

# Traffic Management and Networking for Autonomous Vehicular Highway Systems

Izhak Rubin<sup>a</sup>, Andrea Baiocchi<sup>b</sup>, Yulia Sunyoto<sup>a</sup>, Ion Turcanu<sup>b,c</sup>

<sup>a</sup>*Electrical and Computer Engineering Department,  
University of California Los Angeles (UCLA), Los Angeles, CA, USA*  
<sup>b</sup>*Dept. of Information Engineering, Electronics and Telecommunications,  
University of Rome Sapienza, Italy*  
<sup>c</sup>*Interdisciplinary Centre for Security, Reliability and Trust (SnT),  
University of Luxembourg, Luxembourg*

---

## Abstract

We develop traffic management and data networking mechanisms and study their integrated design for an autonomous transportation system. The traffic management model involves a multi-lane multi-segment highway. Ramp managers regulate admission of vehicles into the highway and their routing to designated lanes. Vehicles moving across each lane are organized into platoons. A Platoon Leader (PL) is elected in each platoon and is used to manage its members and their communications with the infrastructure and with vehicles in other platoons. We develop new methods that are employed to determine the structural formations of platoons and their mobility processes in each lane, aiming to maximize the realized flow rate under vehicular end-to-end delay constraints. We set a limit on the vehicular on-ramp queueing delay and on the (per unit distance) transit time incurred along the highway. We make use of the platoon formations to develop new Vehicle-to-Vehicle (V2V) wireless networking cross-layer schemes that are used to disseminate messages among vehicles traveling within a specified neighborhood. For this purpose, we develop algorithms that configure a hierarchical networking architecture for the autonomous system. Certain platoon leaders are dynamically assigned to act as Backbone Nodes (BNs). The latter are interconnected by communications links to form a Backbone Network (Bnet). Each BN serves as an access point for its Access Network (Anet), which consists of its mobile clients. We study the delay-throughput performance behavior of the network system and determine the optimal setting of its parameters, assuming both TDMA and IEEE 802.11p oriented wireless channel sharing (MAC) schemes. Integrating these traffic management and data networking mechanisms, we demonstrate the performance tradeoffs available to the system designer and manager when aiming to assure an autonomous transportation system operation that achieves targeted vehicular flow rates and transit delays while also setting the data communications network system to meet targeted message throughput and delay objectives.

---

## 1. Introduction

Autonomous driving is taking momentum at a vastly growing pace. Early implementations for various applications, such as taxi services [1], are being introduced. A full understanding of large scale deployment of systems of autonomous vehicles is a key component of a safe growth of the Intelligent Transportation Systems (ITS) as a critical component in the development of a paradigm for the architecture of the next generation smart city. A wide range of mechanisms, technologies and regulation procedures need to be developed and enacted under this perspective, including technical, legal, economic and social impact factors.

Among the multitude of technical aspects that dominate the design, management and control of high performance autonomous driving systems, the following two

elements are of paramount importance: vehicular traffic management and data networking – the latter involving communications and data dissemination among vehicles, i.e. Vehicle-to-Vehicle (V2V), and between vehicles and fixed infrastructure, i.e. Vehicle-to-Infrastructure (V2I). A common characteristic of these traffic management and data communications networking elements is that they address coordination and reliable interactions among distributed entities, moving in space and time, at a potentially large scale. For the specific case of autonomous driving, the inter-dependence of those two components are stronger than ever. High performance reliable communications networking of data messages among moving vehicles is a key ingredient in enabling the learning by autonomous vehicles of their surrounding environment, the rapid reaction to critical events assure the safety of the operation, coordination among vehicles to synchronize mobility and maneuvering. They also aid in the execution of processes that serve to optimize travel safety, reduce transit delays, disseminate sensed data reports to nearby vehicles that include envi-

---

*Email addresses:* rubin@ee.ucla.edu (Izhak Rubin),  
andrea.baiocchi@uniroma1.it (Andrea Baiocchi),  
yulia.sunyoto@ucla.edu (Yulia Sunyoto), ion.turcanu@uni.lu  
(Ion Turcanu)

ronmental and congestion conditions, uploading data to remote cloud servers, and enabling remote monitoring, supervision, management and control of the autonomous vehicular system.

As for vehicular traffic management, the formation of a dynamically controlled intelligent autonomous transportation system that properly manages autonomous vehicles leads to a host of applications that will induce a paradigm shift in transportation systems.

While human controlled vehicles are often guided to make mobility decisions based on traffic advisories, the final decision as to the route, the driving style and the speed to use is in human hands, leading to wide variability. In turn, an autonomous driving system can be managed to produce a controlled high performing mobility operation for each vehicle and for the overall transportation system. Such an intelligent regulation can be used to meet the travel objectives of individual drivers, yield a much improved transportation safety and efficiency, enhanced vehicular throughput rates, reduced travel latencies and lower polluting emissions and/or improved utilization of energy resources.

In this paper, we consider a highway populated with autonomous vehicles [2]. We study the optimized design of vehicular traffic flows along a highway by regulating vehicles that move in each lane into platoon formations. Such formations are advantageous in that an elected vehicle within the platoon, identified as the Platoon Leader (PL), can be used to rapidly and efficiently regulate mobility and maneuvering of its platoon vehicular members. Furthermore, we make use of these formations in synthesizing a dynamically adaptive V2V networking architecture. We develop models that provide system design guidelines and that demonstrate the performance tradeoffs characterizing the system when integrating the embedded traffic management and data communications networking operations.

The major innovative contributions of this work are summarized as follows:

- We state and solve optimization problems that aim to maximize the achievable vehicular throughput rate carried by the highway, under a vehicular delay constraint. We account for on-ramp queueing delays as well as for transit delay across highway lanes. We determine the best settings to be used in forming platoons along each lane, including parameters that account for vehicular speeds and inter-vehicular spacing levels. We also determine the proper scheme to be used in routing admitted vehicles into faster or slower lanes.
- We investigate the design of a V2V data communication networking system that is aided by the formation of autonomous vehicular platoons along the highway. For this purpose, we introduce a two layer hierarchical network architecture that consists of a dynamically formed Backbone Network (Bnet) and Access Networks (Anets). The Bnet is dynamically synthesized

through the election of PLs that are set to serve as Backbone Nodes (BNs). Each BN manages its Anet, which consists of its mobile clients. We examine the use of mixes of candidate Medium Access Control (MAC) schemes across the Bnet and the Anets, in considering Time-Division Multiple Access (TDMA) and IEEE 802.11p based protocols. We carry out extensive analyses to determine the optimal cross-layer parameter settings for each candidate scheme, determining the highest data communication throughput levels that are attainable in each case, under specified packet delay limits. To demonstrate the underlying design options, we assume data applications that require the packets generated by a vehicle that is active in producing a data flow to be disseminated across a specified distance span.

- We study the tradeoffs involved in integrating the mechanisms developed by us to induce an effective traffic management operation that strives to produce high vehicular flows along the highway under vehicular delay constraints, and those developed by us to attain a high capacity V2V data communications network system.

The remainder of the paper is organized as follows. In Section 2, we present an overview of related works and discuss the motivation for platoon-based traffic management and for the employed networking schemes. The traffic management system model is described in Section 3. The associated system optimization problems and solutions for the traffic management system are presented in Section 4. The architecture, schemes and protocols used for dissemination of data flows across the V2V data communications network system are presented in Section 5. In Section 6, we present the parameters employed by the network systems as well as define the metrics used to characterize the performance behavior of the networked system. Performance behavior characteristics for the data communications network systems under consideration are presented in Section 7. In Section 8, we discuss and illustrate the design tradeoffs available to the system manager when traffic management and data networking performance objectives are combined. Conclusions are drawn in Section 9.

## 2. Related Work

The impact of vehicular platoon applications on traffic flows and road safety have attracted significant interest. Published works have shown that organizing vehicles in platoons can improve traffic flows and reduce fuel consumption [3, 4, 5], while increasing safety and enhancing driving experience [6, 7, 8]. Generally, platoon-related published works can be divided in two main categories: (i) studies that focus on traffic management and flow optimization by considering such problems as platoon lane assignment, speed control and on-ramp admission, and (ii) works that

study the impact of the network communication technology on platoon stability and vice versa. A summary of related works is presented in Table 1.

### 2.1. Network communication schemes

With respect to the synthesis of the data communications network, a common approach found in the literature is to dynamically select a subset of vehicles to act as relay nodes in forwarding data messages, serving to reduce the network data traffic load [9, 10, 11, 12, 13, 14]. In [9], the authors describe a high throughput dissemination scheme for Vehicular Ad Hoc Networks (VANETs) based on dynamic formation of a multi-hop backbone network. Analysis and design is performed of such a networking protocol when used to broadcast message flows that are generated by a Roadside Unit (RSU) along a linear road, forming a VANET structure that is identified as a Vehicular Backbone Network (VBN). Please see this paper for references to other studies that develop clustering based networking schemes for non-autonomous vehicular systems.

The VBN scheme, originally developed by Professor Izhak Rubin based on his development of the Mobile Backbone Networks (MBNs) scheme for use by mobile ad hoc wireless networks, is based on using vehicles that reside at preferred locations to act as backbone relay nodes. While an MBN involves the mobility of vehicles over a two-dimensional plane, or over a three-dimensional space when Unmanned Aerial Vehicle (UAV)-aided, a VBN involves mobility of vehicles across a multi-lane linear single-dimensional road. Consequently, a more powerful networking scheme has been developed for the latter. The networking scheme developed in this paper inherits key features of the VBN scheme, yet it modifies it by incorporating new mechanisms that are enabled by the autonomous mode of operation, leading to simplified implementation and enhanced performance behavior of its data networking operation.

As noted above, under the VBN approach, vehicles that reside at preferred locations are elected to act as backbone relay nodes. These relay nodes are optimally configured to operate at designated link data rates, modulation/coding schemes and spatial reuse factors. Significant performance upgrade is achieved under a VBN architecture in attaining enhanced throughput capacity rate and in providing for a wider dissemination coverage span. Admitted packets are assured low end-to-end packet delays through the employment of a flow control scheme. Furthermore, distinct savings are attained in energy consumption, as only selected relay nodes that act as backbone nodes need to engage in data packet forwarding processes.

In [10], a topology synthesis algorithm is presented for the synthesis of the Bnet for an MBN based mobile ad hoc network. The algorithm is shown to be highly scalable and fast converging to a solution of an effective connected backbone network that consists of elected mobile nodes. The use of multicast protocols over a mobile ad hoc network is discussed and studied in [11]. The advantage of

multicasting packet flows through their dissemination over a MBN-oriented synthesized backbone is well demonstrated.

In [12], the VBN concept is introduced, presenting our approach to a GPS-aided synthesis of a cross-layer backbone. The study in [13] provides performance analysis and design of the VBN system, as it is optimally configured to broadcast message flows from a source RSU to highway vehicles using flow pacing mechanisms. It also shows that under such flow regulation, an IEEE 802.11p MAC scheme can achieve throughput rate performance that is similar to that attained under a TDMA scheme. A VBN-based scheme is also employed in [14], where the authors exploit the created backbone network to periodically disseminate and collect floating car data from vehicles roaming inside an area of interest.

For platoon-based communications networking studies, whereby a TDMA wireless access scheme is used, we note the following references. In [15], a TDMA scheme is employed to coordinate the access of vehicles that travel in platoon formations. Time slots are arranged in accordance with the relative distance of the message source to the location of an accident. Vehicles that are located closer to an accident location are granted earlier slots, aiming to reduce data message transport latency. An interference reduction mechanism is also illustrated. The employed protocol uses hop-by-hop routing within a platoon. No use is made of platoon structures to execute inter-platoon communications.

In [3], highway vehicles traveling across a single lane highway are arranged in platoon formations. Platoon formation parameters include inter-platoon and intra-platoon distances, vehicular speeds and the number of vehicles included in a platoon. The desired values of these parameters are determined in relation to performance considerations that are based on the trade-offs between two categories of performance metrics: realized vehicular throughput rate and underlying V2V wireless network performance. A vehicle that produces a packet flow aims to disseminate its packets across a span of the highway whose scope is specified in accordance with the engaged application type. For critical event-driven (Class I) packet flows, the key performance metric is the packet end-to-end dissemination delay. For periodic status-update type (or other stream types of) packet flows (Class II), the attained message throughput rate serves as the key metric. The authors employ a spatial-reuse TDMA protocol, using hop-by-hop communications among vehicles inside and outside a platoon. Multiple information sources can be simultaneously active. Optimal platoon formations are determined, aiming to preserve Class I message delays prescriptions while attaining high Class II message throughput rates and high vehicular throughput rates. Only vehicles that are members of the same platoon are assumed to be synchronized.

In contrast, we assume in this paper, that when TDMA-based transmissions between platoons are executed, time synchronization information is made available to platoon leaders (e.g., by using a Control Channel (CCH) or a

| Focus                                  | Approach                           | References              | Features  |
|--|------------------------------------|-------------------------|---|
| <b>Networking &amp; communications</b> | VBN-based                          | [9, 10, 11, 12, 13, 14] | <ul style="list-style-type: none"> <li>• Focus on dynamic formation of multi-hop vehicular backbone network.</li> <li>• Vehicles residing at preferred locations act as backbone relay nodes and are in charge of handling the network data flows.</li> <li>• The vehicular traffic flow optimization problem is not tackled.</li> </ul>  |
|  | TDMA-based                         | [15, 3, 16]             | <ul style="list-style-type: none"> <li>• Use TDMA wireless access scheme to coordinate the access of vehicles that travel in platoon formations.</li> <li>• Targeted performance metrics include message transport latency and throughput rate.</li> <li>• Only vehicles within the same platoon are assumed to be synchronized.</li> </ul>   |
|  | IEEE 802.11p-based                 | [17, 18, 6]             | <ul style="list-style-type: none"> <li>• Employ IEEE 802.11p schemes to study the impact of different beaconing frequencies, platoon sizes, intra-platoon distances, and various vehicular highway loading levels on the communication channels.</li> <li>• Typical measured metrics are probability of successful transmission and platoon stability.</li> <li>• Communication problems due to inter-platoon coordination are not considered.</li> </ul> |
|  | Hybrid TDMA and IEEE 802.11p-based | [19, 8]                 | <ul style="list-style-type: none"> <li>• Exploit TDMA-like schemes for platoon stability preservation and CSMA/CA for disseminating event-based safety messages.</li> <li>• Vehicular traffic management problems are not considered.</li> </ul>  |
| <b>Traffic flow optimization</b>       |                                    | [4, 20, 5]              | <ul style="list-style-type: none"> <li>• Exploit platoon formation techniques to optimize road network capacities and increase vehicular traffic throughput.</li> <li>• Do not consider the impact of the platoon management on the communication channels.</li> </ul>  |

Table 1: Summary of related works

road side backbone). Also we note that in this paper, we do not impose hop-by-hop transmission of messages among platoon mobiles. In turn, a message transmitted by a platoon leader is broadcasted over a specified range, reaching also its platoon members.

The model mentioned above is extended in [16], using both data networking and vehicular throughput metrics, to include modeling of waiting times incurred by vehicles that form queues across their entrance ramps, prior to their admission to the highway. The paper carries out queuing delay analyses. The employed admission control mechanism is based on having on-ramp waiting vehicles admitted to the highway as they join platoons of vehicles that pass-by the ramp access points, when the latter have platoon membership vacancies.

Relating to papers that employ IEEE 802.11p-based schemes, we note the following studies. The study in [17] presents analytical and simulation models that are used to calculate the probability of successful transmission of beacon messages within a platoon. The study explores the effect on system performance of using different beaconing frequencies, fixed platoon sizes and various vehicular highway loading levels. The minimum beacon rate frequency is determined. This study does not consider the impact of different platoon configurations on data throughput. It also does not involve inter-platoon communications.

The study presented in [18] presents analytical and simulation models for capturing the impact of Coopera-

tive Awareness Messages (CAMs) exchange process, when applied for platooning control. The authors assume both platoon-organized and regular vehicles traveling on a highway and periodically exchanging CAM messages. From a network communication perspective, the probability of a successful transmission is computed under varying traffic density levels. The impact of the communication process on platoon stability is also analyzed. The parameters taken into account include the highway density, intra-platoon distance values, and beacon frequency. The employed analytical model neglects the hidden node terminal problem and considers synchronous communications. This work does not study routing of message flows and optimal platoon formations.

In [6], the authors propose a distributed consensus-based control algorithm for performing vehicular platooning with Inter-Vehicle Communication (IVC). Generally, the coordination of a platoon of vehicles involves the use of a control algorithm that adjusts the relative distance among adjacent vehicles in the platoon, and a communication network that allows vehicles to exchange information. This paper considers the IEEE 802.11p protocol as the main scheme used for executing V2V communications networking. The proposed control algorithm is proved to be robust with respect to variations in logical communications topologies (e.g., predecessor-following, leader and predecessor-following bidirectional), as well as with respect to communication delays and losses. The authors provide

a theoretical analysis of the stability of their proposed algorithm and validate their solution through simulations that consider detailed physical characteristics of vehicular dynamics. In contrast to our paper the inter-platoon coordination and communications problem is not considered.

In noting papers that include the use of TDMA, as well IEEE 802.11p protocols for V2V networking, we observe the following ones. In [19], just two types of communication scopes are involved: broadcasting from a platoon leader to all platoon vehicles, and transmission of a message from a vehicle to its neighboring vehicles. Protocols include slotted beaconing within a platoon (intra-platoon TDMA) and beaconing using Carrier-Sense Multiple Access with Collision Avoidance (CSMA/CA). The number of vehicles on the highway serves as a parameter, under fixed platoon size and intra-platoon distance values. This study also explores the effect of varying the beacon frequency on the minimum intra-platoon distance levels, under targeted data communication delay constraints. The paper does not perform a study on communication between platoons and it does not account for message routing between different platoons. It also does not investigate the setting of platoon configurations for the aim of optimizing communications and/or vehicular throughput metrics, as performed by us in this paper.

The study in [8] presents a consensus-based control algorithm for vehicular platooning and an adaptive message dissemination scheme to guarantee a platoon stability requirement. The study considers a scenario that involves both platoon-organized and individual vehicles traveling along a highway. The theoretical analysis of the proposed control algorithm shows that the state error (position and speed) function incurred between platoon members and platoon leader can converge within a prescribed time bound in presence of beacon packet loss. This bound depends on the platoon size, beacon delivery ratio, and magnitude of the acceleration perturbation component. The proposed message dissemination scheme uses the current channel quality and the leader's dynamics to adaptively select beacon sending time slots for each platoon member. A TDMA-like MAC mechanism is used for intra-platoon beaconing, while a CSMA/CA-based approach is considered for disseminating event-based safety messages. An inter-platoon communication management mechanism is also proposed. The authors perform analytical and simulation-based evaluations to validate the effectiveness of their proposed solution. The traffic management functions studied in our paper and the associated optimization problems are not considered.

## 2.2. Traffic flow optimization schemes

When considering traffic management and flow control schemes that employ platoons, we note the following works. The study in [4] addresses the traffic management problem by exploiting the idea of organizing vehicles into platoons. They propose a vehicular platoon assignment strategy with the objective of maximizing the distance across which platoon formations remain unchanged. Also, they introduce

a lane assignment solution according to which platoons having longer origin/destination distances tend to be assigned to faster lanes. These problems are formulated as optimization problems and are solved by using linear programming based algorithms. Observing simulation results, they note that organizing vehicles into platoons leads to a significant increase in lane capacities and vehicular traffic throughput. Platoon and lane reassignment procedures are repeated every time vehicles approach the start of a new highway segment, which can be challenging when implemented in real-time. The authors assume the presence of inter-vehicular communications component, but do not perform corresponding analysis to determine its impact on the communication channels.

In [20], the authors propose an integrated traffic management and hierarchical traffic control approach for Automated Highway Systems (AHS). They describe a model-based predictive control scheme that determines at which lane and at which speed a platoon should travel within the AHS. They also consider the problem of properly choosing the release times of vehicles that are intending to enter the AHS in order to optimize the overall performance of the traffic system. Their solution is evaluated in terms of total time spent in a traffic network, while considering dynamic speed limits, lane allocation, and on-ramp metering as control measures. This study does not however perform an evaluation of the impact of the system on the performance behavior of the communications network.

The study in [5] explores the idea of using connected vehicle technologies to increase the capacity of an intersection in an urban road system. Three queuing models are considered and analyzed, serving to predict the mobility benefits of organizing vehicles in platoons at intersections. To prove their claims, the authors perform a simulation study that involves a road network near Los Angeles, where vehicles are assumed to have Cooperative Adaptive Cruise Control (CACC) capability and to be organized into platoons. Evaluation results show that the capacity of the road network can be increased by a factor of 3, with no increase in queuing delay or travel time. However, no analysis of the communication network is carried out.

In contrast with the above mentioned studies, we characterize in this paper the optimal setting of vehicular platoon configurations by combining traffic management based vehicular traffic flow rate objectives with data network communications performance behavior aims.

## 3. Traffic Management System Model

We consider a highway system with access/exit ramps and multiple lanes. The highway system is made up of highway span, ranging from one access ramp to the next one, in a given direction. Typical examples are linear highways or ring highways (beltways). Figure 1 illustrates a highway made up of three spans as an example.

To start with, we focus on a unidirectional single-lane highway span, equipped with a single input ramp and a

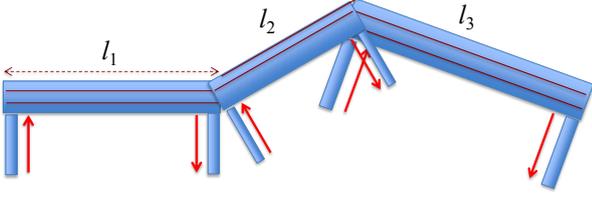


Figure 1: Example of highway span with three input/output ramps.

single output ramp. We refer to the highway span, comprised between the entry point at the input ramp and the exit point at the output ramp, as *highway link*. Later on we will generalize the considered highway to the multiple link multiple lane scenario and introduce further notation accordingly.

Assume a vehicular flow under which vehicles enter at the access ramp and move across the link to the exit ramp, exiting the highway link or continuing to another link thereafter. The vehicular flow entering the highway link is modeled as a Poisson process with mean rate  $\lambda$ . We assume that vehicles are organized in platoons when traveling on the highway. The vehicle that travels at the head of a platoon is identified as the PL.

We introduce the following notation:

- $v_{\max}$  maximum allowable vehicular speed across the highway span.
- $l$  length of the highway link.
- $\lambda$  mean vehicular arrival rate at the entry point.
- $\mu$  mean effective serving rate provided by the highway link; it is defined as the mean rate at which vehicles from the input ramp are admitted into the highway, in joining existing platoons or forming new ones.
- $d_{\text{PL}}$  distance between two successive platoon leaders.
- $d_V$  minimum distance between vehicles that are members of the same platoon.
- $d_P$  minimum distance between platoons, i.e, the minimum distance required between the ending location of a platoon and the starting location of a subsequent platoon; it is set as  $d_P = d_V + g$ , where  $g$  represent an excess gap range, beyond  $d_V$ .
- $g$  excess minimum gap distance between two subsequent platoons; its value is a function of the rate of vehicles that need to merge into and/or out of the lane and/or transit through the lane for the purpose of joining the lane flow or for the purpose of moving across the lane as they maneuver to enter to or exit from other highway lanes; higher such gap margins are typically required for setting platoon formations across slower lanes, as they need to accommodate vehicles that cross the lane as they enter to join flows across a

faster lane, or vehicles that enter the lane from faster lanes as they are in the process of exiting the highway, or switching to another highway.

- $n$  maximum number of vehicles that are allowed to form a platoon.
- $v$  mean velocity of vehicles on the considered highway link.
- $v_{\text{tg}}$  target velocity of vehicles on the considered highway link/
- $s$  mean length of a vehicle.
- $u$  nominal deceleration of a vehicle.
- $\Delta u$  deviation of the deceleration level from its nominal value.

When considering multiple link and/or multiple-lane highways, we add subscripts. A single subscript is used in case of single-lane highway to denote the highway link. When a double subscript is used, the first index refers to the lane, while the second one to the highway link. As a matter of example,  $v_{\text{tg},k}$  is the target velocity on the highway link  $k$  single lane;  $v_{\text{tg},ik}$  is the target velocity of vehicles traveling on lane  $i$  of link  $k$ .

The platoon leader of platoon  $k$ , denoted as  $\text{PL}_k$ , is the leading vehicle in platoon  $k$ . The platoon trailer of platoon  $k$ , denoted as  $\text{PT}_k$ , is the trailing vehicle in platoon  $k$ . The distance spanned between  $\text{PL}_{k-1}$  and  $\text{PL}_k$ , defined as the distance between consecutive platoon leaders,  $d_{\text{PL}}$ , assuming each platoon to have full occupancy of  $n$  vehicles, is given by

$$d_{\text{PL}} = (n - 1)d_V + d_P = nd_V + g = ns + n\zeta \frac{v^2}{2u} + g \quad (1)$$

where the second term in the rightmost side of the equation is calculated by accounting for a safety distance that must be maintained between adjacent vehicles, and using the factor  $\zeta \leq 1$  to account for the range of dispersion of deceleration factor values realized by different vehicles<sup>1</sup>. For notation convenience, we define  $b \equiv \zeta/(2u)$ .

<sup>1</sup>The headway distance used for safety reasons here is set as equal to the braking range, assuming a maximum deceleration of  $u$ . This use is somewhat pessimistic. We could account for the fact that, when a vehicle starts decelerating, it still moves forward and this gives some margin to the subsequent vehicle to brake in turn and avoid a collision. If the decelerations were rigidly constant and reliable, we could theoretically use zero safety gaps, apart from reaction times. However, we have to account for mechanical and performance differences among vehicles. Let  $[u - \Delta u, u + \Delta u]$  be the realized range of decelerations of vehicles. In the worst case, the leading vehicle has the maximum deceleration and the following one has the minimum deceleration. If we aim at guaranteeing the maintenance of a minimum separation  $d_{V,\min}$  between the two vehicles after they have stopped, the initial separation they need is

$$d_V = d_{V,\min} + \frac{v^2}{2u} \frac{2\Delta u/u}{1 - (\Delta u/u)^2}$$

By letting  $d_{V,\min} = 0$  and  $\zeta \equiv 2\Delta u/u/[1 - (\Delta u/u)^2]$  we recover

The time elapsed between the instants of two consecutive platoons passing by a fixed highway location is equal to  $d_{\text{PL}}/v$ . We assume that vehicles are admitted into the highway link at the times at which platoons pass by the access ramp, joining (or forming) platoons. Hence, the maximum “serving rate” of the highway link (realized when every platoon is available to admit into the highway  $n$  vehicles), is given by

$$\mu = \frac{n}{d_{\text{PL}}/v} = \frac{nv}{ns + nbv^2 + g} = \frac{v}{bv^2 + a}.$$

where  $a \equiv s + g/n$ . This equation is a consequence of the basic transportation equation, stating that, under equilibrium conditions, the mean flow equals the product of the mean density and the mean velocity across the road segment. Under our model, the mean density of vehicles along the highway is determined by the configured platoon formations and structures.

The serving rate  $\mu$  is 0 for  $v = 0$ . It grows with  $v$  initially, and then it gradually decreases as  $v$  increases, due to the requirement of maintaining safety spacing ranges between adjacent vehicles.

The key structural parameters of the platoon formation can be further elaborated, by assuming that the parameter  $g$  dictating the excess distance between a platoon trailer and the following platoon leader be a multiple of the inter-vehicle spacing within a platoon (hence, dependent on  $v$ ). We then set  $g = \psi d_V = \psi(s + bv^2)$ . With this setting, the mean serving rate can be written as

$$\mu(v) = \frac{v}{bv^2 + a} = \frac{v}{bv^2 + (1 + \psi/n)s + \psi/n bv^2} = \frac{1}{1 + \psi/n} \frac{v}{bv^2 + s} \quad (2)$$

For fixed  $n$  and  $s$  values, the maximum serving rate is achieved when the velocity dependent component of  $d_{\text{PL}}$  equals the constant component, i.e., when  $s = bv^2$ . Therefore, the maximal value of the serving rate and the velocity at which it is attained are given by

$$\mu^* = \frac{1}{1 + \psi/n} \frac{1}{2\sqrt{bs}}, \quad v^* = \sqrt{\frac{s}{b}}.$$

The parameter  $b$  depends only on the mechanical functionality of the vehicles. In our model,  $b$  is a function of the maximum deceleration parameter  $u$  and the allowance parameter  $\Delta u$  representing an average deviation with respect to the nominal deceleration. Assuming  $u = 3 \text{ m/s}^2$  and  $\Delta u/u = 0.1$ , we find  $b = 0.0337 \text{ s}^2/\text{m}$ . To illustrate, the parameter  $s$  is set to  $s = 6 \text{ m}$ . The resting value of  $v^*$  is  $v^* = 13.3 \text{ m/s}$ .

---

the expression of  $d_V$  used in Equation (1). This formula merely accounts for the braking distance. Most importantly, we note that the dependence on the square of the vehicular velocity remains unaffected by using the deviation factor (though the numerical values of the safety distance for a given velocity are definitely affected).

Thus, platoon formations are characterized by three key parameters: the maximum number  $n$  of vehicles that are members of a given platoon, the inter-platoon distance factor  $\psi$ , which accounts for the inter-platoon gap required to handle add/drop mergers and transit movements, and the velocity parameter  $v$ .

#### 4. Traffic Management System Optimization

Our first aim is to determine the system parameter values that maximize the vehicular throughput rate  $\lambda$  attainable under prescribed vehicular delay constraints. The delay experienced by a vehicle as it arrives to the road across an input ramp and waits to access the road consists of this waiting time component and of the subsequent transit time delay elapsed as the vehicle travels along the highway to its destination. We assume the corresponding system state process to have reached equilibrium. We model the queuing process of vehicles waiting at the input ramp as an  $M/M/1$  queue. Then, the mean access delay time experienced by a vehicle at the input ramp is given as  $1/(\mu - \lambda)$ . The transit time taken by the vehicle in traveling along the highway link is equal to  $l/v$ . In designing the system, and in particular in synthesizing the optimal vehicular speeds and platoon formations, it is often of interest to impose delay constraints on each of these two components, or on their sum.

##### 4.1. Single lane single link case

In the following we determine the optimal synthesis of the system, aiming to maximize the road’s vehicular throughput rate  $\lambda$  under the following composite vehicular delay constraint, accounting for both on ramp queuing delay and highway transit time:

$$\frac{1}{\mu(v) - \lambda} + \frac{l}{v} \leq D_0 = D_{0q} + \frac{l}{v_{\text{tg}}} \quad (3)$$

We have shaped the delay constraint  $D_0$  by setting it equal to  $D_{0q} + l/v_{\text{tg}}$ , where  $v_{\text{tg}}$  is a target velocity level and  $D_{0q}$  is the on-ramp waiting time bound. Note that only the sum of the two components acts as a constraint, not the individual components. As a consequence of this specified delay bound, we obtain the following upper bound on the system’s vehicular throughput rate:

$$\lambda \leq \mu(v) - \frac{1}{D_{0q} + l/v_{\text{tg}} - l/v} = \frac{1}{1 + \psi/n} \frac{v}{bv^2 + s} - \frac{1}{D_{0q} + l/v_{\text{tg}} - l/v} \quad (4)$$

We then proceed to determine the vehicular velocity level  $v$  that maximizes the latter bound on  $\lambda$ . For future reference we define the function

$$f(v, l; \theta) = \mu(v) - \frac{1}{D_{0q} + l(1/v_{\text{tg}} - 1/v)} \quad (5)$$

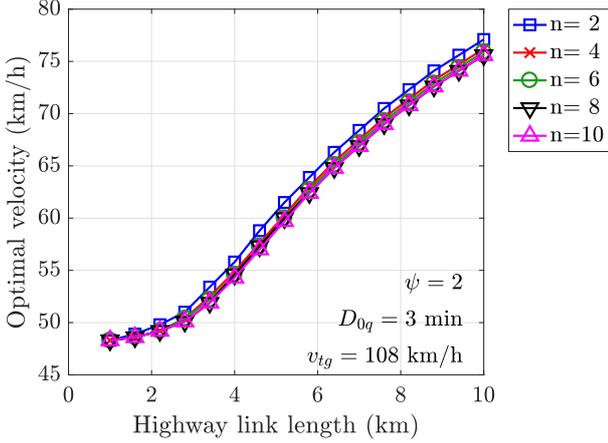


Figure 2: Optimal velocity levels with respect to the highway link length under the mean delay constraint  $D_{0q} + l/v_{tg}$  and different values of  $n$ .

where  $\theta$  is a vector of parameters of the considered link and lane (including  $v_{tg}$ ,  $D_{0q}$ ,  $\psi$ ). It is noted that  $f(v, l; \theta)$  is defined for  $v \in [v_{min}, v_{max}]$ , where  $v_{min}$  is the minimum level of velocity  $v$  such that  $f(v, l; \theta)$  is non-negative. In this range  $f$  has a global maximum as a function of  $v$ .

We note that the constraint in the form  $D_{0q} + l/v_{tg}$  captures the fact that the end-to-end travel time of a vehicle, from when it joins an access ramp queue, until when it arrives at its destination exit and leaves the highway, depends on two distinct effects: (i) waiting time at the access ramp, until when the vehicle is permitted to enter the highway and join a platoon; (ii) travel time along the highway from the access point to the exit point. It is sensible to allow the overall delay requirement to depend on the distance traveled ( $l$  in this case). The delay requirement is surely met if a vehicle spends a time  $D_{0q}$  on the average in the access ramp line and then performs its highway trip at a constant velocity  $v_{tg}$ . However, setting the overall delay requirement as the *sum* of the two delay components, broadens the feasible solution search space of the system optimization, thus giving more flexibility to the design of the highway traffic management system, while at the same time still providing the desired end-to-end average delay.

To illustrate, consider a system characterized by the following parameters:  $u = 3 \text{ m/s}^2$ ,  $\Delta u/u = 0.1$ ,  $s = 6 \text{ m}$ ,  $\psi = 2$ ,  $D_{0q} = 3 \text{ min}$ ,  $v_{tg} = 30 \text{ m/s}$ . We let  $l$  vary on the  $x$ -axis of the graphs and  $n$  act as a parameter, ranging from 2 to 10. We solve the optimization problem and analytically derive the performance results. We plot the following three functions:

1.  $\tilde{v}$ , the velocity level at which the vehicular throughput is maximized under the specified mean delay constraint (Figure 2).
2.  $\tilde{\lambda}$ , the maximum achievable vehicular throughput, under the specified mean delay constraint (Figure 3).
3.  $\tilde{d}_{PL}$ , the value of the inter-platoon distance, whereby

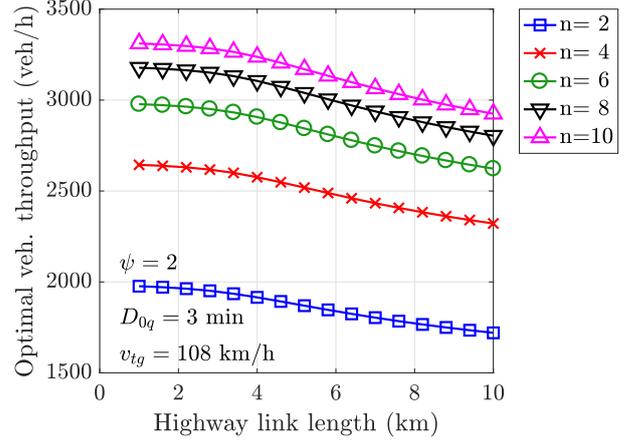


Figure 3: Optimal vehicular throughput with respect to the highway link length under the mean delay constraint  $D_{0q} + l/v_{tg}$  and different values of  $n$ .

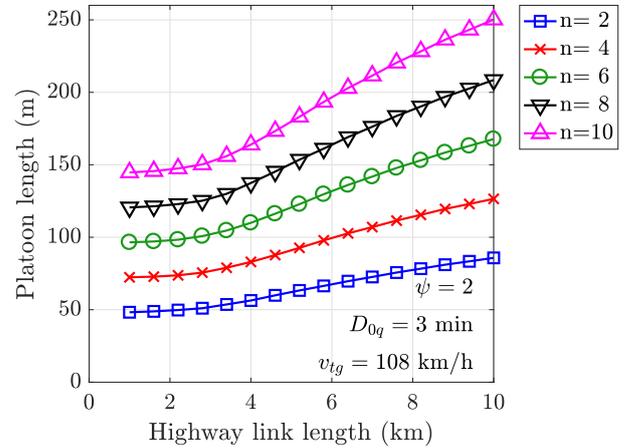


Figure 4: Platoon length with respect to the highway link length under the mean delay constraint  $D_{0q} + l/v_{tg}$  and different values of  $n$ .

each platoon consists of the maximum allowed number of vehicles, at the optimal velocity level (Figure 4).

As for  $\tilde{v}$ , it appears that the optimal level of velocity varies widely as the length of the highway link grows. On the contrary, the size of the platoon, i.e., how many vehicles it is composed of, has a very minor impact on the setting of the optimal velocity.

As expected, the platoon size impacts more significantly the average realized vehicular throughput under the optimal setting of the average velocity. However, as  $n$  increases, a saturation effect is noted. The curves in Figure 3 are observed to get closer and closer as  $n$  increases, thus yielding vanishing returns in terms of the attained average vehicular throughput rate.

We also observe that the length of a platoon under optimal parameter setting ranges between 50–250 m when  $l$  and  $n$  are varied over the entire considered range. Those values

of platoon lengths are well suited to accommodate high performance networking protocols, such as those described and evaluated in later sections.

#### 4.2. Multiple lane single link case

Consider a single highway link comprising  $\ell$  lanes. Lane 1 is the fastest, while lane  $\ell$  the slowest. Let  $v_{\text{tg},i}$  be the target velocity on lane  $i$  and  $\mu_i$  be the serving rate provided by lane  $i$ . Platoon formation structures can change from lane to lane. We expect slower lanes to have to accommodate a higher rate of mobile maneuvers, sustaining a higher rate of lane changing and merging. For example, slower lanes tend to experience more transits by vehicles that aim to merge into faster lanes or temporarily switch to slower lanes in the process of leaving the highway. To accommodate such higher maneuvering rates, the inter-platoon distance gap parameter  $g$  for slower lanes should be set to a higher value than the gap configured for faster lanes. This setting impacts the parameter  $\psi_i$  for lane  $i$ .

Let  $r_i$  be the fraction of the input vehicular flow entering link  $k$  that is directed to lane  $i$ . The mean delay experienced by a vehicle routed through the  $i$ -th lane is given by

$$\frac{1}{\mu_i(v_i) - r_i\lambda} + \frac{l}{v_i} \leq D_{0q} + \frac{l}{v_{\text{tg},i}}, \quad i = 1, \dots, r, \quad (6)$$

where the left-hand part of the inequality represents the delay per lane. By following the same approach as that presented for the single lane case, we express an upper bound on the vehicular throughput flowing along the lane as:  $r_i\lambda \leq \mu_i(v_i) - \frac{1}{D_{0q} + l(1/v_{\text{tg},i} - 1/v_i)} = f(v_i, l; \theta_i)$ . The velocity level per lane,  $v_i$ , is set to maximize the latter bound on  $r_i\lambda$ . Consequently, incorporating the derived maximum velocity levels,  $\tilde{v}_i$ , the overall vehicular throughput rate attained along the multi-lane road is upper bounded as follows:  $\lambda \leq \sum_{i=1}^{\ell} f(\tilde{v}_i, l; \theta_i)$ . Assume the design objective is to maximize the overall vehicular throughput rate  $\lambda$ , under the above prescribed vehicular delay constraints. We readily conclude that this optimization objective is attained when the latter upper bounds on the throughput rate per lane are set to assume the same value. This is achieved by setting

$$r_i = \frac{f(\tilde{v}_i, l; \theta_i)}{\sum_{j=1}^{\ell} f(\tilde{v}_j, l; \theta_j)}, \quad i = 1, \dots, \ell. \quad (7)$$

In conclusion, assuming a road link that employs  $\ell$  lanes and permits the formation of separate on-ramp queues, the vehicular throughput of the single link is maximized by performing the following:

1. Find the per-lane vehicular velocity  $\tilde{v}_i$  that maximizes the function

$$f(v, l; \theta_i) \equiv \frac{v}{(1 + \psi_i/n_i)(bv^2 + s) - 1} \quad (8)$$

$$D_{0q} + l(1/v_{\text{tg},i} - 1/v)$$

where the  $i$ -th lane parameters are  $\theta_i = (\psi_i, v_{\text{tg},i})$ .

2. Set the probability of routing a vehicle to lane  $i$  to  $r_i$ , whereby the latter is calculated by using eq. (7). Then, the maximum vehicular throughput rate that can be accommodated under the delay constraint is  $\lambda = \sum_{i=1}^{\ell} f(\tilde{v}_i, l; \theta_i)$

#### 4.3. Single lane multiple link case

We now move on to consider a highway system composed of a number of links. To that end, we introduce some more notation.

- $\mathcal{F}$  set of vehicular end-to-end flows accessing the highway system.
- $\mathcal{L}$  set of highway links making up the highway system.
- $\lambda_s$  mean vehicle arrival rate of external vehicle flow  $s \in \mathcal{F}$ ; the overall offered vehicular rate is denoted with  $\lambda = \sum_{s \in \mathcal{F}} \lambda_s$ .
- $a_{sk}$  it is 1 if vehicles forming the external flow  $s$  include link  $k$  in their route through the highway system.
- $\nu_k$  mean rate of vehicles entering the highway system through ramp of link  $k$ : note that  $\sum_{k \in \mathcal{L}} \nu_k = \lambda$ .
- $\alpha_k$  fraction of the overall offered vehicular flow rate that enters the highway system through the access ramp located on the highway link  $k$ ; it  $\alpha_k = \nu_k/\lambda$ .
- $\phi_k$  mean vehicle flow in the  $k$ -th highway link, also referred to as internal flows: it is  $\phi_k = \sum_{s \in \mathcal{F}} \lambda_s a_{sk}$ .
- $p_{hk}$  probability that a vehicle, having completed its transit thorough the highway link  $h$ , will move forward to the highway link  $k$ ,  $h, k \in \mathcal{L}$ . Note that  $\sum_{k \in \mathcal{L}} p_{hk} \leq 1$ , being equal to 1 only if no vehicles leaves the highway system after having traveled on link  $h$ .

Based on the definitions above, at equilibrium we can write the following balance equations:

$$\phi_k = \nu_k + \sum_{h \in \mathcal{L}} \phi_h p_{hk}, \quad k \in \mathcal{L}, \quad (9)$$

This is but a flow balance equation made at the point where the access ramp of the  $k$ -th highway link is located (at the beginning of the  $k$ -th highway link). To state the balance equation as in Equation (9), using flow independent routing probabilities, we model the routing of vehicles as a memoryless process. Each vehicle, having traveled up to and including link  $h$ , can move forward to link  $k$  with probability  $p_{hk}$ , independent of previous routing choices and of all other vehicles. This kind of routing is akin to the one hypothesized in Jackson queuing networks [21]. It is a reasonable approximation of reality when each single end-to-end flow contributes a little part of the overall flow traveling on a generic highway link. Obviously, the probabilities  $p_{hk}$  depend on the topology of the highway system, i.e.,  $p_{hk} = 0$  if link  $h$  termination is not connected to link  $k$  starting point.

Consider the following highway synthesis problem. The aim is to maximize a weighted function of the vehicular flows through the highway system under a mean vehicular delay constraint. At equilibrium, the mean vehicular delay through the highway system is proportional to the mean number of vehicles in the highway system, according to Little's law [22].

The delay constraint is imposed separately for each highway link. The mean number of vehicles associated with link  $k$  (including those waiting on ramp and those traveling across the link) is written as

$$\frac{\nu_k}{\mu(v_k) - \phi_k} + \phi_k \frac{l_k}{v_k} \quad (10)$$

where  $\nu_k$  is the mean vehicular arrival rate at ramp  $k$  that is feeding vehicles to highway link  $k$  and  $\phi_k$  is the overall vehicular flow rate on link  $k$  (the latter being the sum of the rate of the flow arriving from ingress ramp  $k$  plus the rate of the flow arriving from upstream links connected to link  $k$ ).

We set the delay constraint as the sum of a term proportional to the access ramp delay  $D_{0q}$  and a term proportional to the travel time on link  $k$ . By converting mean delay into the mean number of vehicles in the  $k$ -th link, according to Little's law, we have

$$\frac{\nu_k}{\mu(v_k) - \phi_k} + \phi_k \frac{l_k}{v_k} \leq \nu_k D_{0q} + \phi_k \frac{l_k}{v_{tg}}, \quad k \in \mathcal{L}, \quad (11)$$

Dividing both sides by  $\nu_k$ , we re-state the constraint as:

$$\frac{1}{\mu(v_k) - \phi_k} + \frac{x_k l_k}{v_k} \leq D_{0q} + \frac{x_k l_k}{v_{tg}}, \quad k \in \mathcal{L}, \quad (12)$$

where  $x_k \equiv \phi_k / \nu_k$ . This inequality is observed to have the same format as the one presented for the single link case, except that: (i) the link flow rate  $\phi_k$  replaces the arrival rate  $\lambda$ ; (ii) the link length is modified to  $x_k l_k$ . The  $x_k$  parameters depend only on the routing matrix and on the load distribution across the links. By re-arranging the linear equation system in (9), we write:

$$x_k = \frac{\phi_k}{\nu_k} = 1 + \sum_{h \in \mathcal{L}} \frac{\phi_h}{\nu_h} \frac{\nu_h}{\nu_k} p_{hk} = 1 + \sum_{h \in \mathcal{L}} x_h \frac{\alpha_h}{\alpha_k} p_{hk}, \quad (13)$$

$$k \in \mathcal{L},$$

hence

$$\alpha_k x_k = \alpha_k + \sum_{h \in \mathcal{L}} \alpha_h x_h p_{hk}, \quad k \in \mathcal{L} \quad (14)$$

Let  $\alpha$  be a row vector of the coefficients  $\alpha_k$  and  $\mathbf{D}_\alpha$  a diagonal matrix whose diagonal is composed of the elements of  $\alpha$ . The vector  $\mathbf{x} = [x_k, k \in \mathcal{L}]$  can be expressed as  $\mathbf{x} = \alpha(\mathbf{I} - \mathbf{P})^{-1} \mathbf{D}_\alpha^{-1}$ , where  $\mathbf{P}$  is the square matrix collecting the probabilities  $p_{hk}$  and  $\mathbf{I}$  is the identity matrix with the same size as  $\mathbf{P}$ . The inverse of  $\mathbf{I} - \mathbf{P}$  exists since  $\mathbf{P}$  is sub

stochastic. Moreover, the inverse is also a non-negative matrix entry-wise.

Once the fractions  $\alpha_k, k \in \mathcal{L}$ , and the routing probabilities  $p_{hk}$  are given, the coefficients  $x_k$  are computed by solving the linear equation system in Equation (14). These coefficients do not depend on the intensity of the vehicular traffic, rather they depend only on the distribution of the vehicular traffic across the inputs of the highway system and on the routing of vehicles through the highway system.

Assume our aim is to design the system in a manner that optimizes an objective function that is expressed as the weighted sum of the end-to-end vehicular flow rates carried along the highway, whereby each end-to-end flow rate is weighted by a route-dependent payoff  $L(s)$  (from the point of view of the highway system manager). In turn, the payoff of a given route is assumed to be the sums of the payoffs induced by a vehicle traveling on each link forming the route, i.e.,  $L(s) = \sum_{s \in \mathcal{F}} a_{sk} q_k$ , where  $q_k$  is the payoff of link  $k$ . An example of payoff is obtained by considering toll fees that are proportional to the distance traveled: then  $q_k = c \cdot l_k$ , where  $l_k$  is the length of the highway link  $k$ .

This objective function is expressed as follows:

$$F \equiv \sum_{s \in \mathcal{F}} \lambda_s L(s) = \sum_{s \in \mathcal{F}} \lambda_s \sum_{k \in \mathcal{L}} a_{sk} q_k = \sum_{k \in \mathcal{L}} q_k \sum_{s \in \mathcal{F}} \lambda_s a_{sk} = \sum_{k \in \mathcal{L}} q_k \phi_k \quad (15)$$

Summing up, the optimization problem is

$$\max_{\phi_k} F(\phi) = \sum_{k \in \mathcal{L}} q_k \phi_k \quad (16)$$

$$\text{s.t.} \quad \frac{1}{\mu(v_k) - \phi_k} \leq D_{0q} + x_k l_k \left( \frac{1}{v_{tg}} - \frac{1}{v_k} \right) \quad k \in \mathcal{L} \quad (17)$$

The following upper bound is induced by the delay constraint:

$$\phi_k \leq \mu(v_k) - \frac{1}{D_{0q} + x_k l_k (1/v_{tg} - 1/v_k)} = f(v_k, x_k l_k; \theta) \quad (18)$$

where  $f(\cdot, \cdot; \cdot)$  is defined in Equation (5). The objective function  $F$  is maximized by setting the velocity levels in each link to a value that maximizes the right-hand side of Equation (18). This is achieved by finding the value of  $v_k$  that maximizes the function  $l_k f(v_k, x_k l_k; \theta)$ .

Numerical examples are presented in the following for the general multi-lane multi-link case.

#### 4.4. Multiple lane multiple link case

In this section, we generalize the analysis to the case of a highway system which is composed of multiple interconnected multi-lane links. The model assumptions are similar to those described above:

- per-sector, per-lane delay constraint is prescribed;

- memoryless assignment of vehicles to the different lanes, i.e., a vehicle admitted to the highway link  $k$  is assigned to lane  $i$  with probability  $r_{ik}$  independently of all other assignments, either past or future.
- memoryless routing process through the highway system, described by the flow termination probabilities  $p_{hk}$ ;
- the system state process has reached equilibrium.

In the following, we use a double subscript notation to identify certain variables as associated with a specific lane (the first subscript) and across a specific link (the second subscript). To satisfy the delay constraint for vehicles that travel in the  $i$ -th lane of the  $k$ -th link we require that:

$$\frac{\nu_{ik}}{\mu_i(v_{ik}) - \phi_{ik}} + \phi_{ik} \frac{l_k}{v_{ik}} \leq \nu_{ik} D_{0q} + \phi_{ik} \frac{l_k}{v_{tg,i}} \quad (19)$$

Let the amount of vehicular flow entering the access ramp on link  $k$  and sent to the  $i$ -th lane be a fraction  $r_{ik}$  of the overall input vehicular flow rate  $\nu_k$ , i.e.,  $\nu_{ik} = r_{ik}\nu_k$ . We assume that the same fraction  $r_{ik}$  of the overall link carried flow rate  $\phi_k$  pertains to the  $i$ -th lane, i.e.,  $\phi_{ik} = r_{ik}\phi_k$ . Substituting into Equation (19) we find

$$\frac{r_{ik}\nu_k}{\mu_i(v_{ik}) - r_{ik}\phi_k} + r_{ik}\phi_k \frac{l_k}{v_{ik}} \leq r_{ik}\nu_k D_{0q} + r_{ik}\phi_k \frac{l_k}{v_{tg,i}} \quad (20)$$

Note that, under our assumptions, we need to find the rations  $x_k = \phi_k/\nu_k$ . We have pointed out above that those rations could be found by knowing the routing probabilities and the external vehicular traffic load profile at the access ramps, the  $\alpha_k$ . From practical deployment point of view, determining the  $x_k$  can be obtained by monitoring the current link flow  $\phi_k$  and the overall rate of vehicles entering through the ramp  $k$ ,  $\nu_k$ . Both monitoring operations can be carried out by using simple technologies, *given that only aggregated overall flows are required*.

Canceling  $r_{ik}$  on both sides and re-arranging terms, we deduce the following upper bound on the vehicular flow rate:

$$\begin{aligned} \phi_{ik} = r_{ik}\phi_k &\leq \mu_i(v_{ik}) - \frac{1}{D_{0q} + x_k l_k (1/v_{tg,i} - 1/v_{ik})} \\ &= f(v_{ik}, x_k l_k; \theta_i) \end{aligned} \quad (21)$$

where  $\theta_i = (v_{tg,i}, \psi_i)$ , and  $x_k = \phi_k/\nu_k$  is the same coefficient as found in the case of multiple links, single lane. Summing over  $i$ , we find  $\phi_k \leq \sum_{i=1}^{\ell} f(v_{ik}, x_k l_k; \theta_i)$ . This upper limit can be achieved by setting

$$r_{ik} = \frac{f(v_{ik}, x_k l_k; \theta_i)}{\sum_{j=1}^{\ell} f(v_{jk}, x_k l_k; \theta_j)} \quad (22)$$

We assume the same target function as that employed in the single lane case, that is to say, a weighted sum of

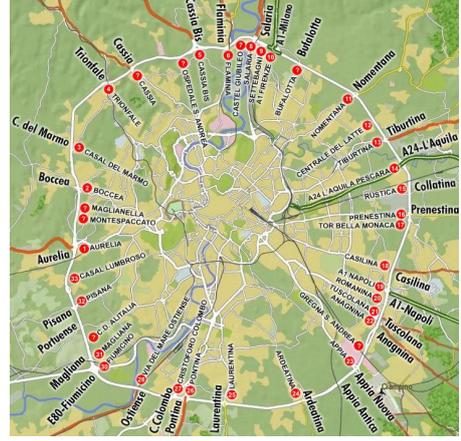


Figure 5: Picture of the GRA layout. Access/exit points are highlighted.

the end-to-end vehicular flows. The weights are interpreted as payoffs for the highway system manager and they are deemed to be the sum of the payoffs of the link traveled by a vehicle. In the end, the target function of the optimization problem is

$$F(\phi) = \sum_{i=1}^{\ell} \sum_{k \in \mathcal{L}} q_{ik} \phi_{ik} \quad (23)$$

where  $q_{ik}$  is the payoff for a vehicle traveling on lane  $i$  of link  $k$ .

The maximum value achieved by this objective function is obtained by performing the following steps.

1. Compute the fraction  $x_k$  given the routing probabilities  $p_{hk}$  and the loading profile  $\alpha_k$ .
2. Set the lane and link velocity levels to values that maximize the function  $f(v_{ik}, x_k l_k; \theta_i)$ . Let  $\tilde{v}_{ik}$  denote the resulting velocity levels.
3. Set  $r_i = f(\tilde{v}_{ik}, x_k l_k; \theta) / \sum_{j=1}^{\ell} f(\tilde{v}_{jk}, x_k l_k; \theta_j)$ .

#### 4.5. Performance behavior

To illustrate the performance behavior of the system under the traffic management models presented above, we apply in the following the optimization procedure to a multi-link multi-lane highway. We consider the ring highway surrounding the city of Rome, Italy – the so called Great Ring Connection Highway (Grande Raccordo Anulare, GRA) (see Figure 5). It is a 69 km closed-ring, three lane highway with 29 main access/exit ramps. The highway links delimited by two successive ramps are numbered from 1 to 29. We associate the access/exit ramp at the *beginning* of the highway link  $k$  with link  $k$ .

In the following we focus on the inner ring direction, i.e., the ring highway where vehicles travel clockwise. By exploiting sample field measurements, we can extrapolate a  $29 \times 29$  Origin/Destination (OD) vehicular flow matrix [23]. The generic element of the OD matrix  $\lambda_{ij}$  gives the

| $k$ | $l_k(\text{km})$ | Exit name                | $k$ | $l_k(\text{km})$ | Exit name            |
|-----|------------------|--------------------------|-----|------------------|----------------------|
| 1   | 3.4              | Via Aurelia              | 16  | 3                | Tor Bella Monaca     |
| 2   | 9.2              | Via Boccea               | 17  | 1.2              | Casilina             |
| 3   | 3.4              | Via Cassia (+ Trionfale) | 18  | 2.1              | A1 Napoli            |
| 4   | 1.6              | Via Cassia Bis           | 19  | 2.9              | Tuscolana/Anagnina   |
| 5   | 1.7              | Via Flaminia             | 20  | 3.6              | Appia                |
| 6   | 0.85             | Castel Giubileo          | 21  | 4.5              | Ardeatina            |
| 7   | 0.6              | Salaria                  | 22  | 2.3              | Laurentina           |
| 8   | 1.1              | Settebagni               | 23  | 1.3              | Pontina              |
| 9   | 4.9              | A1 Firenze               | 24  | 1.5              | Cristoforo Colombo   |
| 10  | 1.8              | Nomentana                | 25  | 2.7              | Via del mare         |
| 11  | 1.5              | Centrale del Latte       | 26  | 1                | Autostrada Fiumicino |
| 12  | 1.6              | Via Tiburtina            | 27  | 2.9              | Magliana             |
| 13  | 1.6              | A 24 Teramo              | 28  | 2.2              | Pisana               |
| 14  | 1.4              | La Rustica               | 29  | 1.8              | Casal Lumbroso       |
| 15  | 0.6              | Pretestina               |     |                  |                      |

Table 2: Name of GRA exits and link lengths.

mean vehicular flow rate entering the access ramp of link  $i$  and leaving through the exit ramp of link  $j$ . Given the unidirectional ring topology, routing has no choice but going through all links between the  $i$ -th one (included), up to the  $j$ -th one (excluded, since the exit ramp is located just at the beginning of link  $j$ ). The OD matrix thus gives enough information to estimate routing probabilities  $p_{hk}$  and load factors  $\alpha_k$ . We have  $\nu_k = \sum_{j=1}^{29} \lambda_{kj}$  for  $k = 1, \dots, 29$ . The link flow rates  $\phi_k$  can be calculated by using the routing of the vehicular flows. This is simple, given the unidirectional ring topology. The flow rate on link  $k$  is the sum of all OD flows  $\lambda_{ij}$  such that  $i \leq k < j$ , if  $i < j$ , or such that  $i \leq k \leq 29$  and  $1 \leq k < j$ , for  $i > j$ . Since the topology of the considered highway system is a ring, we have  $p_{k,k+1} = (\phi_{k+1} - \nu_{k+1})/\phi_k$ , for  $k = 1, \dots, 28$  and  $p_{29,1} = (\phi_1 - \nu_1)/\phi_{29}$ . All other entries of the routing probability matrix are equal to 0.

For the GRA beltway we have  $\ell = 3$  lanes. We label lanes with integers from 1 to 3 according to the maximum allowed traveling speed (1 being the slowest lane, 3 the fastest). We assume that  $v_{\text{tg},1} = 20$  m/s,  $v_{\text{tg},2} = 30$  m/s and  $v_{\text{tg},3} = v_{\text{max}} = 40$  m/s. As for the platoon structure, we assume it is the same throughout the GRA in each lane, but it has different parameters for each distinct lane. Specifically, we let  $\psi_1 = 3$ ,  $\psi_2 = 2$ , and  $\psi_3 = 1$ . This setting is consistent with the fact that lane 1 is the closest to access/exit ramps, hence it is involved in most vehicle maneuvers, whereas lane 3 is only affected by platoon add/drop of vehicles that need to travel on that lane. As for the platoon size  $n$ , we let it vary between 1 and 10. The lengths of the 29 highway links are listed in Table 2.

To illustrate, we use measurements performed over GRA to configure the above-mentioned parameter settings in order to demonstrate the system's performance behavior, as obtained by applying the above derived technique for throughput optimization under the mean delay constraints (imposed per lane and per link). Here we assume  $D_{0q} = 3$  min. Figure 6 plots the mean vehicular link flows as a function of the distance from exit  $k = 1$  on GRA. The link flow rates are obtained according to the optimization

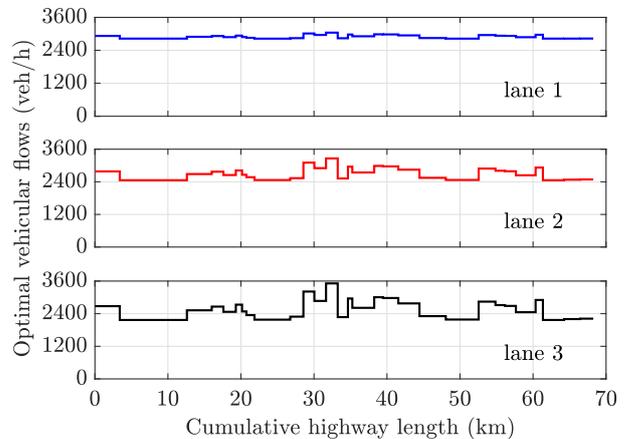


Figure 6: Mean link vehicular flow rates  $\phi_{ik}$  for lane  $i = 1, 2, 3$  as a function of the distance from exit  $k = 1$  on GRA ( $n = 10$ ,  $D_{0q} = 3$  min).

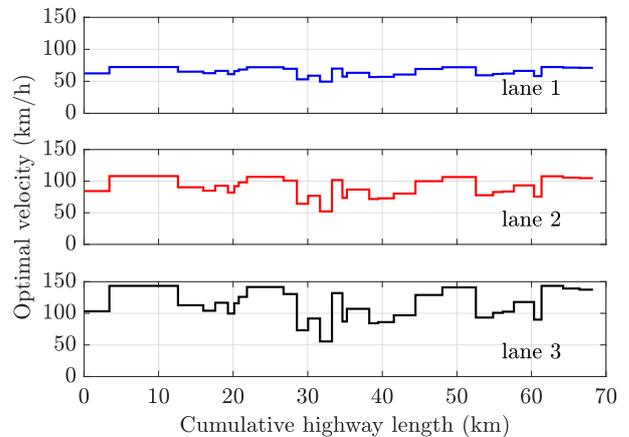


Figure 7: Mean vehicle velocity  $v_{ik}$  for lane  $i = 1, 2, 3$  as a function of the distance on GRA starting from exit  $k = 1$  ( $n = 10$ ,  $D_{0q} = 3$  min).

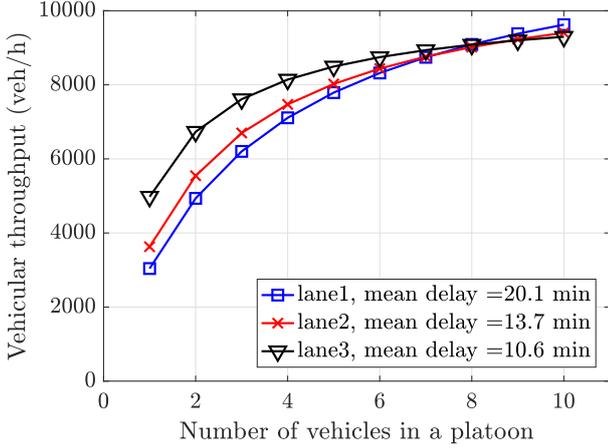


Figure 8: Overall vehicular throughput per lane as a function of the platoon size  $n$ . The corresponding realized end-to-end mean delay is shown in the legend.

procedure. A mean flow rate value is obtained for each link. The flow rate values are depicted as function of the coordinate along GRA to account for different lengths of the GRA links. The flow rates are higher along the slowest lane and lower across the fastest lane. This is consistent with the results highlighted in Figure 7, which shows the mean vehicle velocity levels assigned to platoons per lane as a function of the coordinate along the GRA, starting from exit  $k = 1$ . It is apparent that lane 3 is actually the fastest and lane 1 is the slowest, though the assigned velocity levels vary significantly from one link to the other. This is induced by the outcome of the optimization process that aims to accommodate the maximum allowable vehicular flow, under the mean delay constraint, given the routing matrix and the load profile used by the numerical evaluation process, which is consistent with real traffic measurements collected on the GRA.

The overall mean throughput realized on each lane as a function of the platoon size is shown in Figure 8. As expected, it grows with  $n$ . We observed that the mean delay corresponding to the optimized offered flow and link vehicular rates has negligible variations with  $n$ , independently of the considered lane. For lane 1, the mean delay varies from 20.16 min to 20.03 min when the platoon size  $n$  ranges between 1 and 10. Similarly, for lane 2 and lane 3, we have negligible variations of the mean delay: namely, it decreases monotonically from 13.81 min down to 13.73 min (lane 2) and from 10.65 min down to 10.62 min (lane 3) for  $n$  ranging from 1 to 10. For this reason, we report just the value of the mean delay averaged over all outcomes obtained as  $n$  varies from 1 to 10.

The target function of the optimization for the GRA is the weighted sum of end-to-end vehicular flows, with weights equal to the distance travelled by each vehicle. In other words, we have chosen payoffs proportional to the distance traveled by a vehicle on the highway. We have already noted that the target function is then proportional

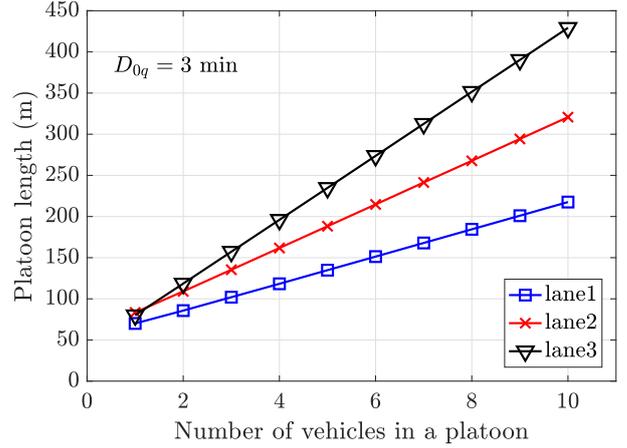


Figure 9: Platoon length averaged over all highway links as a function of the platoon size  $n$ .

to  $\sum_{i=1}^{\ell} \sum_{k \in \mathcal{L}} l_k \phi_{ik} = \sum_{k \in \mathcal{L}} l_k \phi_k$ . To avoid having unit-measure dependent values, we normalize the target function by the overall length of the GRA beltway. Consequently, the target function for the GRA optimization process is given as

$$F(\phi) = \sum_{k \in \mathcal{L}} \frac{l_k}{\sum_{h \in \mathcal{L}} l_h} \phi_k \quad (24)$$

It is found that  $F(\phi)$  grows monotonically from 0.91 to 2.23 as  $n$  varies between 1 and 10.

In Figure 9, we show the variation of the platoon length, averaged over all highway links in each lane, as a function of  $n$ . It is important to note that by using platoon-oriented formations, and thus empowering the platoon leader to manage its platoon members and coordinate operations with leaders of neighboring platoons, the ensuing mechanism achieves enhanced energy consumption rates. Vehicles that are not elected to serve as platoon leaders realize reduced processing rates and thus experience an improved energy aware operation. Also, by using the underlying platoon architecture to effectively regulate the access of vehicles to the highway, routing them to designated lanes and organizing them into preferred vehicular formations in each lane, end-to-end travel delays are much enhanced, further contributing to an improved energy efficient operation. We note that platoon leaders use periodically broadcasted system status message flows to make use of current conditions to best configure system parameters. Accordingly, when advantageous, vehicles that are identified with higher energy resources can be given preferential weights for selection as platoon leaders.

## 5. Networking Architecture and Protocols

In the following sections, we illustrate the configuration and performance behavior of a V2V wireless network system that we synthesize by taking advantage of the organization of vehicles flowing along the highway lanes into

platoon formations. As often assumed, the wireless medium assigned for communications among vehicles is divided into distinct communication channels. Included in our models are thus a Control Channel (CCH) and a Service Channel (SCH). As performed by IEEE 802.11p systems, vehicles use the CCH to periodically (such as every 100 ms) transmit and disseminate status updates to other vehicles in their neighborhood. The CCH under consideration can be employed in a manner similar to that used by the WAVE system [24], as well as by other architectures. In these systems, CCH is used by vehicular and by infrastructure entities to broadcast in each area status, management and control messages [25, 26, 27]. It is also used at times to announce resource allocations, such as the assignment of time slots for use by contention-free MAC schemes [28, 29].

To demonstrate the performance behavior of a platoon-based V2V communications network that we hereby introduce, we consider two classes of event-driven messages produced by highway vehicles that are targeted for dissemination to other highway vehicles over a SCH. To illustrate, we assume that each such message has a dissemination span of a range of at least  $d_{\text{span}} = 300$  m. We note that for protocol simplification, the dissemination span may at times somewhat exceed this range. Also, to simplify we assume in the following that messages produced by a source vehicle are disseminated to vehicles that travel in the specified span behind the source vehicle (i.e., behind its direction of motion). Our protocols and basic models are noted to readily extend to apply to scenarios whereby the dissemination span bi-directionally covers vehicles that move behind and ahead of the source vehicle. We assume Class 1 messages to include critical (e.g., safety) messages and to thus be granted higher priority. The throughput rate of such event-induced message flows is relatively low. However, such a message should be disseminated across its span aiming to achieve a Packet Delivery Ratio (PDR) across the complete span that is at least 0.95. Successfully delivered messages should complete their dissemination within strict time latencies. Class 2 message flows are of lower priority. Yet, they tend to produce message flows that impose high throughput rate requirements on the wireless network, while also generally imposing packet time delay requirements. For either class, we assume in this paper a 90 to 95 percentile time delay requirement of 50 ms.

The network architectures and protocols presented in this paper, extend the concepts and techniques introduced by us in developing a class of mobile ad hoc networks that move in two or three dimensions, identified as MBNs and a class of V2V wireless networks identified by us as VBNs. Yet, in contrast, the network systems introduced in this paper involve the autonomous mobility of highway vehicles and are based on the formation of vehicular platoons across highway lanes. Otherwise, the networking systems studied in this paper assume a two layer hierarchical architecture similar to that used in our MBN and VBN-based systems: certain vehicles are elected to act as BNs forming an interconnected wireless Bnet; each BN manages the access

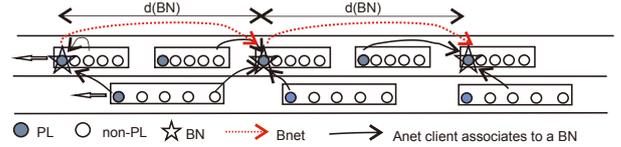


Figure 10: Network Architecture: BNs and associated Anet clients. The solid arrows represent association of Anet clients to BN and the dotted arrows represent Bnet communications

of its client mobiles across its wireless Anet. Figure 10 illustrates the network systems which we use in this paper. We assume the wireless communications channels forming the Bnet and Anet systems to be shared in accordance with one of the following joint MAC protocols:

- 1 Demand-assigned (DA) TDMA scheme across each Anet and TDMA scheme across the Bnet (MAC 1);
- 2 IEEE 802.11p oriented protocol across each Anet and TDMA scheme across the Bnet (MAC 2);
- 3 IEEE 802.11p oriented protocol across each Anet and IEEE 802.11p protocol across the Bnet (MAC 3).

### 5.1. MBN-based networking protocol

As noted above, the network architecture studied in this paper consists of the Bnet and its BNs and the corresponding Anets. An Anet client (i.e., a vehicle) associates itself with a BN which it determines to be the closest to it (or, generally, from which it receives signals at the highest power or signal quality, as often done in a cellular system). The messages produced by a vehicle that is a client of a BN are transmitted by the source vehicle across its Anet to its associated BN (identified as the corresponding source BN). The later disseminates the messages it slates to relay over the Bnet across a multi-hop path that consists of BN-to-BN links, covering each flow's targeted  $d_{\text{span}}$  range. This path consists of  $n_{\text{hops}} = \lceil d_{\text{span}}/d_{\text{BN}} \rceil$  hops. The last BN across a path which receives a flow's packets (but does not further relay them), following their transmission across the last Bnet link over their path, is identified as the destination BN for this flow. Messages transmitted across each Bnet link are assumed to be also received by the mobiles that overhear these transmissions. For this purpose, each mobile is programmed to listen and prepare to receive messages (which are of interest to it) that are disseminated across the Bnet. In particular, message transmissions that are sent across the last Bnet link in a flow's dissemination path, sent from the next-to-last BN to the last BN in the path, are assumed to be received by all vehicles that hear them. The latter vehicles are generally client members of the Anets that are associated with either the next-to-last BN or the last BN.

By using BNs (among elected PLs) to forward messages, rather than performing hop-by-hop forwarding, we reduce in a significant manner the effective number of hops that are

used for the V2V dissemination of data messages and thus increase the energy efficiency of the networking operation.

Under a borderline case, such as when  $d_{\text{BN}} > d_{\text{span}}$ , the destination Anet client is a member of the source Anet. The performance metrics are calculated in a similar manner for this case under the setting of  $d_{\text{BN}} = d_{\text{span}}$ . The overall system throughput is scaled by using a factor that is proportional to the number of system BNs,  $N_{\text{BN}}$ , similarly for the case where the source and destination mobile nodes are members of the same Anet, sharing the same BN. In case the distance from Anet client to its BN is greater than its  $d_{\text{span}}$ , the Anet client can proceed to directly reach its destination through a sole transmission across its Anet. To simplify, we focus on dissemination span coverage ranges that exceed the Anet diameter span.

## 5.2. Synthesis of the hierarchical network architecture

### 5.2.1. Platoon management for network synthesis

We assume that vehicles traveling across each lane of the highway are organized into platoon formations. Each platoon is managed by its PL. Alternatively, another member of a platoon may also be elected to act as the PL. The corresponding management architecture includes the following management entities (whose functionality may be physically realized to reside in a single or separate nodes): the Highway System Manager (HSM) and the Ramp Access Manager (RAM).

HSM collects and stores status and capability information about vehicles under its supervision; for this purpose, it also monitors the periodic status messages disseminated by highway vehicles. The information includes vehicles' location coordinates, speeds, destination, communications channel, radio, memory, processing and energy resource capabilities, environmental status, and management responsibilities, including whether it currently serves as a PL. This allows the system manager to also regulate the admission process of vehicles waiting on-ramp for being admitted into the system. It also computes optimum traffic management parameters including:  $d_V$ ,  $d_P$ ,  $n$ ,  $d_{\text{PL}}$ . These parameters can be adapted to different vehicular traffic loads. The status of vehicle  $i$  (among  $N$  vehicles currently admitted into the highway segment under consideration) is represented by a status vector that includes the following components:  $x_i = (pl_i, c_i, bn_i, v_i)$ ,  $i = 1, \dots, N$ . The  $pl_i$  state variable indicates whether vehicle  $i$  is a PL,  $c_i$  represents the coordinates of the location of vehicle  $i$ ,  $bn_i$  indicates whether a vehicle currently serves as BN,  $v_i$  represents the vehicle's velocity.

In considering the mobility of platoons relative to an access point of a ramp (that is used for vehicles entering the highway as well as for vehicles departing the highway), we identify the following platoon status indicators. A departing platoon is identified as a platoon which is either in the process of passing through a ramp's access point while its tail has not yet passed this access point, or has completely passed this access point (so that its tail vehicle

passed it as well). An arriving platoon is identified as a platoon whose leader has not yet arrived at the ramp's access point (which is located upstream relative to its mobility).

The state of the system at ramp  $k$  at a time  $t$  that corresponds to the time that a platoon within a highway reaches the ramp and thus may incur vehicular departures (identified therefore as a potentially departing platoon) is represented by the vector  $X_t = ((P_d, P_a)^{(k)})$ . The components of the vector  $P_d(n_d, c_d, v, bn_d)$  identify the states of a departing platoon. The component  $n_d$  represents the number of vehicles belonging to a departing platoon,  $c_d$  represents the coordinates of the last vehicle of this platoon,  $bn_d$  indicates whether the platoon leader serves as a BN. The vector  $P_a(n_a, c_a, v, bn_a)$  stores the states of the arriving platoon, whereby  $n_a$  is the number of vehicles within the arriving platoon,  $c_a$  are the coordinates of leading vehicle of the platoon,  $bn_a$  indicates whether the platoon leader serves as a BN.

The HSM computes and disseminates the optimal values of the communications network parameters, such as  $d_{\text{BN}}$ , in considering the system's desired vehicular traffic and communications networking performance features, using the formulas and approaches described in this paper.

The RAM receives state information from the HSM, as well as the outcome of its computations that convey the desired system traffic management parameters, including recommended platoon formation parameters, the optimal networking parameters (including  $d_{\text{BN}}$ ), and the states of vehicles which are subjected to its local management. The RAM uses the local vehicular state updates that it periodically receives from the HSM to regulate the admission of vehicles into the highway, to determine the structure of new platoons that it forms, to control the admission of vehicles in joining existing platoons, merge platoons, and to elect PLs and BNs.

For purpose of forming platoons and managing vehicular admissions, the RAM computes several time spans. It calculates the time elapsed between the tail of a departing platoon and the time of arriving PL at its access point, denoting it as  $T_{d,a} = \frac{c_d - c_a}{v}$ . It also calculates the time parameter  $T_p$  that represents the minimum time required to elapse between the traversal time at its access point of the tail vehicle of a departing platoon and the time of arrival of the subsequent leader of an arriving platoon;  $T_p = \frac{d_{\text{PL}} + d_P}{v}$ , corresponding to the gap required to be set between such instants of time. The later parameter is used in its decision as to whether there exists a sufficient gap that allows it to form a new platoon.

### 5.2.2. Synthesis of platoon formations

The principles that we have used for platoon formation include the following assumptions:

- The RAM considers only arriving platoons in determining the admission of vehicles into the highway. It does not consider (though generally feasible) the

admission of a new vehicle with the purpose that it would accelerate to catch up with a departing platoon.

- Vehicles within a platoon move forward to occupy vacant platoon spaces, if any. Vehicles which are admitted from the ramp join a platoon in occupying its tail positions.
- The computed optimal platoon parameters apply to all platoons and vehicles that travel in the same lane, but can be different for different lanes.
- When there are multiple lanes, an admitted vehicle would be routed to a platoon traveling in a designated lane, based on the routing schemes described in the Traffic Management sections of this paper. Yet, due to space limitations, we describe in the networking sections a single lane operation.

The following principles govern the process used by a vehicle admission and platoon formation algorithm (clearly, other versions can be similarly composed):

1. If no platoon is currently passing at a ramp's access point, and if the time to elapse between the last departing platoon and the next arriving platoon is sufficiently long, such that  $T_{d,a} \geq T_p$  and at least one vehicle is queueing in ramp, the ramp manager would then admit queued on-ramp vehicle(s) to form a new platoon, electing the first one admitted under this event as the platoon's PL. The manager would proceed to subsequently admit other queued vehicles to join this platoon, assuring that the new platoon contains no more than a total of  $n$  vehicles.
2. If no platoon is currently passing by the ramp, and  $T_{d,a} < T_p$ , the ramp manager waits until the next platoon arrives at its access point prior to making new admission decisions. If the platoon that will subsequently arrive is not full (accounting for current departures as well), so that  $n_a < n$  and there is at least one queueing vehicle, the ramp manager admits new vehicles to join this in-transit platoon, assuring its membership to not exceed its designated capacity  $n$ .
3. If an arriving platoon is currently passing by the ramp access point, and if there is at least one queueing vehicle, the platoon manager admits new vehicles to join the platoon in accordance with the principles stated above.

### 5.2.3. Backbone network synthesis

In this section, we illustrate a protocol that is used for the formation of a Bnet. Depending upon the managers objectives, other synthesis protocols can be readily applied. Furthermore, RSUs, including Base Station or Access Point nodes, that are installed at road-side may also be employed

and embedded into our models and analyses. To simplify the discussion, we hereby assume that BNs are elected among PLs.

For multi-lane highway systems, BNs can be elected by considering PLs of platoon that travel across any lane or in preference of PLs that travel along certain lanes. For example, the advantage of electing BNs from PLs that are members of platoons that travel along the slow lane is that a lower rate of re-elections of PLs and BNs may be invoked. Slower lane vehicles will then also incur lower rate of re-associations (of a mobile with an elected BN). Furthermore, the slower lane is usually denser, allowing more flexibility when wishing to elect a higher density Bnet. On the other hand, the advantage of electing BNs from PL vehicles that travel along a fast lane is that such vehicles tend to travel longer distances, and thus staying longer in the highway and serving an assigned BN role. For illustrative and discussion purposes, we assume that the RAM assigns higher priority for electing BNs among PLs that travel along a slower lane. When the manager cannot find at an underlying period of time a PL that travels in a slow lane, it proceeds to choose one that travels along a faster lane.

Each RAM receives information from the HSM regarding the status of vehicles that it manages, including the residual lifetime (time to destination) and entry and departure locations of these vehicles. As noted above, BNs are elected from the group of underlying PLs, including PLs which have already entered the highway or vehicles queued on the ramp which will be admitted into the highway. The rules for BN election for association of mobiles with a BN are as follows:

1. If a RAM observes that there is no vehicle that assumes a BN role traveling across its highway link within its communication range, in upstream or downstream direction, it will elect a PL whose location is closest to its access point to serve as a BN, preferring one that travels along the slowest lane.
2. If the RAM detects at least one vehicle that acts as a BN to transit its location but determines that there is no other vehicle that serve as a BN which is located  $d_{BN}$  away from the latter BN, in either upstream or downstream direction, the RAM proceeds to elect a PL to serve as a BN, aiming this PL to be located a range of  $d_{BN}$  away from the preceding detected BN. For this election, it can consider vehicles that travel in either upstream or downstream direction from the existing BN, in its vicinity.
3. If the above process results in  $d_{BN} \leq d_{PL}$ , the RAM attempts to elect a PL from another lane to serve as a BN, provided the latter is at a distance that is closer to the targeted distance of  $d_{BN}$ , to ensure the desired Bnet connectivity throughout the highway. It is noted that in periods during which the vehicular flow density levels along the lanes are low, it could

help to elect BNs from PLs that belong to multiple lanes in aiming to synthesize a high capacity connected Bnet. If no available PL which satisfies the latter  $d_{\text{BN}}$  distance requirement can be detected, PL that travel along the slow lane are configured to serve as BN.

- 4 A PL which is elected to serve as a BN advertises this role by indicating it within its periodic status message. It is thus detected by all neighborhood mobiles (including PLs, BNs and other mobiles).
- 5 The RAM initiates election of a new BN upon departure of a PL which was elected as a BN.
- 6 A non-BN regularly listens to periodic status message advertisements on the CCH. It uses this information to identify the identity of the BN which it receives at the highest communications quality; often the one that is closer to it. It then associates itself with the Anet managed by the latter BN.

## 6. Network Parameters and Performance Metrics

In the following, we study the performance behavior of the network under several Anet and Bnet MAC protocols, determining for each case the best configuration of the platoons and the best inter-BN distances to be configured. Platoon compositions and velocity parameters that correspond to attaining high vehicular flow rates, as determined by the traffic management models presented in this paper, are incorporated. For configuring the communications network to achieve high performance, we aim to synthesize a network system that yields a high message throughput rate under message delay objectives. To characterize the highest achievable delay-aware data throughput capacity rate, we carry out in the following performance analysis and design evaluations under the assumption that vehicular nodes are highly loaded by messages that can belong to both Class 1 and Class 2 application flows. Hence, we assume that packets of either class are provided packet delay dissemination limiting values. When analyzing multiple lane scenarios, a selective set of PLs are elected as BNs. As noted above, we select BNs from PLs that travel along the slow lane. Through our simulation and analysis based evaluations, we have found out that parametric results presented for a single lane can point to performance attained in a multi lane environment. However, due to space limitations, we focus here on performance results for vehicles that travel along a single lane. The total number of clients served on the highway is set to  $N$ . Each BN serves across its Anet  $n_{\text{clients}}$ . We determine the best configuration of each network scenario by varying the inter-PL and inter-BN distance levels, and by varying the employed Anet and Bnet MAC schemes, and their corresponding transmission data rates and reuse (coloring) levels.

To illustrate the performance tradeoffs available to the designer under various network configuration options, our

simulations and analysis based performance evaluations employ the following parameter ranges. We consider Bnet link ranges,  $d_{\text{BN}}$ , to range from 100 m to 500 m. The number of vehicles  $N$  admitted to the single lane system under consideration is set to range from 100 to 500, representing highway congestion levels that vary from light to heavy loading conditions. The inter-vehicular distances within a platoon,  $d_V$ , is varied from  $d_{\text{PL}}/20$  to  $d_{\text{PL}}/n$ , and the transmit power level  $P_{\text{tx}}$  is varied between 23 dBm and 33 dBm.

The simulations for IEEE 802.11p-related MAC schemes are implemented by using the NS-3 simulator. For TDMA-related MAC schemes, we use analytical models and MATLAB based evaluations. For performance comparisons, we focus our evaluations on setting  $d_{\text{BN}}$  levels at 100, 150 and 300 m, noting that the assumed  $d_{\text{span}}$  for flow dissemination is assumed to be equal to 300 m, and under the later  $d_{\text{BN}}$  levels Bnet disseminations traverse an integral number of inter-BN hops. The results however well illustrate the performance trends to be incurred under other setting levels.

We assume throughout a gap factor of  $\psi = 2$ . We found out that when using a larger  $\psi$  value, for example, setting it to 3, yields generally better or similar network performance behavior, for all the MAC schemes that we have studied. Note however that, as shown above in the traffic management sections, higher  $\psi$  values, which may have to be imposed to accommodate trans-lane movements, lead to lower vehicular flow rates.

The signal power received at a node for a given transmission by a node that is located at a distance equal to  $d$ , denoted as  $P_r$ , is calculated by using the communications propagation loss model used in [30] :

$$P_0 = P_{r,\text{dB}}(d_0) = P_{t,\text{dB}} + 10\log\left(\frac{\lambda^2}{(4\pi)^2 d_0^2}\right)$$

$$P_{r,\text{dB}}(d) = \begin{cases} P_0 - 10\gamma_1 \log_{10} \frac{d}{d_0}, & \text{if } d_0 \leq d \leq d_c \\ P_0 - 10\gamma_2 \log_{10} \frac{d}{d_c} - 10\gamma_1 \log_{10} \frac{d_c}{d_0}, & \text{if } d > d_c \end{cases}$$

We aim to maximize the aggregate throughput rate of data packets which are transported across the V2V system,  $TH_{\text{Anet,Bnet}}$ , as realized under prescribed packet delay and PDR levels. We calculate this performance metric as follows. A data packet is transmitted across the network in traversing a path that consists of its source Anet link followed by the Bnet communications links that belong to its dissemination path. Such a packet is said to be successfully transported across the system if it is successfully received by its flow's destination BN within the prescribed delay limit, noting that it will then also be overheard by vehicles that travel along the corresponding path. A flow of such data packets (that are received in a timely fashion, in accordance with the packet delay prescription) is said to be successfully delivered if the fraction of its successfully received data packets is higher than the specified PDR level. We thus aim to maximize the system's throughput rate of such successful flows.

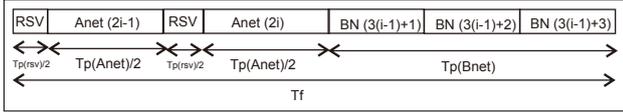


Figure 11: Time Slot Allocation in a DA/TDMA Anet-TDMA Bnet Time Frame for Anet coloring  $k_c = 2$  and TDMA Bnet reuse-M = 3 and  $i \geq 1$

Under the MAC 1 and MAC 2 schemes, we obtain the  $TH_{\text{Anet,Bnet}}$  metric by combining performance results that are obtained by separately analyzing Anet and Bnet operations. Assuming that all allocated system bandwidth is allocated for Anet operations (thus isolating its performance from that induced by Bnet operations), the aggregate throughput rate of data packets supported across the system Anets is denoted as  $TH_{\text{Anet,I}}$ . Similarly, assuming all system bandwidth to be used for Bnet operations, the corresponding Bnet throughput is denoted as  $TH_{\text{Bnet,I}}$ .

The network parameters and performance metrics that we have used are summarized in Table 3.

### 6.1. DA/TDMA Anet - TDMA Bnet (MAC 1)

Under this joint MAC configuration, allocated Anet and Bnet bandwidth resources are shared on a TDMA basis. Also, as shown in Figure 11, we assume the overall bandwidth allocated for data communications to be shared by Anets and the Bnet on a TDMA basis (equivalently, one could use an FDMA configuration to separate Bnet and Anet transmissions). We assume BNs to acquire time synchronization, as required for the implementation of the underlying time division schemes. This can be attained by using the periodic broadcasting of a master clock in beacons that are disseminated across the region by using a control channel, or by such message beacons that are periodically transmitted by road-side or other units. In each time frame, time slots are allocated for data transmissions made by Anet clients and by BN nodes. Also, during Anet periods, time slots are allocated for the transmission of reservation packets (that are embedded as reservation indicators sent to the BN by a client that transmits the first message of a new flow during the reservation slot). Thus, each time frame  $T_f$  consists of a time period allocated to Anet clients to make reservation requests,  $Tp_{\text{rsrv}}$ , time allocated for Anet transmissions,  $Tp_{\text{Anet}}$ , and time allocated for Bnet transmissions,  $Tp_{\text{Bnet}}$ . Hence:

$$T_f = Tp_{\text{rsrv}} + Tp_{\text{Anet}} + Tp_{\text{Bnet}}$$

To reduce interference caused by transmissions that are executed simultaneously across neighboring Anets, a  $k_c$ -coloring MAC scheduling scheme is employed. Hence, during a  $Tp_{\text{Anet}}$  period, clients in a given Anet can transmit for a fraction of time that is equal to  $\frac{1}{k_c} Tp_{\text{Anet}}$ , so that during such a period only 1 out of  $k_c$  consecutive Anets can allow Anet transmissions to take place. We have determined the proper Anet reuse level  $k_c$  to use to mitigate inter-Anet interference, in aiming to assure a high PDR

level,  $PDR \geq 0.95$  while attaining the highest feasible throughput level. The coloring value to be properly used depends on the employed Anet data rate,  $R_{\text{Anet}}$ . Under such higher data rate values, a higher reuse  $k_c$  level is typically required induced by the higher SINR threshold required at the intended BN receiver.

We allocate within each  $\frac{1}{k_c} Tp_{\text{Anet}}$  period, a single Anet time slot of duration  $Ts_{\text{Anet}} = \frac{P}{R_{\text{Anet}}}$  to each one of the  $n_{\text{served}}$  actively served Anet clients, where  $n_{\text{served}}$  represents the configured number of active Anet clients that can be served by a BN during a time frame.

As noted, our throughput analysis has been performed by assuming heavily loaded BNs. The allocation of Anet slots to a newly active application flow is made on a first-come first-serve basis depending on the arrival time of the first packet of a stream flow. Hence, the index of the transmitting Anet client in each Anet is effectively random. Some transmissions may fail due to interference between Anets (as they may then to not meet minimum SINR level requirement).

The Anet MAC scheme is based on the use of a Demand-assigned (DA) TDMA. The Anet data rate used by Anet clients is denoted as  $R_{\text{Anet}}$ . Anet clients need to send reservation requests if the number of active clients associated to a BN is higher than than the number of clients which can be served in a single time frame. When the number of active clients  $n_{\text{active}}$  is lower than the  $n_{\text{served}}$ , all time slots can be allocated to these  $n_{\text{active}}$  mobiles on a dedicated basis. Otherwise, if  $n_{\text{active}} > n_{\text{served}}$ , Anet clients must send reservation packets to the associated BN. Reservation transmissions use the allocated time periods using the bandwidth allocated to the SCH. Class 2 message flows can employ a slotted ALOHA MAC to transmit reservation messages during the reservation slots. In turn, Class 1 messages, may use periods allocated within the CCH to assure a very high success rate in sending reservations. For this purpose, a high priority access mechanism employed by a priority-based IEEE 802.11p protocol can be employed. To well accommodate real-time flows, the BN allocates available data slots for the subsequent transmission of all packets that are members of the flow for which reservation has been successfully carried out. The BN may change the allocation of slots to its client mobiles if slots are required for the support of a new high priority flow.

Across the Bnet links, each BN transmits its packets to its BN neighbor at the  $R_{\text{Bnet}}$  data rate. To mitigate interference among simultaneously transmitting BNs, and reuse-M level is used. We select the optimal M value, considering the most effective levels, which are established to be equal to the set of values  $M = 3,4,5$ . It is noted that for a radio receiver to effectively receive messages at a prescribed bit error rate, for the underlying employed modulation/coding scheme, one needs to assure the receiver to satisfy minimum SINR and minimum receiver sensitivity level requirements.

The throughput rate achieved along the Bnet is calcu-

| Acronym                 | Definition   | Values                            | Reference |
|-------------------------|--|-----------------------------------|-----------|
| $N$                     | total number of vehicles served by the BNs along the highway   | 100–500                           |           |
| $N_{\text{PL}}$         | number of PLs on the highway $\lfloor L/d_{\text{PL}} \rfloor$   |                                   |           |
| $d_{\text{BN}}$         | distance between BNs along and parallel to the highway   | 100–500 m                         |           |
| $N_{\text{BN}}$         | number of BNs on the highway $\lfloor L/d_{\text{BN}} \rfloor$   |                                   |           |
| $R$                     | transmission rate used by the Anet $R_{\text{Anet}}$ or Bnet $R_{\text{Bnet}}$   | 6, 12 and 24 Mbit/s               | [31]      |
| $SINR_t$                | minimum SINR to receive packets correctly for different $R$  | 7, 11 and 20 dB                   | [32, 33]  |
| $rv_t$                  | minimum receiver sensitivity for 10 MHz channel for different $R$  | −85, −82 and −77 dBm              | [31]      |
| $P$                     | packet length  | 3024 bit                          | [34]      |
| $d_{\text{span}}$       | targeted message dissemination span opposite to travel direction   | 300 m                             | [16, 35]  |
| $P_{\text{tx}}$         | transmission power used by all communicating nodes   | 23 and 33 dBm                     | [36]      |
| $P_n$                   | noise power  | −104 dBm                          |           |
| $M$                     | reuse-M factor used for the Bnet using TDMA  | 3, 4 and 5                        |           |
| $k_c$                   | coloring applied on the Anet using TDMA  | 1, 2, 3 and 4                     |           |
| $CS_t$                  | carrier sensing threshold  | −85 dBm                           | [31, 30]  |
| $d_{\text{CS}}$         | carrier sensing radius assuming single farthest transmission sensed with $CS_t$  | 346 m ( $P_{\text{tx}} = 23$ dBm) |           |
| $\gamma_1, \gamma_2$    | path gain coefficients used in the used propagation loss model   | 1.9 and 3.8                       | [30]      |
| $d_0, d_c$              | reference distance and cut-off distance in the used propagation model  | 10 and 80 m                       | [30]      |
| $TH_{\text{Anet,I}}$    | aggregate data throughput achieved by Anet clients that transmit their packets to their associated source BNs when the Anets occupy the complete system bandwidth (isolated from the Bnet)       |                                   |           |
| $TH_{\text{Bnet,I}}$    | aggregate data throughput attained by transmissions across the Bnet from source BNs to destination BNs when the Bnet is allocated the complete bandwidth of the system (isolated from the Anets) |                                   |           |
| $TH_{\text{Anet,Bnet}}$ | aggregate data throughput attained by vehicle-to-vehicle packet transmissions when the Anets and Bnet operate jointly in sharing the allocated communications bandwidth resources                |                                   |           |
| $D_P$                   | end-to-end communication delay from source node to destination node  |                                   |           |
| $D_{P,\text{max}}$      | maximum end-to-end packet delay  | 50 ms                             | [16]      |
| $PDR$                   | packet delivery ratio: fraction of flow packets successfully received at a flow's destination BN   | $\geq 0.95$                       |           |

Table 3: Network parameters and performance metrics.

lated as follows:

$$TH_{\text{Bnet,I}} = \begin{cases} R_{\text{Bnet}} \frac{N_{\text{BN}}}{ME[n_{\text{hops}}]} & , \text{ if (1) and (2)} \\ 0 & , \text{ otherwise} \end{cases}$$

The aggregate Bnet-only throughput  $TH_{\text{Bnet,I}}$  is computed based on the assumption that the following conditions are met under the selected reuse- $M$  and  $d_{\text{BN}}$  values: (1)  $SINR > SINR_t$  setting to meet minimum SINR value requirements. For example, to meet this requirement under  $R_{\text{Bnet}} = 6$  Mbit/s, a minimum reuse  $M = 3$  level is required; under  $R_{\text{Bnet}} = 12$  Mbit/s, a minimum  $M = 4$  value is required; under  $R_{\text{Bnet}} = 24$  Mbit/s, a minimum  $M = 5$  value must be configured; (2)  $P_r > rv_t$  is required to satisfy minimum receiver sensitivity value. For example, under  $R_{\text{Bnet}} = 6$  Mbit/s the latter requirement induces a maximum transmission range (under the employed transmit power level of  $P_{\text{tx}} = 23$  dBm) of about 340 m, while under  $R_{\text{Bnet}} = 24$  Mbit/s, the maximum transmission range is reduced to about 220 m.

For design purposes, we consider a conservative case whereby all BNs are assumed to be continuously residing in transmission mode during  $TP_{\text{Bnet}}$ . The SINR level measured at each active BN is then readily calculated, under each reuse- $M$  level, noting neighboring BNs to be at fixed distance  $d_{\text{BN}}$  from each other. The lowest  $M$  value that meets the minimum SINR level requirement is then selected. This configuration assures Bnet transmissions to

be successfully executed under the prescribed packet error rate.

We coordinate the joint allocation of bandwidth resources to the Anet and Bnet components, noting that the Bnet must carry the traffic that is fed to it by Anet clients. Hence, the corresponding time periods  $TP_{\text{Anet}}$  and  $TP_{\text{Bnet}}$  allocated for Anet and Bnet transmissions in each frame, are set to the proper values, such that one guarantees that the average rate of data messages successfully received by a BN from its Anet mobile clients, supplemented by the rate of disseminated messages that arrive from its neighboring BNs, is equal to the average rate at which data can be served by a BN.

Consequently, to maximize the throughput rate of the shared communications media, yielding full utilization of the allocated per-frame time periods, the ratio of the corresponding time period allocations is set to satisfy the following:

$$\frac{TP_{\text{Bnet}}}{TP_{\text{Anet}}} = \frac{E[n_{\text{hops}}]R_{\text{Anet}}PDR_{\text{Anet}}M}{R_{\text{Bnet}}k_c}$$

The time allocated for Bnet operation, as it applies for transmissions carried out by each BN, has been calculated to include the time taken to serve the traffic arriving from its Anet and the multi-hop traffic arriving from other BNs. Based on the location of the associated BN relative to its Anet clients, the data traffic disseminated across the Bnet is distributed by using a number of hops that vary between

$n_{\text{hops}}$  and  $n_{\text{hops}} + 1$ . In our simulation based analyses, the actual measured average value of the latter variable has been used to allocate the corresponding  $Tp_{\text{Anet}}$  and  $Tp_{\text{Bnet}}$  periods. This average value is denoted as  $E[n_{\text{hops}}]$ . It is noted that we compute the time period durations based on average loading calculations (including the average number of Bnet hops incurred by packet belonging to a specific flow). Consequently, the system should employ a flow control mechanism that serves to block the admission of flows which induce overloading. Accordingly, in considering a BN, the above noted ratio is determined by the time that it requires to transmit messages across the Bnet (which is proportional to  $\frac{ME[n_{\text{hops}}]}{R_{\text{Bnet}}}$ ) vs. the time needed to receive messages transmitted by its Anet clients (which is proportional to  $\frac{k_c}{R_{\text{Anet}} PDR_{\text{Anet}}}$ ), where  $PDR_{\text{Anet}}$  designates the probability that a packet is successfully transmitted across its source Anet, which occurs when it is received at an acceptable SINR level at its targeted BN.

It is noted that active BNs are allocated Bnet slots on a reuse-M basis so that the corresponding data rate and M values are set to assure the BN receiver's SINR is higher than the minimum level required to assure a specified error rate level. In turn, active Anet clients may be located at varying distances from their associated BN. By properly selecting the number of colors  $k_c$ , one can ensure that at least 95 % of the packets will meet the minimum SINR (and thus targeted error rate) requirement.

Assuming the system and access channel to be highly loaded, so that admitted Anet clients are continuously busy, we deduce the following. The aggregate Anet throughput, when considering its operation to be isolated from that of the Bnet in that it is calculated by assuming it to occupy the full bandwidth of the communications system under consideration, is equal to:  $TH_{\text{Anet,I}} = N_{\text{BN}} \frac{R_{\text{Anet}} PDR_{\text{Anet}}}{k_c}$ . The aggregate Bnet throughput is given by:  $TH_{\text{Bnet,I}} = N_{\text{BN}} \frac{R_{\text{Bnet}}}{ME[n_{\text{hops}}]}$ . The aggregate joint Anet-Bnet throughput rate is equal to the average data rate received at the destination BNs. It is equal to:  $TH_{\text{Anet,Bnet}} = TH_{\text{Anet,I}} \frac{Tp_{\text{Anet}}}{T_f}$ , representing the effective data rate carried by the Anets.

To assess the packet delay components, we note the following. The Frame Latency (FL) delay component represents the time elapsed between the arrival (or production) time of a packet at the source Anet client and the time instant at which the mobile transmits its reservation packet. The average FL value is set equal to approximately half of the time frame. FL is incurred only at the start time of a stream flow. We assume that the first packet piggybacks a reservation request. Subsequent flow packets will not experience such FL delays, as we assume that the flow is allocated slots at a rate that matches its requested rate. To state an upper bound, we set a worst case frame latency value of  $FL = T_f$ .

In calculating the end-to-end delay incurred by packets across the system, we account for the delay incurred in traversing the source Anet,  $D_{\text{Anet}}$ , and for the delay associated with the dissemination of the packet across the

Bnet and reception by the destination Anet client,  $D_{\text{Bnet}}$ .

For example, consider an Anet client that produces a stream flow whose application produces packets at a rate of a single packet every  $r_{\text{app}}$  time frames. Such a flow requires the allocation of a single Anet slot every  $r_{\text{app}}$  time frames. The allocation will last for the duration of the flow; say, for  $n_{\text{flow}}$  packet transmissions. The first packet will incur reservation delay, including a FL component and further delay until its reservation packet is successfully received at its BN (using, for example, a slotted ALOHA reservation access protocol). Once the reservation / first packet has been successfully sent, subsequent flow packets are transmitted at periodically assigned time slots, so that no further latencies are incurred. Other models, including such that account for random reservation delays, could be readily included as well. For example, when its Anet is not highly loaded, the BN can proceed to announce also idle service slots to be available for the transmission of reservation packets, yielding a much reduced reservation latency. As the traffic loading increases, the BN can announce the assigned reservation slot(s) to be available for reservations made by only the highest priority client mobiles, assuring a targeted lower reservation delay level for such messages. Clearly, as a higher capacity is allocated for reservations, a lower residual capacity remains available to support the transmissions of data packets with reservations. In the performance analyses presented in the following sections, we however focus on determining an upper bound on the throughput performance rate attainable under the demand-assigned scheme. Hence, we assume there that just a single reservation slot per frame is allocated, when the system is subjected to high loading. Furthermore, by implementing the above described dynamic scheme to allocate reservation slots, by employing flow admission controls, and by assuming flow durations to be of the order that assures otherwise high probability of success in reservation transmissions across the allocated reservation slot in each frame, we further assume there that the targeted reservation latencies are met. Other analyses are readily carried out when making other reservation process or delay assumptions.

By using the following flow control scheme, we provide admitted packets with a bounded delay level. All packets that reach a BN during a period (of time frame duration) that precedes the Bnet service period (whose duration is equal to  $Tp_{\text{Bnet}}$ ) and are targeted for transmission across the Bnet, are targeted for service during the current Bnet service period. Such packets include all packets that arrive across the Bnet from neighboring BNs and must be forwarded to other BNs, which can be provided higher priority for service by the BN. It also includes Anet packets that are received by the BN from its Anet clients during the preceding Anet period (whose duration has been set to  $Tp_{\text{Anet}}$ ). By using flow admission control for regulating the admission of new Anet flows to a BN, using sliding window measurements of transit traffic and thus concluding the rate available to support newly admitted flows, one can assure an end-to-end packet delay of the order of

$D_p \leq FL + T_f \leq 2T_f$  (plus the impact of reservation delay when applicable, normalized in relation to the average flow duration as only the first packet of the flow incurs such delay). In turn, if a higher end-to-end packet delay level is acceptable, a more relaxed access regulation scheme can be employed.

## 6.2. TDMA Bnet - IEEE 802.11p Anet (MAC 2)

In this section, we discuss the performance features of a system that employs an IEEE 802.11p type CSMA/CA MAC scheme for Anet access and a TDMA scheme in sharing the Bnet communications resources. Anet clients thus use a contention based CSMA/CA access scheme. We assume the carrier sensing range to be set at  $d_{CS}$ . A mobile that is currently in the process of transmitting its message, prevents other mobiles (which may be located in the same Anet or in neighboring Anets) that are within carrier sensing range from it from initiating new transmissions. To reduce interference signals caused by simultaneous transmissions occurring in neighboring Anets, we employ a reuse- $k_c$  Anet coloring scheme. Through the execution of system simulations, we have determined that the highest achievable Anet throughput rate,  $TH_{Anet,I}$ , is attained when using  $k_c = 1$  or  $k_c = 2$  scheme. Furthermore, the ensuing throughput rate values were noted to be about the same. At low loading rates, the carrier sensing blocking effect is noted to be equivalent to a coloring level of approximately  $d_{CS}/d_{BN}$ . At high data traffic rate, the PDR value attained by using  $k_c = 2$  is noted to be about twice higher but the nodes are noted to be able to initiate transmissions only half the time, resulting in comparable overall throughput rates.

The Bnet is configured to use a properly configured TDMA based reuse-M and  $R_{Bnet}$  levels, considering  $M = 3, 4$  and  $5$ . As discussed above, the reuse-M level is configured to attain sufficiently high SINR and received power, while achieving high throughput rates.

Anet and Bnet periods share the SCH resources on a time division basis. Thus, each time frame  $T_f$  consists of disjoint time slots and periods allocated for Bnet and Anet transmissions. The durations of Anet and Bnet time periods are denoted as  $Tp_{Anet}$  and  $Tp_{Bnet}$ , respectively.

Corresponding time periods are allocated within each frame for Anet and Bnet packet transmissions:

$$T_f = Tp_{Anet} + Tp_{Bnet}$$

The corresponding periods are sized to accommodate the transmission of admitted Anet and Bnet packets.

To simplify the implementation of the access scheme,  $Tp_{Anet}$  and  $Tp_{Bnet}$  are allocated on a fixed ratio basis, which is calculated based on  $TH_{Anet,I,max}$ , which represents the highest attainable Anet throughput rate over the loading range under consideration when the Anet system is operated separately from the Bnet. The allocation is set to guarantee that the average data rate which is received by a source BN from its Anet clients,  $TH_{Anet,I,max}/N_{BN}$ ,

supplemented by the data rate for packets received from neighboring BNs, is equal to the average rate at which data can be served (i.e., transmitted) by a BN.

For a given  $R_{Anet}$ , as the data load produced by Anet clients is increased, the aggregate Anet throughput rate (in isolation),  $TH_{Anet,I}$ , increases accordingly until it is measured to reach its maximum value, denoted as  $TH_{Anet,I,max}$ . This throughput performance is impacted by the stochastic occurrence of collisions and re-transmissions across the 802.11p-channel. It is further noted that packets that are re-transmitted four times are subsequently discarded and assumed to not be successfully delivered.

Accordingly, we set the ratio of durations of the corresponding periods to satisfy the following relationship:

$$\frac{Tp_{Bnet}}{Tp_{Anet}} = \frac{TH_{Anet,I,max}E[n_{hops}]M}{N_{BN}R_{Bnet}}$$

where  $E[n_{hops}]$  is defined as stated for the MAC 1 scheme. To evaluate the performance of the MAC 2 scheme, we carry out Monte Carlo simulations for the 802.11p-based Anet system in isolation. We set mobiles that belong to distinct Anets to be simultaneously active to account for the impact of inter-Anet interference signals. The performance of the TDMA based Bnet system in isolation is carried out by using the mathematical expressions presented in Section Section 6.1. The resulting performance metrics are then used to calculate the Anet and Bnet per-time-frame periods  $Tp_{Anet}$  and  $Tp_{Bnet}$  by using the mathematical relationship stated above. The realized combined system throughput rate is then given by

$$TH_{Anet,Bnet} = TH_{Anet,I} \frac{Tp_{Anet}}{T_f}$$

The data traffic loading each BN consists of data packets received from the Anet mobiles associated with this BN and of data packets that are received from neighboring BNs. Hence, the aggregate data packet arrival rate at a BN is equal to  $\lambda_{BN} = E[n_{hops}] \frac{TH_{Anet,I}}{N_{BN}} \frac{Tp_{Anet}}{T_f} \frac{1}{P}$  (packets/sec). It is noted that the packet arrival rate at each BN depends on the number of dissemination hops taken by a flow, on the coloring factor  $k_c$  and on the employed Anet data rate  $R_{Anet}$ , through the realized  $TH_{Anet,I}$ , and the per-frame period allocated for Anet transmissions. The service rate provided by a BN,  $\mu_{BN}$ , is expressed as  $\mu_{BN} = \frac{R_{Bnet} Tp_{Bnet}}{M(T_f)} \frac{1}{P}$  (packets / sec). It is noted that  $\mu_{BN}$  depends on the proportion of time allocated for Bnet transmissions within the time frame, on the reuse-M factor used by the Bnet's TDMA mechanism, and on the employed Bnet data rate  $R_{Bnet}$ .

The time delay incurred by a packet across its source Anet, denoted as  $D_{Anet}$ , is measured from the instant of its generation to the instant that it is successfully transmitted to the associated BN (given that its flow has been admitted) is obtained through the execution of Monte Carlo simulations. The delay time incurred by a packet at each BN, denoted as  $d_{BN}$ , has been approximated by using analytical calculations.

The waiting time performance incurred by a packet at each BN, denoted as  $W_{\text{BN}}$ , has been evaluated by using the following approximations. Within its Bnet per-frame service period, each BN forwards an average fraction  $(E[n_{\text{hops}}] - 1)/E[n_{\text{hops}}]$  of the packets that it receives from another BN. A fraction of the packets that it handles which is equal to  $1/E[n_{\text{hops}}]$  is received from its Anet clients. The same fraction also represents the rate of packets that it transmits to the destination mobiles which reside within its Anet.

To reduce packet latency levels while being disseminated across the Bnet, we capitalize on the TDMA structure of the Bnet MAC, by employing a priority oriented service policy at the BN. The BN queueing processor grants higher service priority to transit packets (which arrive from other BNs), while serving packets that it receives directly from its Anet at a lower priority level. Assuming a fixed dissemination span setting for all packet flows, packets that a BN receives from its neighboring nodes for relay to other backbone nodes are perfectly scheduled for transmission by the BN in accordance with the setting of the reuse-M TDMA operation cross the backbone, and thus do not incur queueing delays at intermediate backbone nodes. In turn, packets received by a BN from its Anet clients arrive to the BN in a stochastic manner, which is impacted by the random times at which successful transmissions occur across the IEEE 802.11p based channel sharing scheme, and hence incurred queueing delays at their source backbone node. To analyze this priority queueing system, we carry out an approximate calculation of the queueing (i.e., waiting time) delays incurred by second priority Anet packets at a BN by modeling the associated BN queueing system that is used to serve the latter Anet packets as the following modified M/M/1 queueing system, whereby the service rate of lower priority packets is reduced by that dedicated for servicing higher priority packets. Consequently, the cumulative distribution function of the Anet packet waiting time ( $W$ ) at a BN, denoted as  $W_x = P(W \leq x)$ , is given by:

$$W_x = 1 - \rho_{\text{BN},s} e^{-\mu_{\text{BN},s}(1-\rho_{\text{BN},s})x}$$

where  $\mu_{\text{BN},s}$  and  $\rho_{\text{BN},s}$  denote the service rate granted to and traffic intensity offered by lower priority packets, respectively, arriving from Anet sources at their source BNs. We have:  $\mu_{\text{BN},s} = \frac{1}{E[n_{\text{hops}}]} \mu_{\text{BN}}$  and  $\rho_{\text{BN},s} = \frac{\lambda_{\text{BN}}/E[n_{\text{hops}}]}{\mu_{\text{BN}}/E[n_{\text{hops}}]}$ .

In carrying out the system's performance evaluations, we use the above noted expression for the packet waiting time distribution at the source BN to determine the allowable corresponding queueing delay incurred at the BN, which in conjunction with the delay incurred by a packet in traversing the source Anet, and in conjunction with the involved fixed latencies involved in disseminating a packet across its path, leads to the calculation of the end-to-end packet delay value.

The targeted end-to-end packet delay,  $D_P$ , is then calculated as the sum of the FL delay, the ( $D_{\text{Anet}}$ ) delay time incurred across the source Anet, the waiting time incurred

at the queue of the source BN, and the total transmission time experienced in disseminating the packet across the Bnet, which is equal to  $E[n_{\text{hops}}] \frac{P}{R_{\text{Bnet}}}$ , yielding:

$$D_P = FL + D_{\text{Anet}} + D_{\text{Bnet}}$$

In our simulations runs, we examine varying  $TH_{\text{Anet},I}$  levels. For each iteration, we calculate the 95-percentile delay incurred by a packet traversing its Bnet path, using the priority queueing based analysis presented above. By subtracting this delay value from the specified end-to-end level, we obtain a value that is used as the targeted 95-percentile delay to be incurred by packets transmitted across their source Anet. It is noted that, as an approximation, when Anet and Bnet packet delays are assumed to behave as independent random variables, the attained aggregate end-to-end packet delay value is then guaranteed to hold for at least 90-percent of the packets.

For illustrative purpose, we assume to size one time frame  $T_f$  to accommodate  $M$  Bnet slots. This serves to reduce the  $FL$  induced latency component.

As noted above for the previously considered MAC scheme, a flow control scheme is enacted to ensure that traffic with higher than average number of hops is admitted only if under current conditions it can be served while incurring acceptable packet delay levels.

### 6.3. IEEE 802.11p Anet and IEEE 802.11p Bnet (MAC 3)

The third MAC scheme examined assumes the use of an IEEE 802.11p protocol in sharing the Anet channels and the Bnet links. We assume that the employed bandwidth is shared by all BNs and non-BNs (clients) on an IEEE 802.11p contention access basis. An entity that wishes to gain access to the channel for the successful transmission of its packet must contend with currently active nodes that reside within a range of  $d_{\text{CS}}$ . Active nodes that are located further away may also cause interference at the intended receiver when the total interference energy detected at the receiver is sufficiently high. To simplify our evaluation, we assume that the same transmission data rates are used across each Anet and the Bnet.

## 7. Performance Behavior of the Data Network Systems

### 7.1. DA/TDMA Anet - TDMA Bnet (MAC 1)

In this section, we discuss the performance behavior of the DA/TDMA-Anet / TDMA-Bnet scheme. First, we display the performance of a sole Anet system, showing the dependence of the attained throughput  $TH_{\text{Anet},I}$  on the data rate  $R_{\text{Anet}}$  and  $k_c$  reuse factor. In Figure 12, we show the corresponding performance, assuming the following parameter values:  $N = 250$ ,  $l = 5$  km,  $\psi = 2$  with  $d_{\text{BN}} = 100, 150$  and  $300$  m, inducing the respective number of clients served per BN to be  $n_{\text{served}} = 5, 7$  and  $15$ .

The performance evaluations carried out to obtain  $TH_{\text{Anet},I}$  are based on running a Matlab-based Monte Carlo

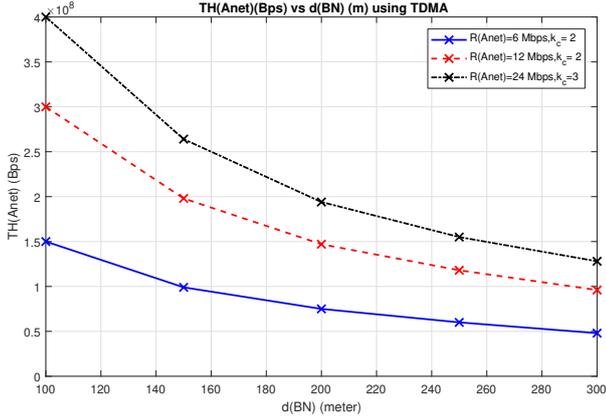


Figure 12: Aggregate Anet Throughput  $TH_{Anet,I}$  (bit/s) vs  $d_{BN}$  (m) using DA/TDMA,  $l = 5$  km

simulation. The simulation has been used to determine the power of the interference signal induced by simultaneous uplink transmissions executed by multiple Anet clients. It is noted that Anet and Bnet transmission operations are carried out over separate time periods. Simultaneous transmissions by multiple vehicles are performed in accordance with the employed reuse- $k_c$  scheme. The  $k_c$  reuse level is selected to assure a  $PDR \geq 0.95$  level.

The higher the  $R_{Anet}$  value, the higher one must set the number of Anet  $k_c$ , as a higher  $R_{Anet}$  value requires the receiver to detect a higher minimum SINR value, noting that a higher Anet coloring value leads to better mitigation of inter-Anet interference. For example, specifying a  $PDR \geq 0.95$ , and setting  $R_{Anet} = 24$  Mbit/s, induces an optimum setting (yielding the highest Anet throughput) of  $k_c = 3$  colors. The latter value is higher than that required when a lower data rate  $R_{Anet} = 6$  Mbit/s is used, which requires setting  $k_c = 2$  colors. As shown in Figure 12, the throughput  $TH_{Anet,I}$  is reduced as the inter-BN range  $d_{BN}$  increases. This main cause for this drop is attributed to the ensuing decrease in the number of elected BNs, which in turn reduces the total Anet and Bnet communications capacity that is made available since each BN makes use of its own associated transmission channel resources. Under a set  $d_{BN}$  value, the attained  $TH_{Anet,I}$  increases as the employed data rate  $R_{Anet}$  increases. Yet, this increase is moderated by the need to use a higher number of Anet colors.

The behavior of the attained Bnet only throughput rate as a function of the inter-BN distance is depicted by Figure 13. The corresponding performance results have been obtained by using the mathematical formula presented in subsection 6.1. For inter-BN ranges of  $d_{BN} = 100$  m and  $d_{BN} = 150$  m, the highest throughput is obtained by setting  $R_{Bnet} = 24$  Mbit/s,  $M = 5$ . In turn, when configuring  $d_{BN} = 300$  m, the highest throughput is obtained by setting  $R_{Bnet} = 6$  Mbit/s,  $M = 3$ . We further note that the corresponding curves for the throughput rate exhibit some performance fluctuations as the inter-Bnet range increase.

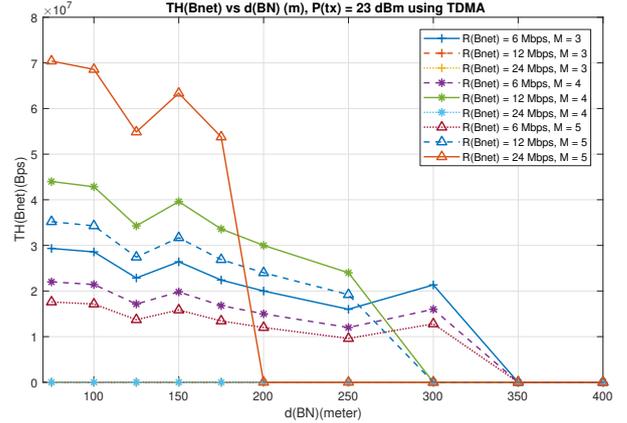


Figure 13:  $TH_{Bnet,I}$  (bit/s) vs  $d_{BN}$  (m) for  $P_{tx} = 23$  dBm

These variations are explained by noting that as the inter-BN distance is raised, the number of dissemination hops traversed across the Bnet may decrease (showing quantization oriented fluctuations since the later number assumes integral values), serving to reduce the internal traffic produced across the Bnet links. In turn, the corresponding number of employed BNs also decreases causing reduction in the total bandwidth made available to support Bnet transmissions.

We also note that as the  $d_{BN}$  range is further increased, say to 300 m,  $R_{Bnet}$  must be decreased to 6 Mbit/s. This is caused by the ensuing decrease in the received power level, as the minimum received power level may then be reduced below the required receiver sensitivity value. To resolve such an issue, we have studied a system that uses an increased transmission power level, setting  $P_{tx} = 33$  dBm. We observed the increased transmit power system to not lead to increased throughput performance when setting shorter inter-BN range distances, such as  $d_{BN} = 100$  m or 150 m. However, an increase in the transmit power was noted by us to yield a significant throughput upgrade under longer inter-BN distances such as  $d_{BN} = 300$  m, under  $R_{Bnet} = 24$  Mbit/s, as it enabled the system to then meet the required minimum receiver sensitivity level. Also, using higher transmit power levels could also be useful when considering longer dissemination spans than those considered in our illustrative scenarios in this paper. It is also noted that by increasing the Bnet data rate from 6 Mbit/s to 24 Mbit/s, the resulting throughput rate increases in a less than proportional manner due to the ensuing increase in the reuse- $M$  level.

In Figure 14, we show the variation of the optimum end-to-end combined Anet-Bnet throughput rate,  $TH_{Anet,Bnet}$ , as a function of the configured inter-BN distance value. The performance of this system was obtained by using a hybrid simulation and analytical process, as expressed by the mathematical formula for  $TH_{Anet,Bnet}$  presented in subsection 6.1. We specify the targeted PDR to be higher than 0.95, and the packet delay to be lower than 50 ms for 95 % of the served packets. The highest throughput rate is

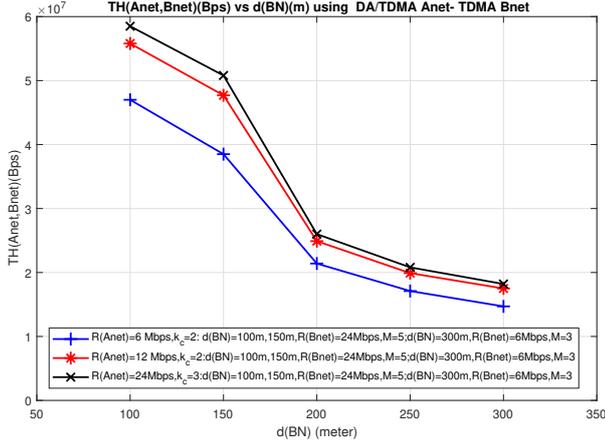


Figure 14:  $TH_{Anet,Bnet}$  (bit/s) vs  $d_{BN}$  (m) using DA/TDMA-TDMA with  $P \geq 0.95$  ( $D_P \leq 50$  ms) and  $PDR \geq 0.95$

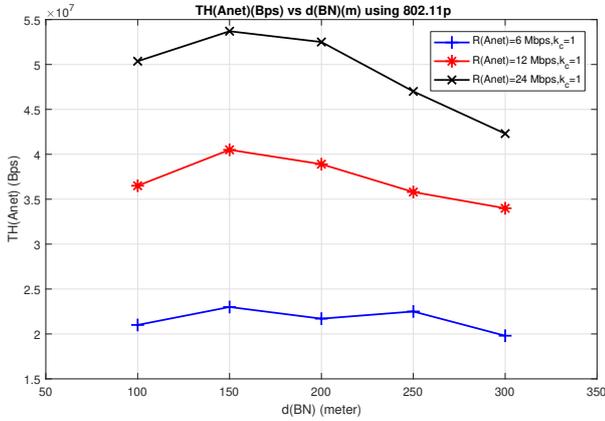


Figure 15: Maximum delay-capped  $TH_{Anet,I}$  (bit/s) and  $PDR \geq 0.95$  vs  $d_{BN}$  (m) for IEEE 802.11p Anet

noted to be attained when setting  $R_{Anet} = 24$  Mbit/s,  $k_c = 3$ ,  $R_{Bnet} = 24$  Mbit/s,  $M = 5$  at  $d_{BN} = 100$  m. As  $d_{BN}$  increases from 100 m to 150 m, the system's throughput rate  $TH_{Anet,Bnet}$  is reduced due mainly to the lower number of employed BNs. As  $d_{BN}$  is further increased to 300 m, a more significant throughput drop is observed, as now a lower  $R_{Bnet} = 6$  Mbit/s must be used to meet the minimum SINR requirement at the targeted receiver. We have examined setting different Anet and Bnet data rates to determine the highest achievable system throughput rate. We note that for  $d_{BN} = 100$  m, setting  $R_{Anet} = 6$  Mbit/s yields a lower throughput rate than that achieved when setting  $R_{Anet} = 24$  Mbit/s; yet, the difference is low as the performance when using  $R_{Anet} = 6$  Mbit/s is then dominated by the setting of the same Bnet parameters,  $R_{Bnet} = 24$  Mbit/s,  $M = 5$  and a lower Anet-coloring value  $k_c = 2$ .

## 7.2. TDMA Bnet and IEEE 802.11p Anet (MAC 2)

The attained maximum  $TH_{Anet,I}$  as calculated without imposing a packet delay limit is used to calculate the Anet and Bnet time periods configured within each frame. We have noted  $R_{Anet} = 24$  Mbit/s corresponding attained

optimum throughput values of 69.8 Mbit/s at  $d_{BN} = 100$  m, 66.7 Mbit/s at  $d_{BN} = 150$  m, 53 Mbit/s at  $d_{BN} = 300$  m.

The delay constrained throughput performance is shown in Figure 15. Results have been obtained through the execution of a NS-3 based simulation coupled with analytical evaluations. Under the Monte-Carlo simulation process, the Anet client packet generation rate is gradually increased, leading to variations in the resulting Anet throughput rate. The latter induces the packet arrival rate levels loading the source BN, as generated by the BN's Anet clients, from which the traffic intensity parameter  $\rho_{BN,s}$  presented in Section 6.2 is analytically calculated. It is subsequently used to compute the ensuing end-to-end packet delay  $D_P$  level. The realized Anet throughput rate under which the prescribed delay  $D_P$  level is satisfied represents the  $TH_{Anet,I}$  value shown in Figure 15. As expected, for this case, a lower than  $TH_{Anet,I,max}$  Anet throughput rate  $TH_{Anet,I}$  is realized when aiming to satisfy delay constraint.

The Bnet throughput  $TH_{Bnet,I}$  performance behavior for this scheme is the same as that exhibited above under the DA/TDMA Anet - TDMA Bnet scheme, shown in Figure 13. We have studied the setting of possibly different Anet and Bnet data rate values in aiming to achieve the highest system throughput rate.

The maximum delay-capped  $TH_{Anet,I}$  value achievable for  $R_{Anet} = 24$  Mbit/s is equal to about 54 Mbit/s at  $d_{BN} = 150$  m. As  $d_{BN}$  increases from 100 m to 150 m, the attainable  $TH_{Anet,I}$  increases because the dissemination span involves a lower number of Bnet hops, inducing lower backbone data loading rate. The lower number of incurred Bnet hops used across the path of a flow further induces lower per-BN packet delay requirement. This consequently allows Anets to carry higher traffic loads. As  $d_{BN}$  further increases to 300 m, attainable  $TH_{Anet,I}$  values decrease. At this distance range, a lower data rate across the Bnet must be used by setting  $R_{Bnet} = 6$  Mbit/s, thus reducing the effective service rate available at a BN for the transmission of arriving packets, causing the waiting time incurred at source BN to increase. Consequently, to meet packet delay prescriptions, the rate of packet flows arriving to a BN from its Anet mobiles must be lowered, resulting in a lower  $TH_{Anet,I}$  level. Such mobile transmissions also tend then to become more sensitive to signal interference caused by simultaneously occurring transmissions that take place outside the underlying carrier sensing range  $d_{CS}$ , which is equal to about 346 m when  $P_{tx} = 23$  dBm. As  $d_{BN}$  increases, the average ratio of the power received from an intended transmitter to the power received from signal interference decreases.

The  $TH_{Anet,Bnet}$  performance behavior is shown in Figure 16. These results have been obtained by using the mathematical formula presented in Section 6.2, which includes parameters evaluated by using the NS-3 based simulation process for the Anet system coupled with the use of analytical calculations for the Bnet system.

As shown in Figure 16, the setting of  $d_{BN} = 150$  m,  $R_{Anet} = 24$  Mbit/s,  $R_{Bnet} = 24$  Mbit/s,  $M = 5$ , yields

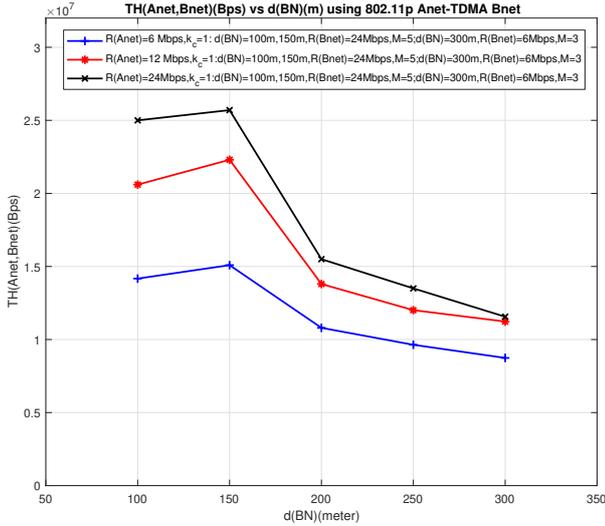


Figure 16: Delay-capped  $TH_{Anet,Bnet}$  (bit/s) vs  $d_{BN}$  (m) for IEEE 802.11p Anet-TDMA Bnet

the highest delay-capped system throughput  $TH_{Anet,Bnet}$  value. It is followed by the setting of  $d_{BN} = 100$  m and then  $d_{BN} = 300$  m. The  $TH_{Anet,Bnet}$  value is not sensitive to changes in  $d_{BN}$  when  $d_{BN}$  is small; however, a significant drop is noticed at longer  $d_{BN}$  ranges such as 300 m. The  $TH_{Anet,Bnet}$  values are insensitive when increasing  $d_{BN}$  slightly from  $d_{BN} = 100$  m to  $d_{BN} = 150$  m as inducing a slight  $TH_{Anet,Bnet}$  increase from 25 Mbit/s to 25.3 Mbit/s, although the number of hops is reduced by 1. To explain, we note that low  $d_{BN}$  values such as 100 m and 150 m are within the  $d_{CS}$  range, resulting in an overall lower number of packet collisions. A key factor impacting the attained end-to-end delay level is noted to be contributed by the waiting time value incurred at the queueing module of the source BN, rather than the delay level incurred across the Anet channel,  $D_{Anet}$ , and the latency values associated with transmissions across Bnet links. When compared to a configuration that uses  $d_{BN} = 100$  m, as we increase the latter to  $d_{BN} = 150$  m, the rate of transit packet flows loading a BN queueing system is noted to be reduced. Hence, packets arriving to a BN from its Anet are served at a higher rate. Consequently, the waiting time incurred by such packets at their source BN is improved. The throughput rate supported across each source Anet can consequently be allowed to increase (as packets can then be allowed to incur higher Anet delays). However, since the number of Anets is reduced, the attained aggregate throughput rate is not expected to change in a noticeable manner, as confirmed by the depicted simulation results.

Hence, the net delay gain achieved by the reduced number of hops is not significant, while just a slight increase in the  $TH_{Anet,I}$  level is attained. Furthermore, as  $d_{BN}$  is increased from 100 m to 150 m, the SINR level recorded at a BN receiver may be reduced as the receiver may be located closer to sources of signal interference that are located outside  $d_{CS}$ . We also note that at higher inter-BN

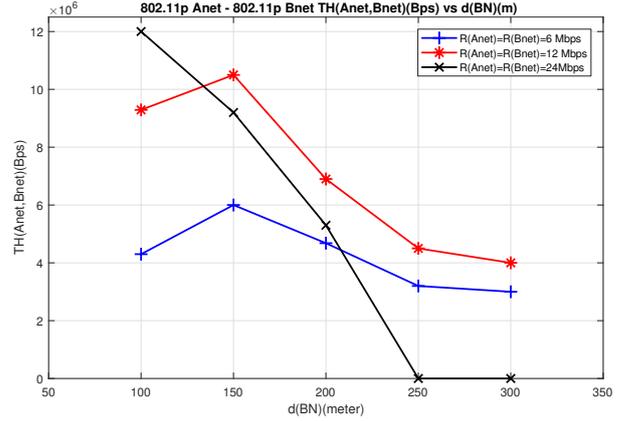


Figure 17:  $TH_{Anet,Bnet}$  (bit/s) vs  $d_{BN}$  (m) at various  $R$  (Mbit/s)  $P \geq 0.95$  ( $D_P \leq 50$  ms),  $PDR \geq 0.95$  for IEEE 802.11p Anet - IEEE 802.11p Bnet

ranges, each Anet would need to support a higher number of client mobiles, which can lead to higher collision rates. In addition, since a lower number of BNs are then employed, an overall lower Anet bandwidth level is available. The performance results depicted in Figure 16 show that the setting of  $R_{Anet} = 24$  Mbit/s leads to a higher delay-capped system throughput  $TH_{Anet,Bnet}$  values.

A significant drop in the throughput rate is observed as  $d_{BN}$  increases from 150 m to 300 m under  $R_{Anet} = 24$  Mbit/s. This behavior is mainly due to the lower value exhibited by the maximum throughput rate attainable by data packets that flow from Anet vehicles to their source BNs  $TH_{Anet,I}$ , under the specified packet delay constraint, as shown in Figure 15, induced by the low Bnet data rate,  $R_{Bnet}$ , needed to satisfy the SINR value required under the long  $d_{BN}$  level. Furthermore,  $d_{BN} = 300$  m is noted to be close to the carrier sensing range,  $d_{CS}$ . Coupled with the observation that the SINR is then very low and is quite close to the minimum SINR required for operation at  $R_{Anet} = 24$  Mbit/s, which induces a low PDR value. The  $TH_{Anet,Bnet}$  value attained when no packet delay limits are imposed, has been determined by us (not shown) to be 15–35% higher than the corresponding allowed delay-capped throughput rate.

Our performance evaluation results show that the attained throughput values are insensitive to variations in the number  $n_{clients}$  of Anet clients. A higher number of such clients does not lead to a significant increase in the packet collision rate, due in a large extent to the impact of the carrier sensing range. We have also noted that the maximum delay-capped  $TH_{Anet,I}$  throughput level does not vary much as the number of Anet clients increases, provided the overall loading rate per Anet is controlled (so that the maximal allowed total offered loading rate is regulated).

### 7.3. IEEE 802.11p Bnet and IEEE 802.11p Anet (MAC 3)

In Figure 17, we show the variation of the system throughput rate under IEEE 802.11p Anet and Bnet MAC

schemes vs. the setting of the inter-BN range. These performance results have been obtained by performing Monte Carlo NS-3 based simulations. We focus on the performance realized by Class 2 messages, assuming Class 1 message delay and PDR requirements to have been met, assuming a single lane loading of  $N = 250$ . The setting of a (Anet and Bnet) data rate  $R = 24$  Mbit/s and  $d_{BN} = 100$  m is shown to yield the maximum throughput rate. Under this high data rate, the throughput rate experiences significant degradation as the the inter-BN distance is increased. In turn, we observe the attained throughput values under  $R = 12$  Mbit/s to be less sensitive to changes in  $d_{BN}$  due to its lower minimum SINR requirement. As  $d_{BN}$  is increased from 150 m to 300 m at  $R_{Bnet} = 24$  Mbit/s, the steep drop in data throughput is caused by the increased sensitivity to signal interference measured at the BN receiver due to a higher minimum SINR requirement, and by the low value of the received power value of the intended signal used for data transmissions between BNs.

The observed performance result is insensitive to variation of  $d_{BN}$  from 100 m to 150 m level, for  $R_{Anet} = R_{Bnet} = 6$  and 12 Mbit/s. Although the average number of hops decreases from 3.5 hops to 2.5 hops, longer  $d_{BN}$  value means that an intended transmission by some Anet clients will become more sensitive to signal interference caused by simultaneous transmissions occurring due to sources outside the  $d_{CS}$  range while also a reduced overall Anet bandwidth is available due to the reduced number of BNs configured along the highway. Hence, the gain in throughput attributed to the employed smaller number of hops is partially offset by the lower SINR levels that can be induced.

#### 7.4. Comparison between the MAC schemes

In Figure 18, we compare the delay-capped highest aggregate system throughput rate performance attained under the three MAC schemes under consideration. We note that the TDMA Anet-TDMA Bnet scheme yields the highest throughput rate (which is equal to about 58.5 Mbit/s). It is followed by the Bnet TDMA-Anet/IEEE 802.11p scheme, achieving a corresponding throughput rate of about 25.6 Mbit/s, and then by the joint-IEEE 802.11p scheme, which produces a throughput rate of about 12 Mbit/s). The highest throughput rate is attained by the TDMA-Anet/TDMA-Bnet scheme when configured to operate at  $R_{Anet} = 24$  Mbit/s,  $R_{Bnet} = 24$  Mbit/s,  $M = 5$ ,  $d_{BN} = 100$  m,  $k_c = 3$ . When examining the sensitivity of the highest attainable throughput rate to variation in the setting of the inter-BN distance, we conclude that the TDMA/TDMA scheme generally yields the best performance when  $d_{BN}$  is set to a value that is lower than 300 m.

By examining the delay capped throughput performance results exhibited in Figures 14, 16 and 17, we note the following. The highest throughput rate is attained by setting  $d_{BN} = 100$  m,  $R_{Anet} = 24$  Mbit/s,  $R_{Bnet} = 24$  Mbit/s,  $M = 5$ ,  $k_c = 3$ ,  $n_{active} \leq 10$ , yielding a throughput rate that is equal to 58.5 Mbit/s, under the Anet

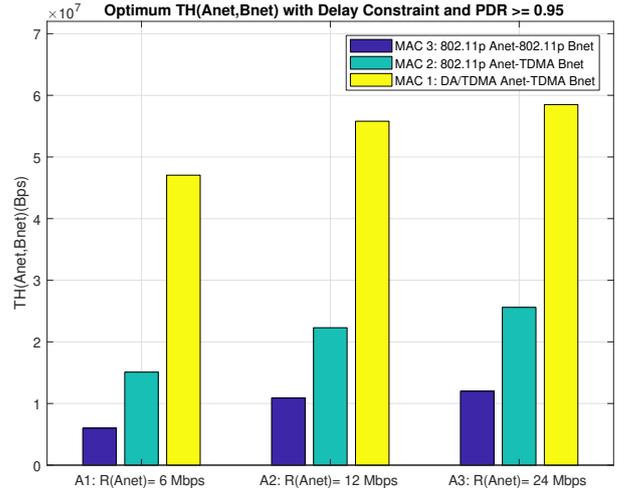


Figure 18: Comparison Delay-capped  $TH_{Anet,Bnet}$  (Mbit/s) vs  $R_{Anet}$  (Mbit/s) for  $N = 250$ . Bar explanation: (A1/A2/A3,MAC3):  $R_{Anet} = R_{Bnet}$ ; (A1/A2/A3,MAC2):  $R_{Bnet} = 24$  Mbit/s,  $M = 5$ ,  $d_{BN} = 100$  m; (A1/A2,MAC1):  $k_c = 2$ ,  $R_{Bnet} = 24$  Mbit/s,  $M = 5$ ,  $d_{BN} = 100$  m; (A3,MAC1):  $k_c = 3$ ,  $R_{Bnet} = 24$  Mbit/s,  $M = 5$ ,  $d_{BN} = 100$  m

TDMA - Bnet TDMA scheme. The next highest throughput when using the Anet TDMA - Bnet TDMA scheme, yielding  $TH_{Anet,Bnet} = 50.9$  Mbit/s, is attained by setting  $d_{BN} = 150$  m, in using the following configuration:  $R_{Anet} = 24$  Mbit/s,  $R_{Bnet} = 24$  Mbit/s,  $M = 5$ ,  $k_c = 3$ ,  $n_{active} \leq 13$ . The third highest attainable throughput rate is  $TH_{Anet,Bnet} = 25$  Mbit/s, achieved by using the 802.11p Anet/TDMA Bnet scheme and setting  $d_{BN} = 100$  m,  $R_{Anet} = 24$  Mbit/s,  $R_{Bnet} = 24$  Mbit/s,  $M = 5$ .

## 8. Integrated System Design when joining Traffic Management and Networking Performance Objectives

The results presented by us in the traffic management sections point to the configuration options available to the designer when aiming to achieve high vehicular flow rates under vehicular delay limits. The results presented in the networking sections identify the dependence of the system message communications throughput rate on the system traffic parameter settings. In this section, we demonstrate how these results are combined to deduce an integrated systems design that jointly meet prescribed vehicular and message communications performance objectives. For illustrative purposes, we consider a single lane single link highway segment of length  $l = 5$  km, as illustrated in Figures 2 to 4. As shown by us in the traffic management part, for  $n = 2$  to  $n = 10$ , one should set the optimal speed to about 60 km/h. The corresponding vehicular flow rate that results varies from 1800 veh/h to 3250 veh/h, and the  $d_{PL}$  ranges vary from 60 m to 175 m. The optimal speed should be configured to about 60 km/h when  $n = 10$  and  $d_{PL} = 175$  m, yielding a vehicular throughput (flow rate) of 3250 veh/h.

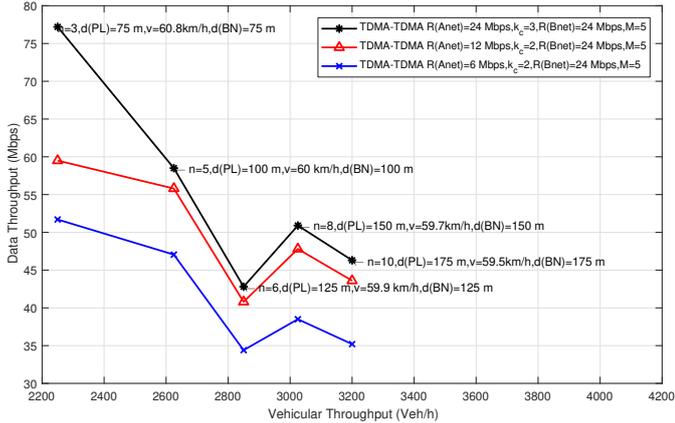


Figure 19: Data Throughput (Mbit/s) vs Vehicular Throughput (veh/h)

To illustrate, assume the system manager aims to configure a system that achieves a vehicular throughput flow rate level that assumes values in a range which includes a high 3000 veh/h level and a medium 2000 veh/h level. At the same time, the system manager aims to attain a sufficiently high delay-capped packet communications throughput performance. This throughput performance accounts for the dissemination of Class 2 messages, assuming that Class 1 messages are guaranteed a high PDR,  $PDR \geq 0.95$ , and an average delay that is not higher than 50 ms for successfully disseminated messages.

In Figure 19, we illustrate the tradeoffs available to the system designer as we observe that different design settings yield different combined (vehicular throughput, communications throughput) operational points. The performance points identified in the figure as we traverse each design combination (involving specified system parameters) by varying the  $d_{PL}$  values over the set 75, 100, 125, 150 and 175 m. For each case, we use the parameter values attained as the result of the optimization process carried out for the DA/TDMA Anet-TDMA Bnet scheme.

To illustrate the underlying performance tradeoff available to the designer, assume first that it is of interest to achieve a high vehicular throughput level, equal to 3000 veh/h. It is attained, as shown in Figure 3, by setting  $d_{PL} = 150$  m and  $n = 8$ . Under these parameter setting, we aim to maximize the data communications throughput performance in disseminating Class 2 message flows, while also meeting Class 1 message communications requirement. Assuming that  $d_{BN}$  is selected from the set 100, 150 and 300 m, we determine that the values  $d_{BN} = 150$  m and  $d_{BN} = 300$  m are feasible as they satisfy  $d_{BN} \geq d_{PL}$ . Accordingly, we determine the setting  $d_{BN} = 150$  m,  $R_{Anet} = 24$  Mbit/s,  $R_{Bnet} = 24$  Mbit/s,  $M = 5$ ,  $k_c = 3$  to yield a higher data throughput performance (when compared with the setting  $d_{BN} = 300$  m,  $R_{Anet} = 24$  Mbit/s,  $R_{Bnet} = 6$  Mbit/s,  $M = 3$ ,  $k_c = 3$ ). In consequently configuring  $d_{BN} = d_{PL} = 150$  m, the communication networking function requires a setting for which we

have  $n_{served} \leq 13$  in considering Anet clients which become active. Hence, all the platoon members can be served by the BN. The joint traffic management and networking parameters are therefore set as follows. The traffic management parameters are set to:  $d_{PL} = 150$  m ( $v = 60$  km/h) and  $n = 8$ . The optimal networking configuration synthesized to achieve a high vehicular throughput implements the following settings:  $d_{BN} = 150$  m,  $R_{Anet} = 24$  Mbit/s,  $R_{Bnet} = 24$  Mbit/s,  $M = 5$ , yielding an aggregate system throughput rate that is equal to  $TH_{Anet,Bnet} = 50.9$  Mbit/s.

In turn, when the system designer is willing to accommodate a medium vehicular throughput rate level of about 2750 veh/h, achievable by setting  $n = 5$  at the optimal velocity level of  $v = 60$  km/h and a  $d_{PL}$  range set to 100 m. The  $d_{BN}$  level should be selected such that the networking configuration satisfies  $d_{BN} \geq d_{PL} = 100$  m (e.g.  $d_{BN} = 100$  m, or 300 m). Considering the latter values, we have determined that setting a TDMA/TDMA scheme at  $d_{BN} = 100$  m,  $R_{Anet} = 24$  Mbit/s,  $R_{Bnet} = 24$  Mbit/s,  $M = 5$ ,  $k_c = 3$ ,  $n_{client} \leq 10$ , would achieve (among the options considered in our analyses) the highest data throughput rate, yielding  $TH_{Anet,Bnet} = 58.5$  Mbit/s. Hence, the joint traffic management and data networking parameters are set as follows. The selected traffic management related parameters are:  $d_{PL} = 100$  m ( $v = 60$  km/h) and  $n = 5$ . The related data networking parameters, selected to realize the networking configuration which achieves a high vehicular throughput rate, are given as:  $d_{BN} = 100$  m,  $R_{Anet} = 24$  Mbit/s,  $R_{Bnet} = 24$  Mbit/s,  $M = 5$ . This setting yields an aggregate data communications throughput rate that is equal to  $TH_{Anet,Bnet} = 58.5$  Mbit/s. As well demonstrated by these design settings, as we lower the targeted vehicular throughput requirement from 3000 veh/h to 2750 veh/h, we are able to increase the data communications network throughput rate from  $TH_{Anet,Bnet} = 50.9$  Mbit/s to  $TH_{Anet,Bnet} = 58.5$  Mbit/s.

It is interesting to note that as  $d_{PL}$  is increased from 100 m to 125 m, the attained vehicular throughput is increased to 2850 veh/h, while the data communications throughput rate is reduced to 35 Mbit/s. However, if  $d_{PL}$  is further increased to 150 m, the system can sustain a higher vehicular throughput rate, of about 3000 veh/h, while also realizing a higher data throughput rate, which is equal to 42 Mbit/s. The joint performance is thus uniformly better than that attained when setting  $d_{PL} = 125$  m. The occurrence of such a performance spike has been explained when discussed in connection with the performance trends exhibited in discussing the results shown in Figures 12 and 13.

## 9. Conclusions

We develop and study in this paper traffic management and data networking mechanisms for an autonomous transportation system. Access and exit ramps are attached to each segment of the highway. A ramp manager is used to regulate the admission of vehicles into the highway and

their routing to designated lanes. Vehicles moving across a lane are organized into platoons. A platoon leader is elected in each platoon and is used to manage, coordinate and synchronize platoon members. The time spent by a vehicle in the system consists of its queueing delay at the ramp supplemented by, once admitted, its transit time across the highway until it reaches its destination exit ramp. We determine the structural formations of platoons in each lane, including their size and vehicular speeds, jointly with the setting of the parameters for the employed admission control and routing, in aiming to maximize the realized flow rate, which represents the vehicular throughput rate across the highway, under vehicular end-to-end delay constraint. The latter sets a limit on the vehicular on-ramp queueing delay and on the (per unit distance) transit time incurred along the highway. We also account for the need to keep sufficient gaps between platoons to accommodate the transition and merging of vehicles moving across lanes, and entering and exiting the highway. Accordingly, wider gaps are provided across the slower lanes. Faster lanes are employed for supporting the mobility of vehicles that travel longer distances towards their destinations. We show that the vehicular flow rate across the highway can be maximized, under delay prescriptions, by the setting of proper platoon formations, vehicular per-lane speeds and admission regulations.

Using such platoon based formations, we develop a V2V wireless data communications networking protocol that is used to disseminate messages produced by highway vehicles and sent to other vehicles that travel within a specified range from the source vehicle. The hierarchical networking scheme that is presented and studied employs algorithms that are used to dynamically synthesize a mobile Backbone Network (Bnet). The latter consists of interconnected platoon leaders that are elected to serve as Backbone Nodes (BNs). Each BN serves as an access point for its Access Network (Anet) mobile clients. We study the performance behavior of the network system and determine the optimal cross-layer setting of its parameters, assuming both TDMA and IEEE 802.11p oriented wireless channel sharing (MAC) schemes for the Anet and Bnet subsystems. Packet delay limits are also imposed. Integrating our mechanisms and schemes developed for the traffic management and for the data networking planes, we demonstrate the performance tradeoffs available to the system designer and to the transportation manager when aiming to assure a transportation system operation that achieves targeted vehicular flow rates and transit delays while also configuring the data communications network system to meet targeted message throughput and delay objectives. For example, we note that often the use of relatively short distances between elected BNs, jointly with the employment of higher transmission data rates across the Bnet and Anet systems, leads to higher delay-capped data throughput rates. In turn, to achieve a high vehicular flow rate, it is often necessary to structure platoons to have wider spans (inducing longer inter-BN ranges). Consequently, the system manager must

select an operating point that is based on a compromise in terms of the ability to meet individual targeted traffic management and data networking performance metrics.

The approaches and mechanisms developed in this paper can also be applied to the design of autonomous highway transportation systems whereby platoon formations are not imposed. For the hierarchical networking protocol, we note that the election of BNs, when coordinated by a regional manager or by a fully distributed algorithm, can be performed in a manner that does not depend on the existence of platoon formations and their identified and announced leaders. BNs (and consequently Anets and the Bnet) can be elected to meet advantageous access and coverage requirements, following analyses similar to those performed in this paper, when not aided by the existence of platoon formations. Also, BNs can be elected, fully or partially, from a set of stationary Roadside Units (RSUs), or access points, when they are (or become) available. Stationary (such as fiber optic based) RSU backbone networks, are then employed, and/or used to supplement the wireless V2V network over certain highway segments. Similarly, it is noted that our traffic management models, methods and results are also applicable when used for non-platoon based vehicular mobility patterns, including when considering semi-autonomous transportation systems. Yet, the platoon formation model is highly advantageous in serving to effectively offer tight coordination and rapid safety based reactive adaptations, when performed among vehicles that move as a group.

## References

- [1] M. Bergen, Alphabet Launches the First Taxi Service With No Human Drivers (Nov. 2017). URL <https://bloom.bg/2Ea2dm1>
- [2] C. Diakaki, M. Papageorgiou, I. Papamichail, I. Nikolos, Overview and analysis of Vehicle Automation and Communication Systems from a motorway traffic management perspective, *Transportation Research Part A: Policy and Practice* 75 (2015) 147–165. doi:10.1016/j.tra.2015.03.015.
- [3] Y.-Y. Lin, I. Rubin, Vehicular and Messaging Throughput Tradeoffs in Autonomous Highway Systems, in: *IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–6. doi:10.1109/GLOCOM.2015.7417014.
- [4] T. S. Dao, C. M. Clark, J. P. Huissoon, Distributed platoon assignment and lane selection for traffic flow optimization, in: *IEEE Intelligent Vehicles Symposium*, 2008, pp. 739–744. doi:10.1109/IVS.2008.4621202.
- [5] J. Lioris, R. Pedarsani, F. Y. Tascikaraoglu, P. Varaiya, Platoons of connected vehicles can double throughput in urban roads, *Transportation Research Part C: Emerging Technologies* 77 (2017) 292–305. doi:10.1016/j.trc.2017.01.023.
- [6] S. Santini, A. Salvi, A. S. Valente, A. Pescapé, M. Segata, R. L. Cigno, A Consensus-Based Approach for Platooning with Intervehicular Communications and Its Validation in Realistic Scenarios, *IEEE Transactions on Vehicular Technology* 66 (3) (2017) 1985–1999. doi:10.1109/TVT.2016.2585018.
- [7] C. Campolo, A. Molinaro, G. Araniti, A. O. Berthet, Better Platooning Control Toward Autonomous Driving : An LTE Device-to-Device Communications Strategy That Meets Ultralow Latency Requirements, *IEEE Vehicular Technology Magazine* 12 (1) (2017) 30–38. doi:10.1109/MVT.2016.2632418.

- [8] B. Liu, D. Jia, K. Lu, D. Ngoduy, J. Wang, L. Wu, A Joint Control-Communication Design for Reliable Vehicle Platooning in Hybrid Traffic, *IEEE Transactions on Vehicular Technology* 66 (10) (2017) 9394–9409. doi:10.1109/TVT.2017.2702650.
- [9] F. Cuomo, I. Rubin, A. Baiocchi, P. Salvo, Enhanced VANET Broadcast Throughput Capacity via a Dynamic Backbone Architecture, *Ad Hoc Networks* 21 (2014) 42–59. doi:10.1016/j.adhoc.2014.04.008.
- [10] H.-J. Ju, I. Rubin, Performance analysis and enhancement for backbone based wireless mobile ad hoc networks, in: 2nd International Conference on Broadband Networks, Vol. 1, 2005, pp. 733–742. doi:10.1109/ICBN.2005.1589679.
- [11] C.-C. Tan, I. Rubin, H.-J. Ju, Multicasting in mobile backbone based ad hoc wireless networks, in: *IEEE Wireless Communications and Networking Conference (WCNC)*, Vol. 2, 2006, pp. 703–708. doi:10.1109/WCNC.2006.1683554.
- [12] I. Rubin, A. Baiocchi, F. Cuomo, P. Salvo, GPS aided inter-vehicular wireless networking, in: *Information Theory and Applications Workshop (ITA)*, San Diego, CA, USA, 2013, pp. 1–9. doi:10.1109/ITA.2013.6502973.
- [13] I. Rubin, Y.-Y. Lin, A. Baiocchi, F. Cuomo, P. Salvo, Rapid Dissemination of Public Safety Message Flows in Vehicular Networks, *Journal of Communications (JCM)* 9 (8) (2014) 616–626. doi:10.12720/jcm.9.8.616-626.
- [14] I. Turcanu, P. Salvo, A. Baiocchi, F. Cuomo, An integrated VANET-based data dissemination and collection protocol for complex urban scenarios, *Ad Hoc Networks* 52 (2016) 28 – 38, Modeling and Performance Evaluation of Wireless Ad Hoc Networks. doi:10.1016/j.adhoc.2016.07.008.
- [15] F. Yu, S. Biswas, Self-Configuring TDMA Protocols for Enhancing Vehicle Safety With DSRC Based Vehicle-to-Vehicle Communications, *IEEE Journal on Selected Areas in Communications* 25 (8) (2007) 1526–1537. doi:10.1109/JSAC.2007.071004.
- [16] Y.-Y. Lin, I. Rubin, Integrated Message Dissemination and Traffic Regulation for Autonomous VANETs, *IEEE Transactions on Vehicular Technology* 66 (10) (2017) 8644–8658. doi:10.1109/TVT.2017.2700399.
- [17] G. Pathak, H. Li, C. B. Math, S. H. de Groot, Modelling of Communication Reliability for Platooning Applications for Intelligent Transport System, in: 84th IEEE Vehicular Technology Conference (VTC-Fall), 2016, pp. 1–6. doi:10.1109/VTCFall.2016.7881092.
- [18] O. Shagdar, F. Nashashibi, S. Tohme, Performance study of CAM over IEEE 802.11p for cooperative adaptive cruise control, in: *Wireless Days*, 2017, pp. 70–76. doi:10.1109/WD.2017.7918118.
- [19] M. Segata, B. Bloessl, S. Joerer, C. Sommer, M. Gerla, R. L. Cigno, F. Dressler, Toward Communication Strategies for Platooning: Simulative and Experimental Evaluation, *IEEE Transactions on Vehicular Technology* 64 (12) (2015) 5411–5423. doi:10.1109/TVT.2015.2489459.
- [20] L. D. Baskar, B. D. Schutter, H. Hellendoorn, Traffic Management for Automated Highway Systems Using Model-Based Predictive Control, *IEEE Transactions on Intelligent Transportation Systems* 13 (2) (2012) 838–847. doi:10.1109/TITS.2012.2186441.
- [21] J. R. Jackson, Jobshop-Like Queueing Systems, *Management Science* 50 (12) (2004) 1796–1802. doi:10.1287/mnsc.1040.0268.
- [22] J. D. C. Little, A Proof for the Queuing Formula:  $L = \lambda W$ , *Operations Research* 9 (3) (1961) 383–387. doi:10.1287/opre.9.3.383.
- [23] A. Baiocchi, F. Cuomo, M. D. Felice, G. Fusco, Vehicular Ad-Hoc Networks sampling protocols for traffic monitoring and incident detection in Intelligent Transportation Systems, *Transportation Research Part C: Emerging Technologies* 56 (2015) 177–194. doi:10.1016/j.trc.2015.03.041.
- [24] IEEE Standard for Wireless Access in Vehicular Environments (WAVE) – Multi-Channel Operation, *IEEE Std 1609.4-2016 (Revision of IEEE Std 1609.4-2010)* (2016) 1–94doi:10.1109/IEEESTD.2016.7435228.
- [25] C. Campolo, A. Molinaro, A. Vinel, To switch or not to switch: Service discovery and provisioning in multi-radio V2R communications, in: 8th International Workshop on Resilient Networks Design and Modeling (RNDM), 2016, pp. 281–287. doi:10.1109/RNDM.2016.7608299.
- [26] Y. Maalej, A. Abderrahim, M. Guizani, B. Hamdaoui, E. Balti, Advanced Activity-Aware Multi-Channel Operations 1609.4 in VANETs for Vehicular Clouds, in: *IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6. doi:10.1109/GLOCOM.2016.7841571.
- [27] M. van Eenennaam, A. Remke, G. Heijenk, An analytical model for beaconing in VANETs, in: *IEEE Vehicular Networking Conference (VNC)*, 2012, pp. 9–16. doi:10.1109/VNC.2012.6407451.
- [28] M. Hadded, R. Zagrouba, A. Laouti, P. Muhlethaler, L. A. Saidane, An Optimal Strategy for Collision-Free Slots Allocations in Vehicular Ad-hoc Networks, in: *Vehicular Ad-hoc Networks for Smart Cities*, Springer Singapore, Singapore, 2015, pp. 15–30.
- [29] H. A. Omar, W. Zhuang, L. Li, VeMAC: A novel multichannel MAC protocol for vehicular ad hoc networks, in: *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2011, pp. 413–418. doi:10.1109/INFCOMW.2011.5928848.
- [30] ETSI, Intelligent Transport Systems (ITS); STDMA recommended parameters and settings for cooperative ITS; Access Layer Part, Tech. Rep. 102 861 v1.1.1, ETSI (Jan 2012).
- [31] IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, *IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012)* (2016) 1–3534doi:10.1109/IEEESTD.2016.7786995.
- [32] M. Sepulcre, J. Gozalvez, B. Coll-Perales, Why 6 Mbps is Not (Always) the Optimum Data Rate for Beaconing in Vehicular Networks, *IEEE Transactions on Mobile Computing* 16 (12) (2017) 3568–3579. doi:10.1109/TMC.2017.2696533.
- [33] W. G. Cassidy, N. Jaber, S. A. Ruppert, J. Toimoor, K. E. Tepe, E. Abdel-Raheem, Interference modelling and SNR threshold study for use in vehicular safety messaging simulation, in: 26th Biennial Symposium on Communications (QBSC), 2012, pp. 52–55. doi:10.1109/QBSC.2012.6221350.
- [34] Vehicle safety communications - applications (vsc-a) final report, U.S. Department of Transportation (2011) 1–102.
- [35] Q. Xu, T. Mak, J. Ko, R. Sengupta, Vehicle-to-Vehicle Safety Messaging in DSRC, in: *Proceedings of the 1st ACM International Workshop on Vehicular Ad Hoc Networks, VANET '04*, ACM, Philadelphia, PA, USA, 2004, pp. 19–28. doi:10.1145/1023875.1023879.
- [36] S. Gräffing, P. Mähönen, J. Riihijärvi, Performance evaluation of IEEE 1609 WAVE and IEEE 802.11p for vehicular communications, in: *Second International Conference on Ubiquitous and Future Networks (ICUFN)*, 2010, pp. 344–348. doi:10.1109/ICUFN.2010.5547184.