v^2 e^{D_0 R/\nu} (D_0)^{-2} PZ \].

Exploiting \((I - P + e^{D_0 R/\nu} P)^{-1} e^{D_0 R/\nu} = Z\) from (10) and applying simple transformations establishes the theorem.

As we did with the mean cluster length, we can also obtain the phase-independent (scalar) second moment of the cluster length by multiplying \(E(G^2)\) by \(\chi\) from the left and by \(I\) from the right.

**Corollary 2** The phase-independent second moment of the cluster length, \(E(G^2)\), is

\[ E(G^2) = 2v^2 (I - P + e^{D_0 R/\nu} P)^{-1} (-D_0)^{-1} E(G) I - 2RE(G) - 2vRZ(-D_0)^{-1} I. \] (18)

Finally, we provide a differential equation for the ccdf \(G(x)\) itself. The motivation is that it is usually easier to solve a differential equation than the integral equation of (12).

**Theorem 4** The phase-dependent ccdf of \(G\) is the solution of the delayed differential equation (DDE)

\[ \frac{d}{dx} G(x) = DG(x) - e^{D_0 R/\nu} D_1 G(x-R), \quad x > R, \] (19)

with boundary condition \(G(x) = Z, x \leq R\).

**Proof** Let us express and manipulate \(G(x+R)\). For \(x > 0\), we need to use the second case of (12) only. Changing variable in the integral gives
\[ G(x + R) = \int_{y=0}^{x+R} e^{D_0 y} D_1 G(x + R - yv) dy = \int_{u=x}^{x+R} e^{D_0 (x + R - u)/v} D_1 G(u) du. \]

Taking the derivative and using the Leibniz integral rule leads to

\[ \frac{d}{dx} G(x + R) = D_1 G(x + R) - e^{D_0 R/v} D_1 G(x) + D_0 \left( \int_{u=x}^{x+R} e^{D_0 (x + R - u)/v} D_1 G(u) du \right) \]

\[ = (D_0 + D_1) G(x + R) - e^{D_0 R/v} D_1 G(x), \]

which establishes the theorem.

4 | ANALYSIS OF THE MESSAGE PROPAGATION

Assume that an event occurs on the highway at position \( A \). In the rest of the paper, this position is assumed to be fixed, and messages advertising the event are generated continuously for a long time. Here we study the right-continuous stochastic process \( \{ D(t), t > 0 \} \), where \( D(t) \) is the information propagation distance, that is the position of the last car measured from \( A \) having the message received at time \( t \), plus \( R \) (the radius of its radio coverage). In Mahmood and Horvath,\(^{18} \) we have studied the same scenario with the Poisson arrival process, where we have shown that due to the PASTA property \( E(D) \) and \( E(G) \) are equal. In the case of MAP, the \( E(D) \) and \( E(G) \) are not the same, and we are going to investigate their relation in this section.

The time evolution of \( D(t) \) is shown in Figure 5. The trajectory of \( D(t) \) consists of alternating intervals. There are intervals where no vehicles hold the information; the length of these intervals (in distance) is denoted by \( H \). Then, a vehicle enters the range of the accident and gets informed, informing its cluster of length \( G \) as well. This informed cluster will leave the accident in time \( G/v \), followed by another uninformed interval, and so forth.

The mean length of the uninformed intervals is easy to derive. The phase of the MAP at the beginning of a cluster is given by vector \( \gamma \); on the other hand, the phase at the end of the cluster is \( \gamma Z \). With this initial phase, the mean time till the MAP generates an arrival is given by \( \gamma Z (\gamma D_0)^{-1} \mathbb{I} \), thus we have

\[ E(H) = \gamma Z (\gamma D_0)^{-1} \mathbb{I}. \] (20)

From the properties of \( H \) and \( G \) derived above, we can characterize the properties of \( D(t) \).

4.1 | Mean information distance

Let us first derive the mean value of \( D = \lim_{t \to \infty} D(t) \), denoted by \( E(D) \).

\[ \begin{align*}
\text{FIGURE 5} & \quad \text{The evolution of the information distance } D(t) \\
\end{align*} \]
Theorem 5  The mean information distance is expressed by
\[ E(D) = \frac{E(G^2)}{E(G)} + R. \]  \hfill (21)

Proof  The process \( \{D(t), J(t)\} \) forms a Markov renewal process at the moments when the current informed cluster leaves the accident. This means that at these time instants, the phase of the MAP characterizes the future of the process completely.

The theorem can be proven by the renewal reward theorem\(^{30}\) as
\[ E(D) - R = \lim_{t \to \infty} \frac{1}{t} \int_{u=0}^{t} \frac{C}{D(u) - R} du, \]  \hfill (22)
where \( C \) denotes the time duration of a cycle, which is the sum of the informed and the subsequent uninformed intervals. Hence, (22) is the integral of the information distance in a stationary cycle, divided by the mean cycle time.

If the joint density of \( G \) and the corresponding phase transition is defined by matrix \( g(x) = -\frac{d}{dx} G(x) \), \( E(C) \) is calculated as
\[ E(C) = \int_{x=0}^{\infty} \int_{y=0}^{\infty} \frac{x+y}{v} g(x)e^{D_0y}D_1 dy dx + \int_{x=0}^{\infty} \int_{y=0}^{\infty} yz g(x)e^{D_0y}D_1 dy dx \]
\[ = \frac{1}{v} \int_{x=0}^{\infty} xz g(x) P_1 dx + \int_{y=0}^{\infty} g(x) dy \quad D_1 \hat{1} \quad (D_0)^{-1} \quad E(G) + E(H)/v, \]  \hfill (23)

where \( y \) represents the duration of the uninformed interval and \( x \) is the cluster length in a specific cycle, and \( g(x)e^{D_0y}D_1 \hat{1} \) is their joint pdf.

The numerator of (22) can be calculated as
\[ \int_{0}^{C} (D(u) - R) du = \int_{x=0}^{\infty} \frac{x^2}{2v} dx = \frac{1}{2v} E(G^2), \]  \hfill (24)
since \( x^2/2v \) is the area below one “triangle” in Figure 5. Substituting (23) and (24) into (22), we get
\[ E(D) = \frac{\int_{0}^{C} (D(u) - R) du}{E(C)} + R = \frac{\frac{1}{2v} E(G^2)}{E(G)/v + E(H)/v} + R, \]
which equals (21).

4.2  The transient analysis of \( D(t) \)

After the analysis of the mean information distance \( E(D) \), we are going to study the transient behavior of the process \( D(t) \). This is one of the most interesting measures when the effect of an event/accident is analyzed since it is important to know how far the alert message gets \( t \) time after the accident.
Since we have a MAP vehicle arrival process, the analysis of the joint behavior of \{D(t), J(t)\} is easier than analyzing D(t) alone. More precisely, we are going to introduce the random variable \( \hat{J}(t) \) as the phase of the MAP at the moment when the cluster present at time \( t \) will leave the accident, and define the joint ccdf \( F_i(t,x) = P(D(t) > x, \hat{J}(t) = i) \), and the corresponding row vector \( F_i(t,x) = [F_i(t,x)] \). The joint probability of being in the uninformed interval and in a certain phase at time \( t \) is given by row vector \( \beta(t) = [\beta_i(t)] \), with elements \( \beta_i(t) = P(D(t) = R, \hat{J}(t) = i) \). Note that, for the latter quantity, we use \( J(t) \) instead of \( \hat{J}(t) \). Hence, in the uninformed intervals, \( \beta(t) \) follows the evolution of the background Markov chain of the MAP, and when a vehicle enters the range of the accident, we let the MAP generate all the vehicles of the cluster immediately and freeze its phase at the end of the cluster. Hence, as long as the cluster has not left the accident yet, \( F_i(t,x) \) will characterize the information distance and the MAP phase at the end of the current cluster.

**Theorem 6** The transient ccdf \( F_i(t,x), x > R \) and the probability of an uninformed interval \( \beta(t) \) satisfy the partial differential equations (PDEs)

\[
\frac{\partial}{\partial t} F_i(t,x) - v \frac{\partial}{\partial x} F_i(t,x) = \beta_i(t) D_1 G(x-R), \tag{25}
\]

\[
\frac{\partial}{\partial t} \beta(t) = - \frac{\partial}{\partial t} F_i(t,R) + \beta(t) \langle D_0 + D_1 Z \rangle. \tag{26}
\]

**Proof** To prove the theorem, we describe the evolution of \( D(t) \) in an infinitesimally small time period \((t,t+\Delta)\). Since we have a Markovian arrival process, the events to consider must include the phase transitions, too, both those that generate an arrival and those that do not generate any. There are two possibilities leading to \( D(t+\Delta) > x \) and \( \hat{J}(t+\Delta) = i \), for \( x > R \) as follows:

- At time \( t \), there was an informed cluster on the highway already, which moved towards the accident, by a distance of \( v \Delta \). Since \( \hat{J}(t) \) is the phase at the end of the current informed cluster, we have that \( \hat{J}(t+\Delta) = \hat{J}(t) \) in this case (as the current cluster remained the same).
- Or, at time \( t \), the system was in an uninformed interval \((D(t) = R)\) in phase \( \hat{J}(t) = j \), and the MAP has generated an arrival (a vehicle has entered the range of the accident) in \((t,t+\Delta)\), forming a new cluster of informed vehicles. The probability of an arrival in \((t,t+\Delta)\) accompanied by a phase transition from \( j \) to \( k \) is \([D_1]_{jk} \Delta + o(\Delta)\), and the length of the cluster ending in phase \( i \) is given by ccdf \( G_{ki}(x-R) \).

The probability of having multiple events in \((t,t+\Delta)\) is \( o(\Delta) \), for which \( \lim_{\Delta \to 0} o(\Delta)/\Delta = 0 \) holds. Based on these two possibilities, we have that

\[
F_i(t+\Delta,x) = F_i(t,x + v\Delta) + \sum_{k=1}^{N} \sum_{j=1}^{N} \beta_j(t) [D_1]_{jk} \Delta G_{ki}(x-R) + o(\Delta).
\]

With simple algebraic manipulations, we get

\[
\frac{F_i(t+\Delta,x) - F_i(t,x)}{\Delta} = -v \frac{F_i(t,x + v\Delta) - F_i(t,x)}{v\Delta} = \sum_{k=1}^{N} \sum_{j=1}^{N} \beta_j(t) [D_1]_{jk} G_{ki}(x-R) + o(\Delta)/\Delta,
\]

which, tending \( \Delta \) to 0 and switching to matrix notations, is equal to (25). To derive (26), we have to investigate how it is possible to be in an uninformed interval and phase \( i \) at time \( t + \Delta \). There are three cases, as follows:

- The system was in an uninformed interval and phase \( i \) at time \( t \) already, and no phase transitions occurred in the MAP. The probability of the latter event is \( 1 + [D_0]_{ii} \Delta + o(\Delta) \).
The system was in an uninformed interval and phase \( j \) at time \( t \), and there was an internal phase transition in the MAP from phase \( j \) to \( i \), meaning that no vehicles arrived to the range of the accident, with probability \( [D_0]_{ji} \Delta + o(\Delta) \).

At time \( t \), there was a cluster of informed vehicles on the highway (ending with phase \( i \)), which has left the accident in \((t, t + \Delta)\). Hence, a new uninformed interval begins at time \( t + \Delta \). The probability of this event is \( F_i(t, R) - F_i(t, R + v\Delta) \).

Putting all parts together leads to

\[
\beta_i(t + \Delta) = \beta_i(t)(1 + [D_0]_{ii} \Delta) + F_i(t, R) - F_i(t, R + v\Delta) + \sum_{j \neq i}^{N} \beta_j(t)[D_0]_{ji} \Delta
\]

that can be transformed to

\[
\frac{\beta_i(t + \Delta) - \beta_i(t)}{\Delta} = v \frac{F_i(t, R) - F_i(t, R + v\Delta)}{v\Delta} + \sum_{j = 1}^{N} \beta_j(t)[D_0]_{ji},
\]

which, after taking the limit \( \Delta \to 0 \) and using matrices, yields

\[
\frac{d}{dt} \beta_i(t) = -v \frac{d}{dx} F(t, x)|_{x=R} + \beta(t)D_0.
\]

Observing that the derivative with regards to \( x \) on the right-hand side can be expressed using the time derivative based on (25) establishes the theorem.

### 4.3 Stationary analysis

The stationary behavior can be obtained by taking the limit \( D = \lim_{t \to \infty} D(t) \). The following theorem provides the stationary solution of the phase-dependent ccdf of the information distance, denoted by \( F(x) = \lim_{t \to \infty} F(t, x) \), using the stationary phase-dependent ccdf of the cluster length \( G(x) \).

**Theorem 7** For \( x > R \), the stationary phase-dependent ccdf of the information distance is given by

\[
\frac{d}{dx} F(x) = -\frac{1}{E(G) + E(H)} v G(x - R),
\]

and, for \( x = R \), the stationary probability vector of an uninformed interval is the solution of the system of linear equations

\[
\hat{p}(D_0 + D_1 Z) = 0, \quad \hat{p} \mathbb{1} = \frac{E(H)}{E(G) + E(H)}.
\]

**Proof** Taking the limit \( t \to \infty \) in (25) gives

\[
-v \frac{d}{dx} F(x) = \beta D_1 G(x - R),
\]

and taking the limit in (26) leads to

\[
0 = \hat{p}(D_0 + D_1 Z).
\]

In the latter equation, observe that matrix \( D_0 + D_1 Z \) defines a continuous-time Markov chain that characterizes the phase process of the MAP in uninformed intervals. The MAP evolves according to matrix \( D_0 \), and whenever a
vehicle arrives to the accident (matrix $D_1$), the phases are immediately adjusted by matrix $Z$ to reflect the end-of-cluster phases. The stationary solution of this Markov chain provides the phase distribution in the uninformed intervals. From the definition of $E(H)$ and $E(G)$, the stationary (phase-independent) probability of the uninformed intervals, $\beta$, is $E(H)/(E(G) + E(H))$, proving (28).

On the other hand, we can observe that vector $D_1$ is proportional to the phase distribution right at the end of the uninformed intervals when a new cluster gets formed. The same phase distribution is given by vector $G$, too, the only difficulty is to find the scaling factor between them, that normalizes $D_1$. Since exactly one cluster is formed at the end of each uninformed period (that lasts for $E(H)/v$ in time), we have that the scaling factor is $v/E(H)$; thus, (29) becomes

$$\frac{d}{dx} E(x) = -\beta D_1 G(x-R)/v = \frac{E(H)}{E(G) + E(H)} \frac{1}{E(G) + E(H)} - G(x-R),$$

that proves (27).

**Theorem 8** The phase independent cccdf of the information distance, $F(x) = E(x)1$, can be expressed by

$$F(x) = \frac{\int_0^{x-R} G(y) dy}{E(G) + E(H)} + \frac{E(G)}{E(G) + E(H)},$$

for $x > R$. For $x \leq R$, we have that $F(x) = 1$.

**Proof** Multiplying (27) by $1$ from the right leads to scalar ODE

$$\frac{d}{dx} F(x) = -\frac{G(x-R)}{E(G) + E(H)}.$$

Taking the integral of both sides from $y=0$ to $\infty$ gives

$$F(y) = -\int_{x=0}^{\infty} G(x-R) dx + c,$$

where, from $F(R) = P(D > R)$, for the constant, we have that $c = E(G)/(E(G) + E(H))$ (see Figure 5), yielding (30).

Theorem 5 provides a way to calculate the mean information distance $E(D)$. The same quantity can be derived from Theorem 8 as well, if we take the integral of (30) as follows:

$$\int_{x=R}^{\infty} F(x) dx = \frac{1}{E(G) + E(H)} \left( \int_{x=0}^{\infty} \left( -\int_{y=0}^{x-R} G(y) dy + E(G) \right) dx \right),$$

where the integral term in the parenthesis simplifies to

$$(*) = \int_{x=R}^{\infty} \left( -\int_{y=0}^{x-R} G(y) dy + E(G) \right) dx = \int_{y=0}^{\infty} \int_{x=R}^{x-R} G(y) dy dx$$

$$= \int_{y=0}^{\infty} \int_{x=R}^{y+R} G(y) dx dy = \int_{y=0}^{\infty} yG(y) dy = E(G^2)/2,$$
which is in line with (21).

### 4.4 Speed of the information propagation

Making use of the transient distribution, another interesting study can be carried out: the analysis of the speed of the information propagation. From the ccdf $F(t,x)$, the mean information distance at time $t$ can be obtained as

$$E(D(t)) = R + \int_{R}^{\infty} F(t,x) \, dx. \quad (32)$$

Hence, taking the integral of both sides of (25), and multiplying it by 1 from the right, we get that

$$\frac{d}{dt} (E(D(t)) - R) - v \left[ F(t,x) \right]_{R}^{\infty} = \beta(t) D_1 E(G) 1,$$

which leads to

$$\frac{d}{dt} E(D(t)) = \beta(t) D_1 E(G) 1 - v. \quad (33)$$

Equation (33) is easy to interpret and is completely reasonable. The term $-v$ means that the informed cluster moves towards the accident at speed $v$; hence, the information distance decreases at this speed. At the other hand, $\beta(t) D_1$ is the rate of new cluster formation (the rate of a vehicle arrival in an uninformed period), when the information distance jumps up to $E(G) 1$.

The negativity of $\frac{d}{dt} E(D(t))$ is not obvious from the formula. Still, it follows from the behavior of the system, namely that in case of constant speed, only a single informed cluster can exist, and a new one can be formed only after the previous one has left the accident.

### 5 NUMERICAL EXAMPLES

In this section, we investigate the impact of the statistics of the vehicle interarrival times on the cluster size $G$ and the information distance $D(t)$. Our implementation is based on Matlab, and we use the BuTools library for obtaining the MAPs. To validate the correctness of the presented analytical methods, we have developed a custom simulation tool that gives the exact same results in all of the studied cases.

In all of the numerical examples, the vehicle speed is assumed to be $v = 36 m/s$ (which is the typical speed limitation on highways in many countries), and the radio coverage of the communication is $R = 150 m$.

In Sections 5.1 and 5.2, we demonstrate the importance of using appropriate traffic models by showing the impact of the squared coefficient of variation and the lag-1 correlation coefficient on the mean cluster length and the mean

![Figure 6](image_url)
information distance. Section 5.3 studies the transient distribution of the message propagation distance, and Section 5.4 presents some results with real traffic data as well.

### 5.1 Analysis of the cluster length

We have used the procedure of Horvath\(^{26}\) to obtain matrices \(D_0\) and \(D_1\) from the mean arrival rate, the SCV, and the lag-1 correlation, \(\rho\), of the headway times. With several combinations of these statistical parameters, the mean and the SCV of the cluster length \(G\) were computed based on Corollaries 1 and 2.

Setting the mean arrival rate to a low value leads to very short clusters that mostly consist of isolated vehicles. On the other hand, by setting it too high, almost all vehicles will be connected into a single cluster. We found that \(\lambda = 0.65\) (vehicles/second) represents a car density that falls between the two extreme cases, through which the cluster length is worth examining. The SCV of the headway times has been varied in the range of 0.5 (close to deterministic) to 5 (very bursty), where \(SCV=1\) corresponds to the case of the Poisson vehicle arrival process. As for the lag-1 correlation parameter \(\rho\), we studied the effect of no correlation (\(\rho = 0\)), negative correlation (strong negative correlation \(\rho = -0.3\) and moderate negative correlation \(\rho = -0.19\)), and positive correlation (strong positive correlation \(\rho = 0.85\) and moderate positive correlation \(\rho = 0.19\)) as well. Of course, our procedure can handle the full range of these parameters, that is, \(SCV > 0\) and \(-1 < \rho < 1\), the only limitation being that \(\rho > -1/SCV\) must be respected.\(^{26}\)

The results are shown in Figure 6. As visible from the figure, the statistics of the vehicle arrival process has a significant impact on the cluster length statistics. In general, the lower the SCV is, the longer the clusters are, since the variability of the headway times is lower. The combination of high correlation and low SCV leads to the longest clusters, in this case the headway times of many subsequent vehicles are almost identical. When the SCV is higher, the difference between the curves is smaller, but still significant (observe that the scale of the y axis is logarithmic). According to the plot on the right side of Figure 6, the squared coefficient of variation of the cluster length also depends on the SCV of the headway times.

We note that there are some short intervals in Figure 6 where the plots are not perfectly smooth. The reason for this effect is that the procedure we used for creating the MAPs\(^{26}\) has changed the size of the MAPs around these points.

The numerical solution of the DDE defined by Theorem 4 makes it that it is possible to investigate the ccdf of the cluster length as the function of the SCV as well. According to Figure 7, the cluster length is always greater than or equal to \(R\), due to the definition of the system. It is clearly visible that the burstiness of the vehicle flow has a significant impact on the length of clusters of informed vehicles. When SCV is small (more regular headway times), the informed clusters are longer, while higher SCV leads to shorter clusters.

### 5.2 Analysis of the information distance

The main target of the paper is the analysis of the information distance, that is, how far a message (e.g., an alert message) can propagate in the stationary state. On the left side of Figure 8, the mean information distance is depicted assuming different SCV and correlation values, based on Theorem 5. As known from earlier results,\(^{18}\) the mean information distance \(E(D)\) and the mean cluster length \(E(G)\) are equal for the Poisson case, when \(SCV=1\) and \(\rho = 0\). In all the other cases, \(E(D) \neq E(G)\) does not hold. Nevertheless, comparing Figures 6 and 8, it is clear that the difference between \(E(D)\) and \(E(G)\) is small.

![FIGURE 7 The ccdf of the cluster length distribution \(G(x)\) with various SCV parameters](image-url)
According to the right side of Figure 8, higher SCV of the headway time leads to lower message propagation distance, and high correlation decreases the message propagation distance even more. The plots were obtained by the numerical solution of the ODE defined in Theorem 8.

5.3 | Transient analysis

The results of Section 4.2 make it possible to investigate the time-evolution of the message propagation distance, $D(t)$. As a demonstration, we study a case when an accident occurs at time $t=0$, when the state of the MAP is stationary. We assume that no vehicles are informed about the accident have $\beta(0) = 0$ and $F(0, x) = 0, x > R$. From this starting point, the ccdf $F(t,x)+\beta(t)$ can be obtained by the numerical solution of the ODE given in Theorem 6. (For the solution, the step size along the time axis was 1, and along the distance axis it was $1/v$).

The results are visualized in Figure 9 as a heat map. In line with the expectations, the information distance is always at least $R$, and the more time elapses since the accident, the higher the probability is that vehicles farther away from the accident receive the message about it. After $t=300$ (that is, 5 min), the stationary state is almost reached, and the distribution of $D(t)$ does not change significantly.

5.4 | Experiments with real data

We did experiments with real data as well. Two factors make such a study difficult.

| Method  | $E(\hat{g})$ | $E(\hat{g}^2)$ | $E(\hat{h})$ | $E(D)$ |
|---------|--------------|----------------|--------------|---------|
| Simulation | 495          | 8.19134 + 6   | 290          | 5375    |
| Poisson model | 248.6306     | 8.0594e + 04  | 159.9331     | 248.6306|
| MAP model   | 489.4438     | 5.9635e + 06  | 282.2731     | 4,013   |
The first difficulty is that there are very few high-quality headway data traces available publicly. The majority of traffic data contains counts over a certain period of time (e.g., number of vehicles detected in an hour), which is not suitable for MAP fitting. For our procedure, we need the exact arrival times of the vehicles (or equivalently, all headway times). The only relevant data set we found was Coifman and Li,20 which was based on LIDAR measurements (a cheaper method that does not need LIDARs has been published in Zhe et al32). While this is a fairly large data set, it is still not long enough, since the treatment of such seasonal traffic measurements needs a lot of data. Hence, we decided to ignore the seasonal nature of the traffic and cut out a part of the data consisting of around approximately 629,000 samples where the traffic was approximately stationary.

The second difficulty is that fitting MAPs needs a lot of data. Capturing the characteristics of the density function of the marginal distribution can be accurate with less data, too, but for correlation fitting, especially for higher lags, much more data is needed. This observation is reflected by Figure 1, too, where the density is fitted well and the lag-correlations are not matching as well. As more headway time data become publicly available, it will be possible to fit MAPs that better represent the real traffic, making our analytical model more accurate.

Based on the approximately 629,000 samples extracted from the data set,20 we executed the EM-algorithm published in Horváth and Okamura27 to create a MAP with 800 states. Even with that many states, the formulas presented in this paper give instant results. Table 2 compares these analytical results with the simulation results driven by the original measurement data. According to the results, the mean cluster length $E(\mathcal{G})$ is obtained very accurately; the error (the deviation from the simulation results) is below 2%. The Poisson assumption, commonly used in the literature,9,18 gives almost 100% error. The same holds for the mean noninformed periods $E(\mathcal{H})$ as well. However, for the second moment of the cluster length $E(\mathcal{G}^2)$, our method has a significant error, due to the imperfect MAP fitting caused by the overly small data set. Still, the MAP model-based results are not far away from the simulation results, as opposed to the Poisson model-based results, where there is a 100-times difference. The inaccuracy in $E(\mathcal{G}^2)$ implies inaccuracy in $E(\mathcal{D})$, too. Our method gives 4,013 m for the mean message propagation distance instead of 5,375 m, but it is still much better than the Poisson result with 248.6 m.

We believe that, as more data gets available and more mature MAP fitting methods get developed, the practical relevance of our procedure is going to improve in the future.

6 CONCLUSION

The behavior of the message propagation in VANET systems is affected by the traffic model of the vehicles. In this paper, we consider the Markovian arrival process as a traffic model and derive various results related to message propagation. We have derived the moments and the ccdf of the stationary cluster length and the stationary and transient distribution of the information distance. We validated our analytical results with simulation. We conclude that, when the Poisson assumption fails to model the traffic, the Markov arrival process can be a remedy to approximate the message propagation distance in practice.

ACKNOWLEDGEMENTS

Dhari Ali Mahmood would like to thank the Tempus Public Foundation (TPF) – Stipendium Hungaricum program and University of Technology – Iraq for the support for his PhD scholarship. This work is supported by the OTKA K-123914 project and by the National Research, Development and Innovation Fund (TUDFO/51757/2019-ITM, Thematic Excellence Program).

ORCID

Gabor Horvath https://orcid.org/0000-0003-3097-1273

REFERENCES

1. Cunha F, Villas L, Boukerche A, Maia G, Viana A, Mini RA, Loureiro AA. Data communication in VANETs: protocols, applications and challenges. Ad Hoc Netw. 2016;44:90-103.
2. Al-Sultan S, Al-Doori MM, Al-Bayatti AH, Zedan H. A comprehensive survey on vehicular ad hoc network. J Netw Comput Appl. 2014;37:380-392.
3. Cunha F, Villas L, Boukerche A, Maia G, Viana A, Mini R. A. F., Loureiro A. A. F.. Data communication in VANETs: Protocols, applications and challenges. Ad Hoc Networks. 2016;44:90–103. http://dx.doi.org/10.1016/j.adhoc.2016.02.017
4. Etsi T. 102 638: Intelligent transport systems (its); vehicular communications; basic set of applications (BSA). Sophia Antipolis, France. 2009:1.1.1:1-81.

5. Waldeer KT. A vehicular traffic flow model based on a stochastic acceleration process. Transp Theory Stat Phys. 2004;33(1):7-30.

6. Kosun C, Ozdemir S. A superstatistical model of vehicular traffic flow. Physica A: Stat Mech Appl. 2016;444:466-475.

7. Nafi NS, Hasan MK, Abdallah AH. Traffic flow model for vehicular network. In: 2012 International Conference on Computer and Communication Engineering, ICCCE 2012; 2012:738-743.

8. Zhou H, Xu S, Ren D, Huang C, Zhang H. Analysis of event-driven warning message propagation in vehicular ad hoc networks. Ad Hoc Netw. 2017;55:87-96.

9. Miorandi D, Altman E. Connectivity in one-dimensional ad hoc networks: a queueing theoretical approach. Wirel Netw. 2006;12(5):573-587.

10. Rostamzadeh K, Gopalakrishnan S. Analysis of message delivery delay in vehicular networks. IEEE Trans Vehic Technol. 2015;64(10):4770-4779.

11. Frohlich K, Neuts MF. Modelling vehicular traffic using the discrete time Markovian arrival process. Transp Sci. 1995;29(2):109-117.

12. Farkas C, Telek M. Capacity planning of electric car charging station based on discrete time observations and MAP(2)/G/c queue. Periodica Polytechnica Electr Eng Comput Sci. 2018;62(3):82-89.

13. Zhao L, Yan L, Chen KH, Sung DK. Fast, efficient broadcast schemes based on the prediction of dynamics in vehicular ad hoc networks. IEEE J Selected Areas Commun. 2007;25(8):1538-1556.

14. Alfa AS, Neuts MF. Modelling vehicular traffic using the discrete time Markovian arrival process. Transp Sci. 1995;29(2):109-117.

15. Seliem H, Shahidi R, Ahmed MH, Shehata MS. On the end-to-end delay in a one-way VANET. IEEE Trans Vehic Technol. 2019;68(9):8336-8346.

16. Babu A, Naheed Ajeer V. Analytical model for connectivity of vehicular ad hoc networks in the presence of channel randomness. Int J Commun Syst. 2013;26(7):927-946.

17. Zhuang Y, Pan J, Cai L. A probabilistic model for message propagation in two-dimensional vehicular ad-hoc networks. In: Proceedings of the seventh ACM international workshop on Vehicular Networking and Security. 2010:31-40.

18. Mahmood DA, Horvath G. Analysis of the message propagation on the highway in VANET. Arabian J Sci Eng. 2019;44(4):3405-3413.

19. Friis HT. A note on a simple transmission formula. Proc IRE. 1946;34(5):254-256.

20. Cox DR. Renewal theory. The University of Michigan: Methuen; 1962.

21. Horvath G, Telek M. BuTools 2: a rich toolbox for Markovian performance evaluation. In: proceedings of the 10th EAI International Conference on Performance Evaluation Methodologies and Tools; 2013:119-133.

22. Horvath G, Okamura H. A fast EM algorithm for fitting marked Markovian arrival processes with a new special structure. European Workshop Perform Eng. 2013;8168:119-133.

23. Okamura H, Dohi T. Faster maximum likelihood estimation algorithms for Markovian arrival processes. In: 2009 Sixth international conference on the quantitative evaluation of systems; 2009:73-82.

How to cite this article: Mahmood DA, Horvath G. Analysis of alert message propagation on the highway in VANET assuming Markovian vehicle arrival process. Int J Commun Syst. 2020;33:e4501. https://doi.org/10.1002/dac.4501