End-to-End Delay Bound Analysis for Location-Based Routing in Hybrid Vehicular Networks

Konstantinos Katsaros, *Member, IEEE*, Mehrdad Dianati, *Senior Member, IEEE*, Rahim Tafazolli, *Senior Member, IEEE*, and Xiaolong Guo

***Abstract*—There is an ongoing debate in the research and industry communities as to whether IEEE 802.11p or Third-Generation Partnership Project (3GPP) Long-Term Evolution (LTE) should be used for vehicular communications. In this paper, we argue that a hybrid vehicular network combining both technologies can increase the performance of the system. We first propose a mechanism to improve location-based routing in a hybrid vehicular network architecture by data and signaling traffic separation on independent wireless networks. We then develop analytical models to calculate the stochastic upper bound of the end-to-end delay (E2ED) for location-based routing in three different networking architecture alternatives based on a) short-range ad hoc only, b) cellular only, and c) the proposed hybrid ad hoc/cellular network. The analytical approach in this paper is based on the stochastic network calculus (SNC) theory, which provides a solid and uniform framework for analysis of the upper bound of the E2ED in communication networks. It is demonstrated that the proposed hybrid network provides a lower E2ED compared with the other two alternatives. Comparisons of realistic simulation results, carried out in NS-3, and analytical results show that the proposed delay bounds provide relatively tight approximations for the E2ED in the three alternative architectures for vehicular networks investigated in this paper.**

***Index Terms*—End-to-end delay (E2ED), hybrid vehicular networks, location services (LSs), stochastic network calculus (SNC).**

I. Introduction

VEHICULAR networks have attracted increasing attention in recent years. Governmental bodies in the U.S. (National Highway Traffic Safety Administration [1]), Europe (European Commission [2]), and Asia (Ministry of Land, Infrastructure, Transport and Tourism of Japan [3]) are in the phase of standardizing and regulating the different technologies that will facilitate intervehicle and vehicle-to-infrastructure communications. Several communication technologies have been proposed and investigated for vehicular network applications. Each of these technologies has certain properties that make them suitable for particular type of applications, mostly dictated by the end-to-end delay (E2ED) requirements, communication range, and the dissemination mode, i.e., broadcast, geocast, or unicast. However, two technologies, namely, dedicated short-range communication (DSRC), e.g., IEEE wireless access in vehicular environments and Intelligent Transport Systems G5 of the European Telecommunications Standards Institute (ETSI), and cellular, e.g., Third-Generation Partnership Project (3GPP) Long-Term Evolution (LTE), are the most promising potential candidates. Analysis of LTE for safety applications suggests that it may struggle to satisfy delay and capacity requirements due to a relatively large latency imposed by the centralized network architecture and the inability to natively support vehicle-to-vehicle (V2V) broadcast [4], [5]. While DSRC-based networks are shown to be more suitable in single-hop broadcast communications, infotainment and cloud-based applications, which require unicast multihop communications, are challenging. Connection to roadside units (RSUs) and from there to the Internet is intermittent with an average duration period of a few seconds depending on vehicle speed. Handing off between RSUs introduces significant delay for address reconfiguration and authentication in the new access point. On the other hand, LTE has an obvious advantage for communication over longer distances due its larger coverage that reduces the required number of hops and handoffs in the network layer. To this end, it seems like a hybrid networking solution that exploits advantages of both technologies can provide a more effective networking solution for vehicular networks. However, designing and analyzing such a hybrid network architecture is a nontrivial issue and is the focus of this paper.

To the best of our knowledge, hybrid networking architectures have been partially considered in the literature [6]–[8]. In these proposals, cellular networks are utilized in conjunction with ad hoc networks and focus on stripping application data on the two networks for content distribution only from a remote host toward a vehicle. Their performance evaluation is mainly limited to simulation-based analysis of achievable throughput without considering the effect on delay. In particular, such systems may suffer from increased jitter in data dissemination due to inequalities in path latencies.

E2ED is a significant key performance indicator for communication networks, and its modeling is not a trivial issue. Different delay models have been developed in the literature to evaluate a communication network in a systematic approach predominantly using three methodologies: a) Markov model,

Manuscript received March 17, 2015; revised July 19, 2015 and September 17, 2015; accepted September 18, 2015. Date of publication September 25, 2015; date of current version September 15, 2016. The review of this paper was coordinated by Prof. G. Mao.

K. Katsaros, M. Dianati, and R. Tafazolli are with the Institute for Communication Systems, University of Surrey, Guildford GU2 7XH, U.K. (e-mail: K.Katsaros@surrey.ac.uk; M.Dianati@surrey.ac.uk; R.Tafazolli@surrey.ac.uk).

X. Guo is with Huawei Technologies Company Ltd., Shenzhen 518129, China (e-mail: guoxiaolong@huawei.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2015.2482362

b) queuing theory (QT), and c) network calculus (NC). There is a plethora of delay models for ad hoc networks; however, as it is discussed in Section III-A, they mainly concentrate on single-hop scenarios where access delay is the dominant factor. E2ED models for cellular networks are even more scarce.

A. *Motivations and Contributions*

The rationale behind the proposed work is driven by the shortcomings of DSRC to efficiently connect with the Internet through RSUs, specifically for signaling traffic required by location services (LSs). In the proposed architecture, signaling traffic for LS is carried by an LTE-based network and application data are served by an ad hoc DSRC system. By splitting data and signaling traffic, the aim is to take advantage of both networks in terms of ubiquitous coverage of cellular networks and low latency (for small number of hops) of DSRC networks to provide better service to the user, as it is demonstrated in this paper. By defusing the congested ad hoc network from the non-time-critical signaling, we provide lower E2ED for the data. In the meantime, keeping data in the same path does not introduce additional jitter from the path disparity. Our work goes beyond the current literature reviewed in Section III-A by considering the E2ED of location-based routing in hybrid network architectures for vehicular communications. For analytical modeling, we use stochastic NC (SNC) [9] to obtain the upper bounds of the E2ED for three network architectures based on a) only a short-range ad hoc network, b) only a cellular network with a large coverage area, and c) a hybrid network comprising an ad hoc and a cellular network. The contributions of this paper can be summarized as follows.

- A mechanism for location-based routing in an ad hoc/cellular hybrid network architecture is proposed to improve the E2ED performance of communication in vehicular networks, based on the separation of data and signaling on different networks.
- A novel approach based on the SNC theory is introduced to obtain the stochastic upper bounds on E2ED for the three aforementioned network architectures.
- The proposed analytical models are validated using realistic simulation scenarios in NS-3 environment.
- In addition, comprehensive performance analysis is carried out to compare the aforementioned architectures in terms of E2ED of data and signaling traffic, as well as throughput.

B. *Overview of Paper*

Following the brief introduction to vehicular communications, Section II presents the hybrid networking architecture with the corresponding system model and assumptions. Section III gives an overview of *state-of-the-art* delay modeling and the related concepts in SNC that are used in this paper. The proposed models for short-range ad hoc, long-range cellular, and hybrid networks based on the SNC methodology are presented in Section IV. In Section V, the proposed analytical models are validated, and the performance of the three aforementioned network architectures is evaluated in different scenarios. Finally, Section VI summarizes the main conclusions of this paper.

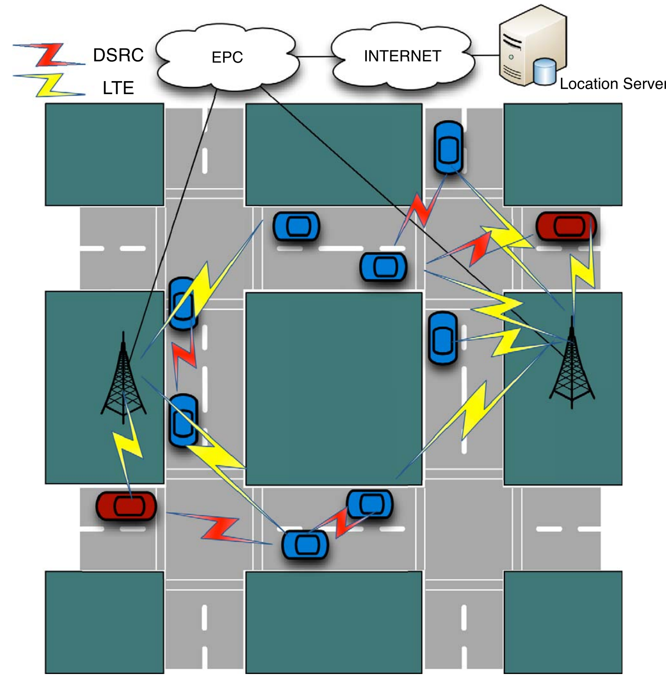


Fig. 1. Network model with hybrid architecture.

II. Hybrid Networking Architecture

Here, we introduce the proposed concept of hybrid networking architecture by first describing the system model and assumptions. Then, we provide an illustrative sequence diagram of the operations performed within the hybrid architecture.

We consider a network that consists of vehicles, ordinary mobile users, base stations (BSs), and a remote host that acts as the location server, as shown in Fig. 1. Vehicles are equipped with two network interfaces, i.e., an IEEE 802.11p-based (11p) network interface for short-range communications and a 3GPP LTE (LTE) network interface to access a cellular network. The 11p network is assumed to be solely used by the vehicles, whereas the LTE network is shared with ordinary mobile users. Following ETSI recommendations, a dual protocol stack is employed with GeoNetworking over 11p and IP over LTE. For short-range ad hoc communications, a location-based routing protocol is considered to handle multihop communications, as proposed by ETSI for unicast communications in vehicular networks [10]. The actual next-hop selection strategy of the routing protocol is out of the scope of this paper; thus, we consider simple greedy location-based forwarding, where the closest-to-destination neighbor is selected as the next hop. All location-based routing protocols require an LS to provide information related to the location of the destination. There are infrastructure-based and infrastructure-less LSs as we describe in [11]. We consider a centralized infrastructure-based LS, exploiting the cellular infrastructure and architecture. Contrary to state-of-art LS implementations that use the same network both for data and LS traffic (in-band signaling), we propose an out-of-band approach where LTE network connectivity is exploited for LS traffic.

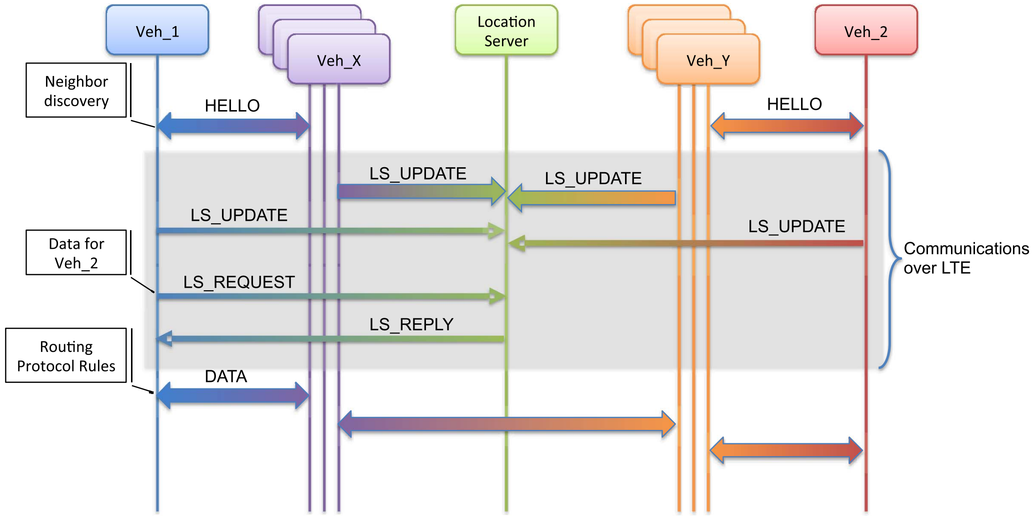


Fig. 2. Message sequence diagram for location-based routing.

A message sequence diagram for location-based routing in vehicular networks is shown in Fig. 2, highlighting those exchanged over the LTE network. Vehicles exchange periodic 1-hop broadcast HELLO messages as a means of neighbor discovery mechanism, whose frequency can be dynamically adjusted to control the signaling overhead [12]. These messages carry position, speed, heading, and other information that depends on the routing protocol design. They are stored for a valid time period calculated as 2× the broadcast interval as specified in similar approaches in HELLO-based protocols, e.g., ad hoc on-demand distance vector [13]. The vehicle should assume that the link to a neighbor is currently lost if it does not receive any message (HELLO or otherwise) for that time. In this simple scenario, Veh_1 has several neighbors represented with Veh_X and Veh_2 has neighboring vehicles represented with Veh_Y, respectively. Between Veh_1 and Veh_2, there is at least one valid path through vehicles X and Y. In addition, LS_UPDATE messages destined to the remote location server are transmitted, which can be triggered by a timer or the distance traveled by a vehicle; in this paper, we consider the timer approach. These messages carry the identifier of the sending node, e.g., IP address, in addition to location-related data, similar to that of a HELLO message. When a vehicle has data to send to another vehicle, it first looks up its own local register for location information on the destination vehicle to start the forwarding process. If the required information is not locally available, a vehicle sends an LS_REQUEST message to the LS server requesting the location information on the destination. This process is also performed at intermediate hops unless the location information is piggybacked to the data packets. These messages are sent to the location server, which replies back with an LS_REPLY message. LS_REQUEST messages contain the identifier of the destination node, and LS_REPLY messages contain the location information that the server holds for that node. The location information is then stored on the local register for a certain valid time period, similar to the valid time of HELLOs. The validity period is related to the interval of LS_UPDATE messages and is calculated as 2× of it. This ensures freshness of the location information on the vehicles.

III. End-to-End Delay Modeling

The performance analysis of different vehicular networking architectures in this paper is primarily based on E2ED bounds. Hence, in Section III-A, we review analytical models for E2ED in wireless communication systems. Then, in Section III-B, we give an overview of the most relevant aspects of the SNC methodology, which will be used later in our analysis.

A. State of the Art in E2ED Modeling

As previously stated in the introduction, there are three main methodologies in the literature to model the delay of a communication system, namely: a) Markov model; b) QT; and c) NC. For short-range communications, most of the existing studies using Markov models are based on extensions of Bianchi's model [14] for saturated data traffic case in IEEE 802.11-based single-hop scenarios. For nonsaturated scenarios, which are more realistic, there are a number of related works in the literature as follows. In [15], Felemban and Ekici have introduced a tight and accurate model for the IEEE 802.11 distributed coordination function (DCF). Using an iterative algorithm to compute the binomial distribution for the contenting nodes, the authors extend the saturated model to unsaturated cases. On the other hand, in [16], Tickoo and Sikdar use QT to model a wireless node by a discrete time $G/G/1$ queue. The model is extended for arbitrary packet size distributions and queue priorities as in the IEEE 802.11e standard. One of the most recently developed models for the IEEE 802.11 DCF has been introduced in [17], which combines the Markov modeling with QT. The backoff transitions are considered a Markov renewal process, and the service is characterized by an $M/G/1$ queue. The Markov renewal process model simplifies the derivation of the closed-form solution for the probability that each station attempts to transmit in a slot. Furthermore, two recent studies have aimed to model IEEE 802.11 using the NC methodology [18], [19]. The first study does not provide an analytical form for the calculations of the upper bound of the service curve, which is numerically evaluated according to a heuristic algorithm. Moreover, both of them are restricted to saturated single-hop scenario deriving their *service curves* from the IEEE 802.11 model of Kumar *et al.* [20]. In [21], Gupta and Shroff presented a model for lower bound delay in multihop scenarios. The authors' aim is to develop a delay-efficient scheduler. They claim that a lower bound technique captures the effect of interference and statistical multiplexing of packets in the system. Finally, Jiao *et al.* [22] use basic probability theory and NC to analyze the delay that a packet experiences at each hop along a path. Then, E2ED is calculated through summing up the per-hop delay along the path. However, it has been shown in [23] that the complexity of a system is proportional to $\mathcal{O}(n^2 \log n)$ when analyzed hop by hop compared with $\mathcal{O}(n \log n)$ when analyzed as one system, as we present in the next section.

Mathematical models of E2ED in cellular networks have not been adequately investigated. A model for 3G cellular technology in [24] analyzes the delays contributed from radio link control and physical layers on IP packets based on stochastic models. A semianalytical Markov model of medium access control (MAC) layer of LTE is presented in [25], where the average delay of packets can be derived but it is limited to uplink traffic. Two studies, i.e., [26] and [27], have used the NC methodology to calculate the delay bounds in the LTE network. The first is restricted to the air interface model of LTE and a specific case of applications related to the Internet of Things. It models the LTE service with a simple Gilbert–Elliot channel without considering delays in the evolved packet core (EPC) and assumes constant traffic from a sensor node to a remote host. On the other hand, in [27], a more generic LTE architecture was presented, which considers a multiple-input–multiple-output air interface and an EPC with multiple routers and strict priority scheduling. Each component is modeled with a stochastic service, and they are combined in a single system. The arrival traffic consists of both real-time and non-real-time flows.

B. SNC Overview

NC is a theory to analyze queuing/flow systems used for modeling communication networks. It originated from the work of Cruz [28], which introduced an alternative to the classical QT for analyzing backlog and delay in communication networks and has split into deterministic NC [29] and SNC [9]. NC employs min +/ max + algebra, which can transform non-linear queuing systems into analytically tractable linear systems. SNC uses more relaxed characterization of distributions, which are defined by violation probabilities of arrival and service processes, thus providing probabilistic bounds for the delay and backlog compared to the exact analysis of queuing theory, which are not tractable for complex real systems.

Definitions and Notation: Consider a service system, as shown in Fig. 3, with input $A(t)$ and output $A^*(t)$ after a variable delay. The following are definitions and notations in the NC framework [9], [29].

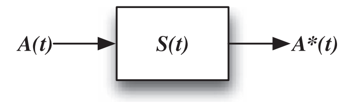


Fig. 3. Basic input–output system.

- The Arrival/Departure Process $A(t)/A^*(t)$ represents the total cumulative number of bits or packets arrived/seen on the input/output flow in the time interval $(0, t]$. In addition, $A(s,t) = A(t) - A(s) \forall\, s < t$.
- Stochastic Arrival Curve (SAC): A flow is constrained by a wide-sense increasing function $\alpha(t)$ if, for all $s \leq t : A(s,t) \leq \alpha(t-s)$, where $\alpha(t)$ is the arrival curve for flow $A(t)$. There are different models to describe SAC, but in this paper, we focus on the virtual-backlog-centric (v.b.c.) model. A flow has a v.b.c. SAC $\alpha(t)$ with bounding function $f(x)$ denoted as $A \sim_{\text{vb}} \langle f, \alpha \rangle$, if $\forall\, t, x \geq 0$

$$P\left\{\sup_{0\leq s\leq t}\{A(s,t) - \alpha(t-s)\} > x\right\} \leq f(x). \tag{1}$$

A list of common arrival processes used in networking that have SAC is presented in Table I.
- Service curve defines the lower bound on the service provided by a server. The system is said to provide to the input a deterministic service curve $\beta(t)$ if

$$A^*(t) \geq (A \otimes \beta)(t)\ \ \forall\, t \geq 0. \tag{2}$$

Here, $\otimes$ denotes the $(\min, +)$ convolution of two functions as follows:

$$(F \otimes G)(t) = \inf_{0\leq\tau\leq t}\{F(\tau) + G(t-\tau)\}\,.$$

A widely used service curve type is the *latency–rate service curve* represented by $\beta(t) = Rt + T$, where R and T are the rate and latency parameters defined by the service process $S(t)$. There are different server models for SNC; however, we only present the *weak stochastic curve* and the *stochastic service curve* (SSC) models that are used later in our analysis. A server $S(t)$ provides a weak SSC $\beta(t)$ with bounding function $g(x)$, denoted by $S \sim_{\text{ws}} \langle g, \beta \rangle$ if, for all $t \geq 0$ and all $x \geq 0$

$$P\{(A \otimes \beta)(t) - A^*(t) > x\} \leq g(x). \tag{3}$$

A server provides an SSC $\beta(t)$ with bounding function $g_t(x)$, denoted by $S \sim_{\text{sc}} \langle g_t, \beta \rangle$ if, for all $t \geq 0$ and all $x \geq 0$

$$P\left\{\sup_{0\leq s\leq t}[A \otimes \beta(s) - A^*(s)] > x\right\} \leq g_t(x). \tag{4}$$

If a server provides to the input a *weak stochastic* service $S \sim_{\text{ws}} \langle g, \beta \rangle$, it provides a *stochastic service* $S \sim_{\text{sc}} \langle g_t^\theta, \beta_{-\theta} \rangle$ with the same service curve $\beta(t)$ and bounding function $g_t^\theta(x)$ equal to

$$g_t^\theta(x) = \left[\frac{1}{\theta}\int_{x-\theta t}^{t} g(y)dy\right]_1 \tag{5}$$

which holds for all $t \geq 0$, $x \geq 0$, and $\theta > 0$, $[z]_1 \equiv \min\{z, 1\}$.

TABLE I
SAC FOR DIFFERENT ARRIVAL TYPES [9]

Type	Arrival Curve $\alpha(t)$	Bounding Function $f(x)$	Comments
Constant Inter-Arrival	$T \cdot L \cdot t$	0	– T packet arrival interval – L packet size
Poisson	$r \cdot t$	$1-(1-a)\sum_{i=0}^{k}\left[\frac{[a(i-k)]^i}{i!}e^{-a(i-k)}\right]$	– $r > \lambda L$ – λ arrival rate – L packet size – $a = \lambda L/r$ – $k = \lceil \frac{x}{L} \rceil$
gSBB [30]	$\rho \cdot t$	me^{-nx}	– ρ upper rate – m, n optimisation parameters

- The *virtual delay* is the delay that would be experienced by a bit or packet arriving at time t if all bits (packets) received before it are served before it and is given by

$$d(t) = \inf\left\{\tau : A(t) \leq A^*(t+\tau)\right\}. \tag{6}$$

There are a number of important theorems in the literature that are often used in SNC. Here, we only introduce the relevant theorems that are used in this paper (the proofs can be found in [9]).

Theorem 1—E2ED Bound: Consider a system with an arrival flow characterized by the arrival curve $\alpha(t)$ with bounding function $f(x)$ and the service has an SSC $\beta(t)$ with bounding function $g(x)$, then the *virtual delay* $d(t)$ satisfies the inequality

$$P\left\{d(t) > h\left(\alpha(t)+x, \beta(t)\right)\right\} \leq (f \otimes g)(x) \tag{7}$$

where $h(a,b)$ is the maximum horizontal distance between functions $a(t)$ and $b(t)$ and is defined as

$$h(a,b) = \sup_{s\geq 0}\left\{\inf\left\{\tau \geq 0 : a(s) \leq b(s+\tau)\right\}\right\}.$$

Theorem 2—Flow Aggregation: Consider N flows with arrival processes $A_i(t)\ \forall\, i = 1,\ldots,N$. Then, the aggregated arrival flow is equal to the sum of all flows

$$A(t) = \sum_{i=1}^{N} A_i(t) \tag{8}$$

and if $\forall\, i\ A_i \sim_{\text{vb}} \langle f_i, \alpha_i\rangle$, then $A \sim_{\text{vb}} \langle f, \alpha\rangle$, where $f(x) = f_1 \otimes f_2 \otimes \cdots \otimes f_N(x)$, and $\alpha(t) = \sum_1^N \alpha_i(t)$.

Theorem 3—Systems in Tandem: If a flow is traversing a sequence of servers $i = 1,\ldots,N$ with constant propagation delay between the servers, each offering an SSC $S \sim_{\text{sc}} \langle g_i, \beta_i\rangle$ with service $\beta_i(t)$ with bounding function $g_i(x)$, the total (network) service curve $\beta(t)$ and bounding function $g(x)$ are given by

$$\beta(t) = (\beta_1 \otimes \beta_2 \otimes \cdots \otimes \beta_N)(t) \tag{9}$$
$$g(x) = (g_1 \otimes g_2 \otimes \cdots \otimes g_N)(x). \tag{10}$$

Theorem 4—Leftover Service: Consider a system with an aggregated arrival of $A(t)$, consisting of two flows $(A_1(t), A_2(t))$ and an SSC $S \sim_{\text{sc}} \langle g, \beta\rangle$. If flow $A_2(t)$ has a v.b.c. SAC, i.e., $A_2 \sim_{\text{vb}} \langle f_2, \alpha_2\rangle$, then the system guarantees to flow $A_1(t)$ an SSC characterized by

$$\beta_1'(t) = \beta(t) - \alpha_{2,\theta}(t) \tag{11}$$
$$g_1'(x) = \left(g \otimes f_{2,t}^{\theta}\right)(x) \tag{12}$$

where $a_{2,\theta}(t) = a(t) + \theta t$, $\theta > 0$, and $f_{2,t}^{\theta} = (1/\theta)\int_{x-\theta t}^{\infty} f_2(y)dy$.

IV. Proposed End-to-End Delay Bounds

Here, we develop upper bound models for the E2ED of location-based traffic, including both data and signaling traffic. The models include three network architectures based on i) only short-range ad hoc wireless communications (e.g., IEEE 802.11p); ii) only long-range cellular communications (e.g., 3GPP LTE); and, finally, iii) the proposed hybrid network, where the short-range ad hoc network is used for data communications and the long-range cellular network is used for signaling.

E2ED is the sum of the delay in different communication layers, in one or multiple hops, depending on the scenario and network architecture. The delay in each hop can be broken down into a number of components. For the short-range communications, based on the IEEE 802.11p technology, the main source of delay for each hop is considered to be the time spent contenting for the shared channel, as well as any queuing delay. In long-range communications, based on the 3GPP LTE technology, the total delay is a combination of delays introduced by the radio access network and the delay in EPC. Processing delays are neglected from our model since they are generally very small in the range of some microseconds [31]. The delay for communication from each RSU or the EPC to the LS server is mainly governed by the *Internet* delay.

We use SNC to analyze the E2ED; thus, we first describe the arrival processes with stochastic curves in Section IV-A. Subsequently, calculations of the service curves and delay bounds according to Theorem 1 for the three aforementioned network architectures are given in Section IV-B–D, respectively.

A. Modeling of Arrival Processes

Data and signaling traffic are characterized by different arrival processes. These processes can be modeled by one of the

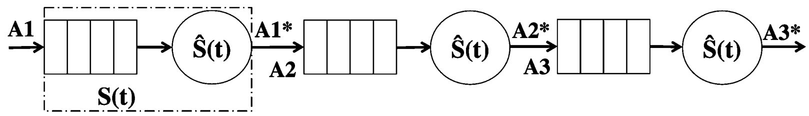


Fig. 4. Abstract scenario for a multihop wireless network.

generic traffic types with the corresponding v.b.c. SAC $A \sim_{\rm vb} \langle f, \alpha \rangle$ from Table I as follows.

- *Application data traffic* depends on the type of application, e.g., Internet access, location advertisement, or other infotainment application. In general, the traffic generated from these applications can be characterized by a generalized stochastically bounded bursty (gSBB) model [30].
- *Neighbor discovery traffic* is generated by periodic broadcast of 1-hop HELLO messages. The interval period can be fixed or dynamically adjusted to control the network overhead. In this paper, we consider fixed interval; thus, the HELLO message traffic can be characterized by a constant interarrival time process. The packet size for HELLOs is 100 bytes.
- *LS traffic* consists of periodic LS_UPDATE messages and asynchronous LS_REQUEST and LS_REPLY messages. The LS_UPDATE messages can be triggered by a timer or by the position changes of a vehicle. We adopt the timer approach; thus, the traffic type is characterized by a constant interarrival time process. We also assume that the LS server answers all requests successfully; hence, LS_REQUEST and LS_REPLY have the same distribution. The LS_REQUEST and LS_REPLY message arrivals are linked to the application traffic and can be generally characterized by a Poisson process. The packet sizes are fixed to 100 bytes for all LS traffic.
- The packet arrival in cellular networks has bursty characteristics with small packet sizes based on the analysis in [32]. Furthermore, as more than 60% of mobile traffic is multimedia content as reported by Cisco [33], we can therefore characterize *background traffic* by a gSBB model with an average rate close to that of video streaming. The packet size is not a relevant parameter for the arrival curve and bounding function of gSBB.

B. E2ED Bound: Ad Hoc Network Scenario

Here, we obtain the upper bound of the E2ED for a pure ad hoc vehicular network architecture based on short-range communication technology. We consider a single-channel network interface similar to IEEE 802.11p for DSRC. Here, each node is modeled by a stochastic process $S(t)$, which comprises a first-in–first-out (FIFO) buffer and the second stochastic process $\hat{S}(t)$ to model the access to the shared channel, as shown in Fig. 4. The SSC for the process $S(t)$ can be obtained from the average service times of the FIFO buffer and access delay, following the methodology in [19]. $\hat{S}(t)$ characterizes the service experienced by a packet that is at the *head-of-line* (HOL) until it is successfully transmitted, otherwise known as the access model. In the IEEE 802.11p access model, access delay is dictated by a multistage binary exponential backoff process, with $R+1$ backoff stages assuming a retry limit of R. We use the model developed in [17] to calculate the mean access delay of a packet at HOL, i.e., $\bar{t}_{\rm serv}$, as follows[1]:

$$\bar{t}_{\rm serv} = \sum_{j=0}^{R} p^j \bar{t}_j \tag{13}$$

where p is the collision probability, and $\bar{t}_j$ is the mean time that a node stays at backoff stage j, which is given by

$$\bar{t}_j = E(b_j) t_B + t_{TX}. \tag{14}$$

$E(b_j)$ represents the number of backoff slots at stage j, t_B is the average length of a backoff slot, and t_{TX} is the average length of a transmission slot. To model the average queuing delay in the FIFO buffer, we resort to basic QT. Following the assumptions of the wireless node for a G/M/1 queue [17] and QT basics (Pollaczek–Khinchin formula), we calculate the mean waiting time in the queue t_q as

$$t_q = \frac{\rho E(\Re)}{1-\rho}. \tag{15}$$

Here, $\rho = \lambda \bar{t}_{\rm serv}$, and λ is the average arrival rate of the packets. $E(\Re)$ is the mean residual processing time, which can be calculated by

$$E(\Re) = \frac{1}{2}\left(c_t^2 + 1\right) \bar{t}_{\rm serv} \tag{16}$$

where c_t^2 denotes the squared coefficient of variation of the processing time.

In terms of SNC, $S(t)$ is described by an SSC $\beta(t)$ bounded by $g_t(x)$, $S \sim_{\rm sc} \langle g_t, \beta \rangle$. Extending the work in [19] to account also for queuing delay and nonsaturated scenarios by using Theorem 3, $\beta(t)$ and $g_t(x)$ are calculated as follows $\forall\, x \geq 0$ if it makes $0 \leq y < 1-q$:

$$\beta(t) = \bar{t}_{\rm serv} \lambda t \otimes t_q \lambda t \tag{17}$$

$$g_t(x) = \left\{ \left(\frac{q}{y}\right)^y \left(\frac{1-q}{1-y}\right)^{1-y} \right\}^K \tag{18}$$

where

$$q = \frac{\bar{t}_{\rm serv} + t_q - t_s}{\mathcal{R} t_c + K \bar{t}_{\rm serv} + \mathcal{B} t_s}, \quad y = \frac{x - K \cdot t_s}{K(\mathcal{R} t_c + K \bar{t}_{\rm serv} + \mathcal{B} t_s)}. \tag{19}$$

Here, K is the queue size, t_s is the average time the channel is found busy due to successful transmission, and t_c is the average time the channel is found busy due to collision. The maximum allowed number of retransmissions is represented by $\mathcal{R}$, and $\mathcal{B}$ is the maximum sum of backoff intervals given by $\sum_{r=0}^{\mathcal{R}} (CW_r - 1)$, where CW_r is the size of the contention window during backoff state r.

[1]Calculations for $p, E(b_j), t_B, t_{TX}$ can be found in [17].

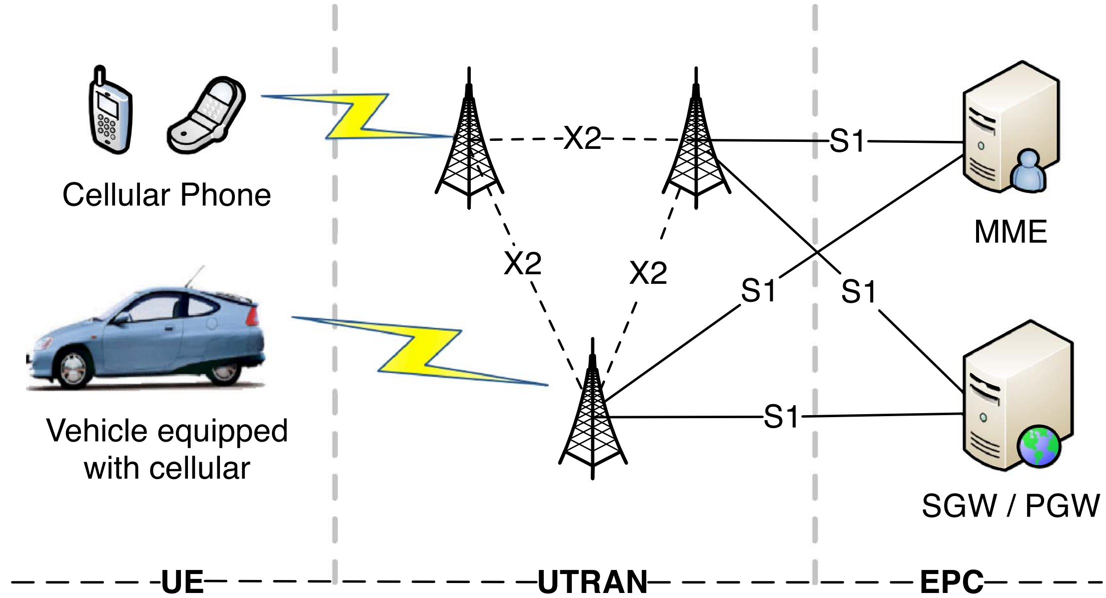


Fig. 5. LTE reference scenario.

The arrival process at each node i represents the aggregate traffic of data and signaling flows, i.e., HELLO messages and LS. Using Theorem 2 for aggregated SAC, $A^i \sim_{\text{vb}} \langle f^i, \alpha^i \rangle$ is calculated as follows:

$$\alpha^i(t) = \alpha^i_D(t) + \alpha^i_{\text{LS}}(t) + \alpha^i_H(t) \tag{20}$$
$$f^i(x) = \left(f^i_D \otimes f^i_{\text{LS}} \otimes f^i_H\right)(x) \tag{21}$$

where (α_D, f_D), $(\alpha_{\text{LS}}, f_{\text{LS}})$, and (α_H, f_H) are the arrival curve and bounding function for data flow, LS flow, and HELLO messages, respectively.

Using Theorem 4 for leftover service, the service that each flow receives on node i can be calculated. For example, the service received by data flow, i.e., $S^i_D \sim_{\text{sc}} \langle g^i_D, \beta^i_D \rangle$, is given by

$$\beta^i_D(t) = \beta^i(t) - \left[\alpha^i_{\text{LS},\theta}(t) + \alpha^i_{H,\theta}(t)\right] \tag{22}$$
$$g^i_D(x) = \left(g^i_t \otimes f^{i,\theta}_{\text{LS},t} \otimes f^{i,\theta}_{H,t}\right)(x) \tag{23}$$

where $\alpha^i_{\text{LS},\theta}(t) = \alpha^i_{\text{LS}}(t) + \theta t$, $\alpha^i_{H,\theta}(t) = \alpha^i_H(t) + \theta t$, $f^{i,\theta}_{\text{LS},t} = (1/\theta)\int_{x-\theta t}^{\infty} f_{\text{LS}}(y)dy$, and $f^{i,\theta}_{H,t} = (1/\theta)\int_{x-\theta t}^{\infty} f_H(y)dy$.

Furthermore, based on Theorem 3 for systems in tandem, the service that a flow will experience after n nodes is $S^{\text{net}}_D \sim_{\text{sc}} \langle g^{\text{net}}_D, \beta^{\text{net}}_D \rangle$, where

$$\beta^{\text{net}}_D(t) = \left(\beta^1_D \otimes \beta^2_D \otimes \cdots \otimes \beta^n_D\right)(t) \tag{24}$$
$$g^{\text{net}}_D(x) = \left(g^1_D \otimes g^2_D \otimes \cdots \otimes g^n_D\right)(x). \tag{25}$$

Finally, the E2ED bound for the data flow is given by Theorem 1 as

$$P\left\{D_D > h\left(\alpha_D(t) + x, \beta^{\text{net}}_D(t)\right)\right\} \le \left(f_D \otimes g^{\text{net}}_D\right)(x). \tag{26}$$

The LS traffic is routed from a vehicle, through an RSU toward the Internet to reach the location server. In a similar way to that of the data traffic in (24) and (25), we calculate the service curve and bounding function for the LS traffic, i.e., $S^{\text{net}}_{\text{LS}} \sim_{\text{sc}} \langle g^{\text{net}}_{\text{LS}}, \beta^{\text{net}}_{\text{LS}} \rangle$, as follows:

$$\beta^{\text{net}}_{\text{LS}}(t) = \left(\beta^1_{\text{LS}} \otimes \beta^{\text{int}}_{\text{LS}}\right)(t) \tag{27}$$
$$g^{\text{net}}_{\text{LS}}(x) = \left(g^1_{\text{LS}} \otimes g^{\text{int}}_{\text{LS}}\right)(x) \tag{28}$$

where $S^1_{\text{LS}} \sim_{\text{sc}} \langle g^1_{\text{LS}}, \beta^1_{\text{LS}} \rangle$ is the SSC of a wireless node for one hop to reach the RSU, and $S^{\text{int}} \sim_{\text{sc}} \langle g^{\text{int}}_{\text{LS}}, \beta^{\text{int}}_{\text{LS}} \rangle$ is the SSC provided by the Internet. Note that the Internet is considered as a set of routers in tandem providing a deterministic ($g^{\text{int}}(x) = 0$) latency–rate service with capacity C, maximum packet size $L^{\max}$, and a strict priority scheduling modeled by the service curve $\beta^{\text{int}}(t) = (L^{\max}/C)t + C$. Thus, the E2ED bound for the LS flow is given by

$$P\left\{D_{\text{LS}} > h\left(\alpha_{\text{LS}}(t)+x, \beta^{\text{net}}_{\text{LS}}(t)\right)\right\} \le \left(f_{\text{LS}} \otimes g^{\text{net}}_{\text{LS}}\right)(x). \tag{29}$$

C. E2ED Bound: Cellular Network Scenario

Here, we examine the E2ED model in a scenario with only a cellular network. We base our analysis on [27] and examine the delay bounds of the LTE network using the reference scenario in Fig. 5.

The arrival process $A(t)$ in this network represents the aggregation of two types of flow: a) background traffic (bg), which is assumed to consume almost 70%–80% of system capacity; and b) the data traffic from the vehicles (veh).[2] The traffic is characterized by a v.b.c. SAC $A \sim_{\text{vb}} \langle f, \alpha \rangle$, which is the aggregation of background and vehicle flows calculated using Theorem 2 as follows:

$$\alpha(t) = \alpha_{bg}(t) + \alpha_{\text{veh}}(t) \tag{30}$$
$$f(x) = (f_{bg} \otimes f_{\text{veh}})(x). \tag{31}$$

[2]The vehicle traffic is the same as in the previous section, i.e., $\alpha_D(t) = \alpha_{\text{veh}}(t)$ and $f_D(x) = f_{\text{veh}}(x)$.

To characterize the service curve for the LTE network, we examine the path that a packet follows for both traffic flows. In the cellular-only network scenario, a node sends the packet through the ingress *eNB* to the EPC and the egress *eNB*, where it is delivered to the destination node. The service curve of this system is the concatenation of three subsystems, namely, uplink, EPC, and downlink. The uplink and downlink are governed by the channel characteristics and the scheduler, whereas the EPC are governed by the underlying network capabilities of the core servers. This can be seen as three systems in tandem; therefore, the total service curve, i.e., $S \sim_{\rm sc} \langle \beta_{\rm net}, g_{\rm net}\rangle$, provided by the LTE network can be calculated using Theorem 3 as follows:

$$\beta_{\rm net}(t) = (\beta_{\rm uplink} \otimes \beta_{\rm EPC} \otimes \beta_{\rm downlink})(t) \tag{32}$$
$$g_{\rm net}(x) = (g_{\rm uplink} \otimes g_{\rm EPC} \otimes g_{\rm downlink})(x). \tag{33}$$

In this paper, we consider a single–single-output air interface for uplink and downlink with round-robin scheduler, where the channel can be modeled as a two-state Markov model: i) ON state where transmission succeeds with probability of 1; and 2) OFF state where a transmitted frame fails with probability of 1. The transition probability matrix is denoted by

$$Q = \begin{bmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{bmatrix}$$

where $q_{ij} \in \{0, 1\}$ denotes the transition probability from state i to state j.

The service curve of the channel is shown to have an SSC $S \sim_{\rm sc} \langle \beta(t), g(x)\rangle$ [34], [35] with

$$\beta(t) = -\frac{1}{\theta_c}\log\frac{\omega(\theta_c)}{2}t$$
$$\omega(\theta_c) = q_{00} + q_{11}e^{-c\theta_c} + \sqrt{(q_{00}+q_{11}e^{-c\theta_c})^2 - 4(q_{00}+q_{11}-1)e^{-c\theta_c}} \tag{34}$$
$$g(x) = e^{-\theta_c x} \tag{35}$$

where $\theta_c > 0$ is the optimization parameter, and c is the number of arrivals at the ON state. Selection of θ_c depends on the constraints for each specific traffic type. In this system model, retransmission until success policy is employed, which means that no packet is dropped because of collision or deep channel fading. Packet losses only happen when the sojourn delay exceeds the delay budget. EPC is considered as a set of routers in tandem with constant rate and a strict priority scheduling, modeled similar to the service curve of the Internet. Since the service is divided among the two traffic flows, the vehicle flow will get a fraction of server capacity according to Theorem 4 and is calculated as follows:

$$\beta_{\rm veh}(t) = \beta_{\rm net}(t) - \alpha_{bg,\theta}(t) \tag{36}$$
$$g_{\rm veh}(x) = \left(g_{\rm net} \otimes f^{\theta}_{bg,t}\right)(x) \tag{37}$$

where $\alpha_{bg,\theta} = \alpha_{bg} + \theta t$, $f^{\theta}_{bg,t} = (1/\theta)\int_{x-\theta t}^{\infty} f_{bg}(y)dy$. Finally, the delay of vehicle traffic in this case is bounded using Theorem 1 and is given by

$$P\{D_{\rm veh} > h\left(\alpha_{\rm veh}(t)+x, \beta_{\rm veh}(t)\right)\} \le (f_{\rm veh} \otimes g_{\rm veh})(x). \tag{38}$$

D. E2ED Bound: Hybrid Network Scenario

In the case of hybrid communications, each vehicle is equipped with two network interfaces, i.e., one for 11p and another for LTE. Data and HELLO messages' traffic is served by the 11p network, whereas LS flows and background traffic are served by the LTE network. According to Theorem 2, the arrival flows have v.b.c. SAC defined for the ad hoc network by

$$\alpha^{\rm ah}(t) = \alpha_D(t) + \alpha_H(t) \tag{39}$$
$$f^{\rm ah}(x) = (f_D \otimes f_H)(x) \tag{40}$$

and for cellular by

$$\alpha^{\rm cell}(t) = \alpha_{bg}(t) + \alpha_{\rm LS}(t) \tag{41}$$
$$f^{\rm cell}(x) = (f_{bg} \otimes f_{\rm LS})(x). \tag{42}$$

The service curve, i.e., $S^{\rm ah} \sim_{\rm sc} \langle g^{\rm ah}, \beta^{\rm ah}\rangle$, for the ad hoc network, following the analysis in Section IV-B, is given by

$$\beta^{\rm ah}(t) = (\beta^1 \otimes \beta^2 \otimes \cdots \otimes \beta^n)(t) \tag{43}$$
$$g^{\rm ah}(x) = (f^1 \otimes f^2 \otimes \cdots \otimes f^n)(x) \tag{44}$$

where n is the number of hops in the path of the flow, and each service curve is calculated based on Theorem 4 for leftover service between the data and HELLO traffic. Thus, the delay bound in this case is given by

$$P\left\{D_{\rm ah} > h\left(\alpha^{\rm ah}(t) + x, \beta^{\rm ah}(t)\right)\right\} \le (f^{\rm ah} \otimes g^{\rm ah})(x). \tag{45}$$

For long-range cellular communications, the service is shared among the background traffic and the LS traffic. The LS requests and updates are forwarded from the vehicles to the EPC. From there, they pass through the Internet toward the LS server; *vice versa* for the replies. Based on Theorem 3, the service provided to the LS flows, i.e., $S^{\rm LS} \sim_{\rm sc} \langle g^{\rm LS}, \beta^{\rm LS}\rangle$, is given by

$$\beta^{\rm LS}(t) = \left(\beta^{\rm cell}_{\rm LS} \otimes \beta^{\rm int}\right)(t) \tag{46}$$
$$g^{\rm LS}(x) = \left(g^{\rm cell}_{\rm LS} \otimes g^{\rm int}\right)(x) \tag{47}$$

where $\langle g^{\rm cell}_{\rm LS}, \beta^{\rm cell}_{\rm LS}\rangle$ is the characteristic of the service provided to the LS flow from the LTE network, and $\langle g^{\rm int}, \beta^{\rm int}\rangle$ is the characteristic of the service provided by the Internet. According to Theorem 4, the service left for the LS traffic in the LTE network is calculated as

$$\beta^{\rm cell}_{\rm LS}(t) = \beta^{\rm cell}(t) - \alpha_{bg,\theta}(t) \tag{48}$$
$$g^{\rm cell}_{\rm LS}(x) = (g^{\rm cell} \otimes f^{\theta}_{bg,t})(x) \tag{49}$$

where $\alpha_{bg,\theta} = \alpha_{bg} + \theta t$, $f^{\theta}_{bg,t} = (1/\theta)\int_{x-\theta t}^{\infty} f_{bg}(y)dy$, and $\langle g^{\rm cell}, \beta^{\rm cell}\rangle$ is calculated for the uplink as

$$\beta^{\rm cell}(t) = (\beta_{\rm uplink} \otimes \beta_{\rm EPC})(t) \tag{50}$$
$$g^{\rm cell}(x) = (g_{\rm uplink} \otimes g_{\rm EPC})(x) \tag{51}$$

and for the downlink as

$$\beta^{\rm cell}(t) = (\beta_{\rm downlink} \otimes \beta_{\rm EPC})(t) \tag{52}$$
$$g^{\rm cell}(x) = (g_{\rm downlink} \otimes g_{\rm EPC})(x). \tag{53}$$

Finally, the delay bound of the LS flow in the hybrid network is given by

$$P\left\{D_{\rm LS} > h\left(\alpha_{\rm LS} + x, \beta^{\rm cell}_{\rm LS}\right)\right\} \le f_{\rm LS} \otimes g^{\rm cell}_{\rm LS}. \tag{54}$$

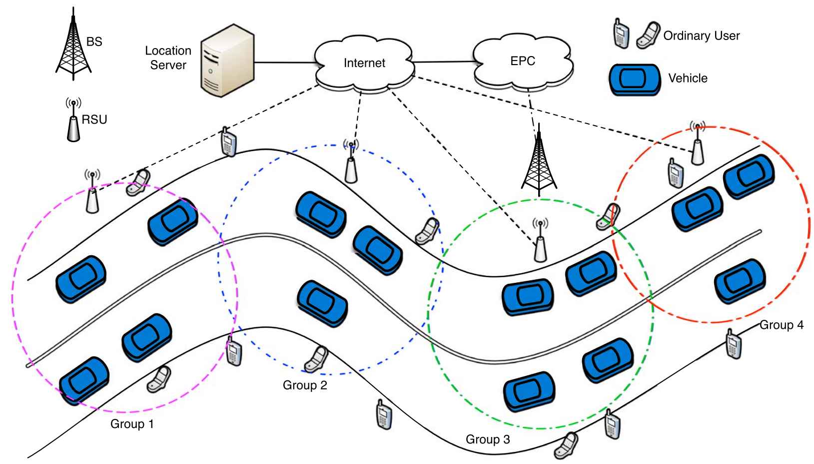


Fig. 6. Reference area for simulations and validation of the model.

V. Model Validation and Performance Evaluation

Here, we numerically evaluate the models for the three networking architectures and validate them through simulations using NS-3 simulator (see Section V-A). In addition, we perform extensive performance evaluation of the aforementioned network architectures in terms of E2ED (upper bounds and statistical characteristics) for data and signaling traffic, and throughput (see Section V-B). For the validation of our model, we have used a highway segment as a reference scenario. This is shown in Fig. 6, where vehicles, ordinary cellular users, RSUs, a BS, and backhaul network are represented. This road segment is 1 km long, and vehicles travel uniformly on two directions with speeds of 50–80 km/h, whereas ordinary users are randomly distributed in the area with fixed positions. Each RSU serves only one group of vehicles (graphically represented with a circle), whereas only one BS serves all vehicles and all ordinary users of the LTE network. Furthermore, for the evaluation, we have used an urban 5×5 Manhattan Grid area, where vehicles travel with speeds of 20–50 km/h. In this scenario, IEEE 802.11p-capable RSUs are located at each intersection and nine BSs cover the area. Vehicles are equipped with IEEE 802.11p and/or LTE communication modules. The distance between each RSU is 300 m, which is also the nominal communication range of the short-range wireless modules of the vehicles. The vehicle application traffic is configured to start from vehicles on the left-hand side of the road segment toward corresponding vehicles in the right-hand side. The configuration parameters for the simulations and the analysis are summarized in Table II.

A. Model Validation

Our first scenario consists of 30 veh/mile, and 10 vehicles from group #1 send data to vehicles from group #4, forming a three-hop communication at the 11p network using 6-Mb/s data rate. Figs. 7 and 8 show the numerical evaluation of the models and simulation results, and as it can be observed, there is a relatively tight approximation of the delay bound. For this scenario, the 11p and LTE networks are closely competing with each other, whereas the proposed hybrid network shows significantly lower bounds. In terms of LS traffic (see Fig. 8), the hybrid network can provide very low bounds compared with the 11p network. The LTE scheduler can provide stricter quality of service as opposed to the enhanced distributed channel access mechanism of IEEE 802.11p. We can observe that the 11p curve has a very long tail, resulting from the blocking of certain packets.

TABLE II
Configuration Parameters

Parameter	Value
Number of vehicles (highway)	30 / 60 veh/mile/direction[1]
Number of vehicles (urban)	100-400 veh.
Number of other users	100 random allocation
Data Packet Size	500Bytes
802.11p Data Rate	6 & 27Mbps
Buffer size (Φ)	100 packets
LTE scheduler / RB alloc.	Round Robin / 25 RBs
Loc. Service Update interval	5sec (time triggered)
Loc. Service Request/Reply	λ=0.1 pkt/sec
HELLO interval	1sec
Vehicle Data Traffic	10/20 V2V connections (10-20kbps/con)
Background Traffic	80 uE-uE connections (200kbps/con)
Internet Capacity/L^{max}	1Gbps / 1500 bytes

[1]Representing Level of Service C and E. The geography of transport systems - Highway speed, Flow and Density [online] http://goo.gl/biXc6d

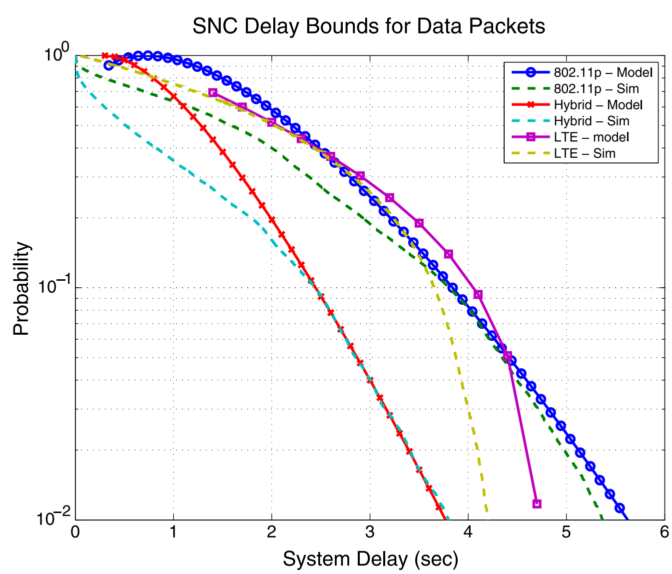


Fig. 7. Comparison of model bound and simulation results for data packets (30 veh/mile and 10 con. scenario).

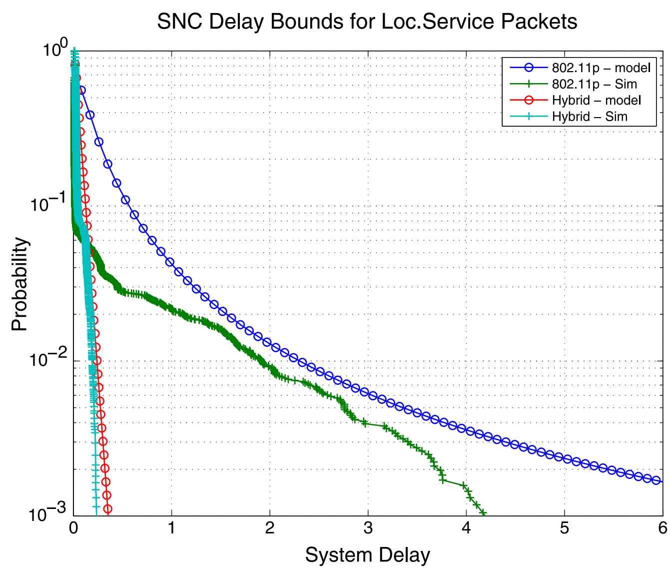


Fig. 8. Comparison of model bound and simulation results for LS packets (30 veh/mile and 10 con. scenario).

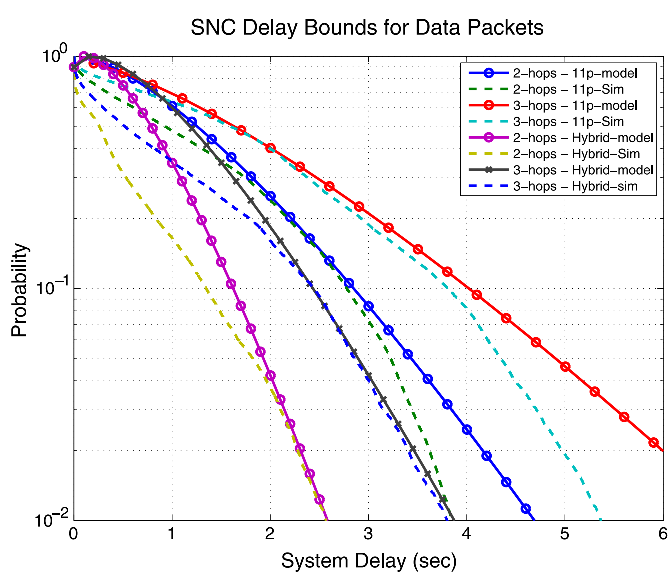


Fig. 9. Comparison of model bound and simulation results for different hop counts (30 veh/mile and 10 con. scenario).

The second scenario evaluates 11p and hybrid networks when different numbers of hops are required to reach the destination. The LTE network is not affected by the number of hops due to the cellular architecture; hence, it is not analyzed in this scenario. For two-hop communications, vehicles from group #1 send data to group #3 and for three-hop communications to group #4. Fig. 9 presents the delay distribution over the 11p and hybrid networks, respectively. It is obvious that more hops result in increased E2ED, but even with only two hops, the proposed hybrid scenario can provide lower bounds compared with the 11p network. This is resulted by the reduced contention levels on the medium due to the separation of signaling and data traffic in the hybrid architecture.

The third scenario evaluates how IEEE 802.11p data rate affects the delay bounds for pure ad hoc and hybrid networks. The increase in available MAC layer data rate from 6 to 27 Mb/s

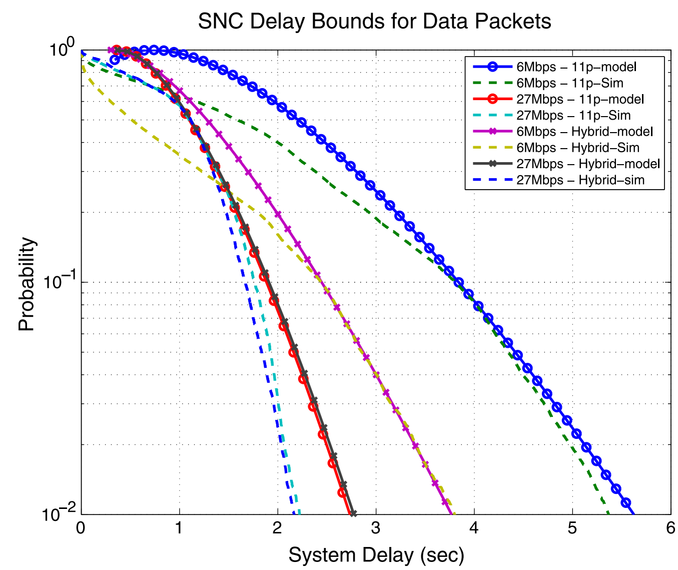


Fig. 10. Comparison of model bound and simulation results for different data rates.

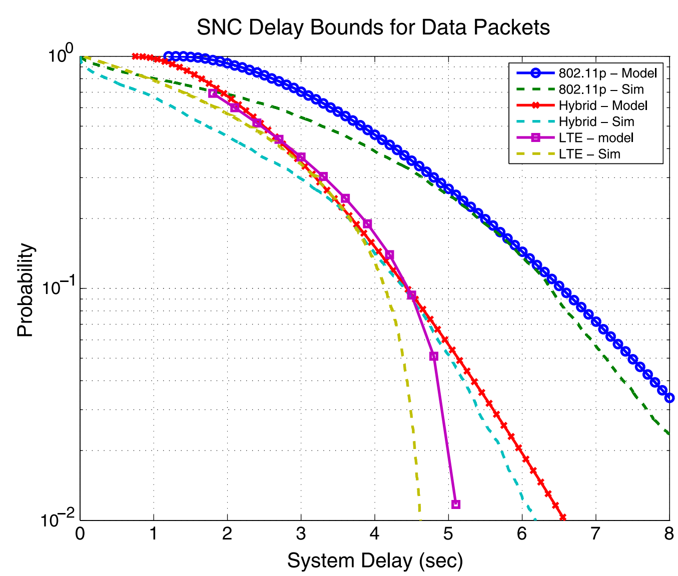


Fig. 11. Comparison of model and simulation results for data packets (60 veh/mile and 20 con. scenario).

has a significant impact in reducing the E2EDs. However, we have to point out that to achieve such high data rates, higher modulation and coding schemes have to be used. These are subject to lower achievable communication range and may result to increasing the number of hops. Nevertheless, for the validation of our model, we configure the scenario such that the number of hops is fixed to three for both data rates. The impact of increased data rate on the E2ED bounds in the 11p and hybrid scenarios is presented in Fig. 10. We observe that, with 27-Mb/s data rate, the two architectures provide relatively similar delay bounds. The available data rate is adequate to carry the data and LS traffic (for the 11p network) without increasing the collision probability dramatically, thus keeping the access delay low. However, in the 6-Mb/s scenario, the bound of the 11p network is significantly higher than the bound of the hybrid network. This reflects the increased contention levels on the shared medium from both data and signaling traffic.

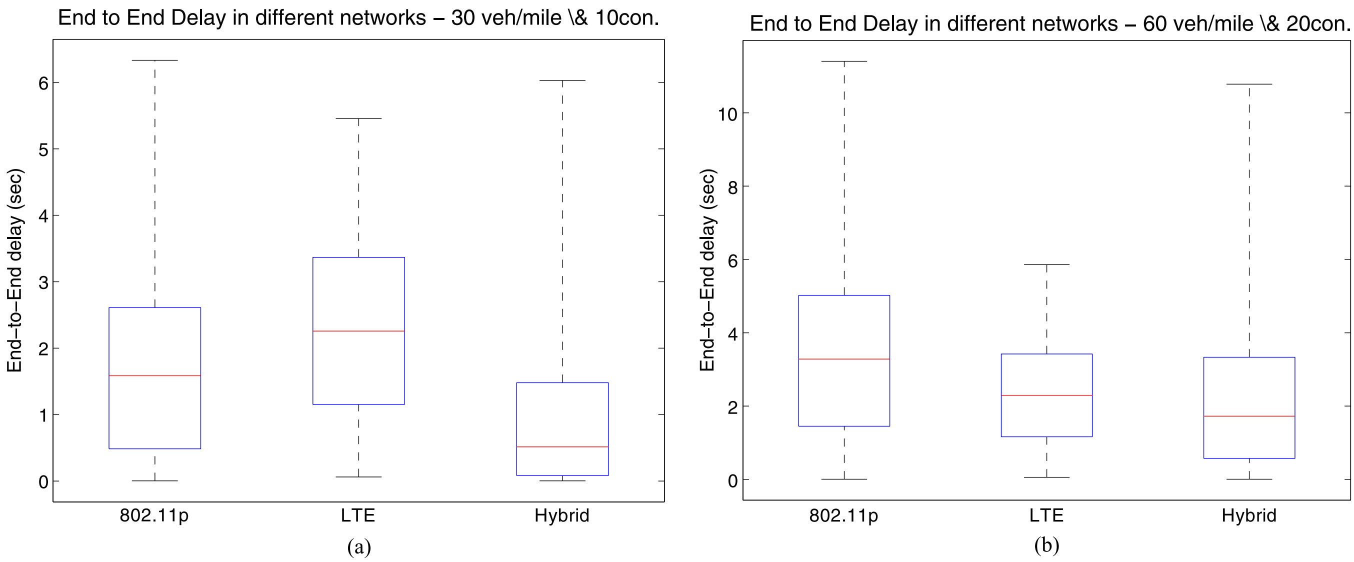


Fig. 12. Comparison of E2ED distribution for data traffic. (a) 30 veh/mile and 10 con. (b) 60 veh/mile and 20 con.

In summary, the validation of the proposed SNC-based models has been successful, providing a relative close fit to the simulation results. In the next section, we provide further performance evaluations of the proposed hybrid architecture in terms of E2ED and throughput.

B. Performance Evaluation

Here, we evaluate the three aforementioned network architectures using both the proposed models and simulations in different scenarios.

We increase the number of vehicles to 60 veh/mile and the number of connections to 20 and evaluate the three-hop scenario, with the results presented in Fig. 11. It is clear that the increase in the number of vehicles and connections affects the contention on the shared channel of IEEE 802.11p and the hybrid network, whereas the pure LTE is less affected because the proportion of contenting nodes does not increase in the same way. In this scenario, the 11p network delay bound is significantly increased, and now, the hybrid network is closely competing with the LTE network. However, the results suggest that the hybrid network still can deliver better E2ED (average and 75th percentile) than the other two network architectures [see Fig. 12(a) and (b)]. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points not considered outliers.

In addition to the data traffic, signaling is also important. We evaluate the average E2ED of LS traffic in IEEE 802.11p and the hybrid networks, accounting both for uplink and downlink flows. In the previous section, we analyzed how the delay bounds for 11p and hybrid with respect to signaling are formed (see Fig. 8) and argued that the hybrid architecture provides lower bounds. Now, we measure the average E2ED in different scenarios. The results presented in Fig. 13 suggest that the 11p network provides lower average delay but with higher standard deviation when the contention levels are low. However, the

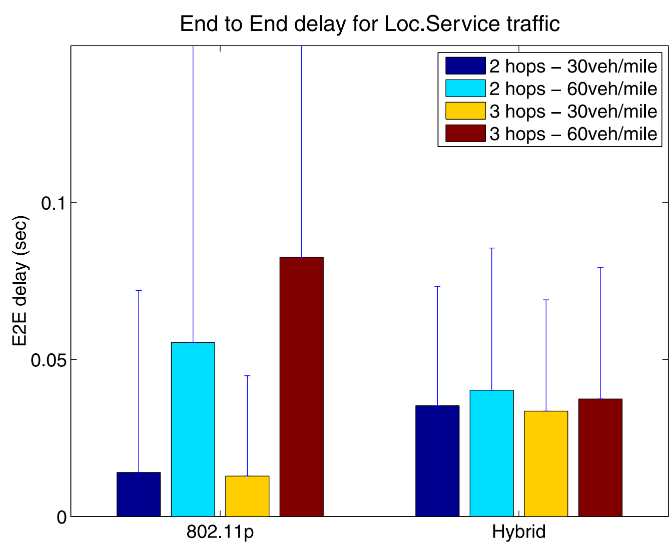


Fig. 13. Comparison of average E2ED for signaling traffic.

hybrid network is affected less by the increase in the number vehicles and can deliver lower E2ED when the number of vehicles increases.

Furthermore, we evaluate the aforementioned network architectures in terms of average normalized throughput on the data flows (see Fig. 14). In scenarios with low contention on the shared channel (30 veh/mile) and when IEEE 802.11p data rate is 27 Mb/s, all three network architectures provide relatively similar results, with 11p and hybrid slightly higher ($< 5\%$) than LTE. Nevertheless, in scenarios with low data rates (6 Mb/s), when the number of contenting nodes increases, both 11p and hybrid show low performance. This degradation of throughput is caused by the increased collision probability in these scenarios.

Finally, we evaluate the aforementioned network architectures in an urban environment in terms of average E2ED and investigate the effect of average vehicle speed, traffic load, and number of vehicles in the reference area. The results presented in Fig. 15(a) suggest that LTE-based networks are not affected

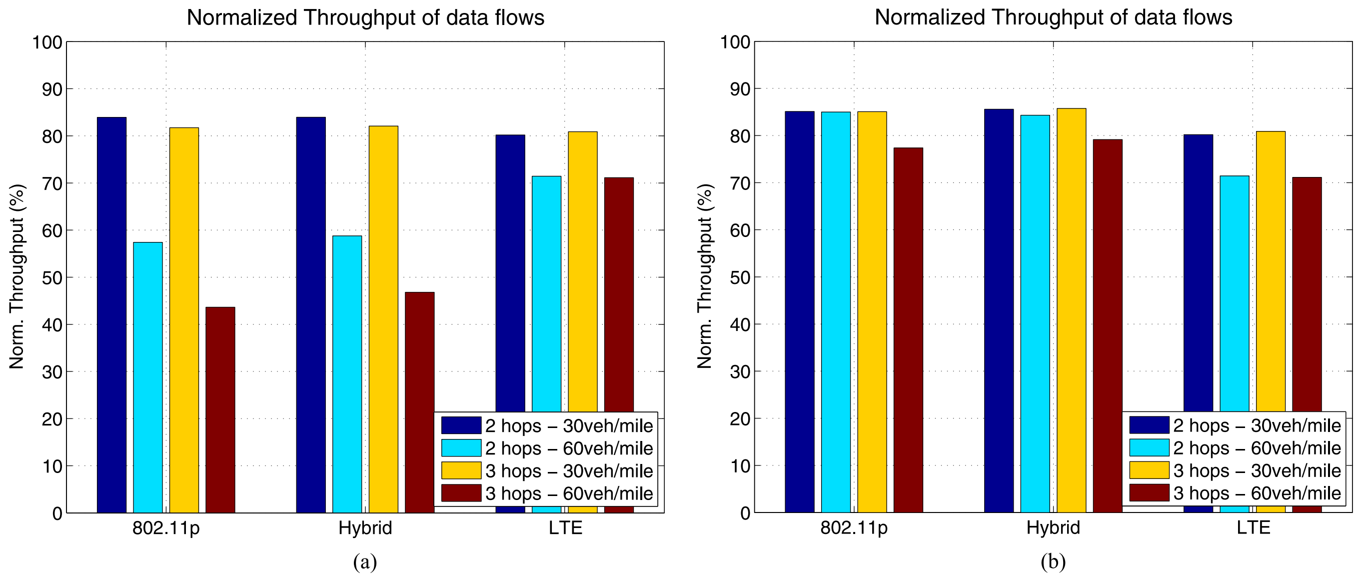


Fig. 14. Comparison of throughput for data rate. (a) IEEE 802.11p @ 6 Mbps. (b) IEEE 802.11p @ 27 Mbps.

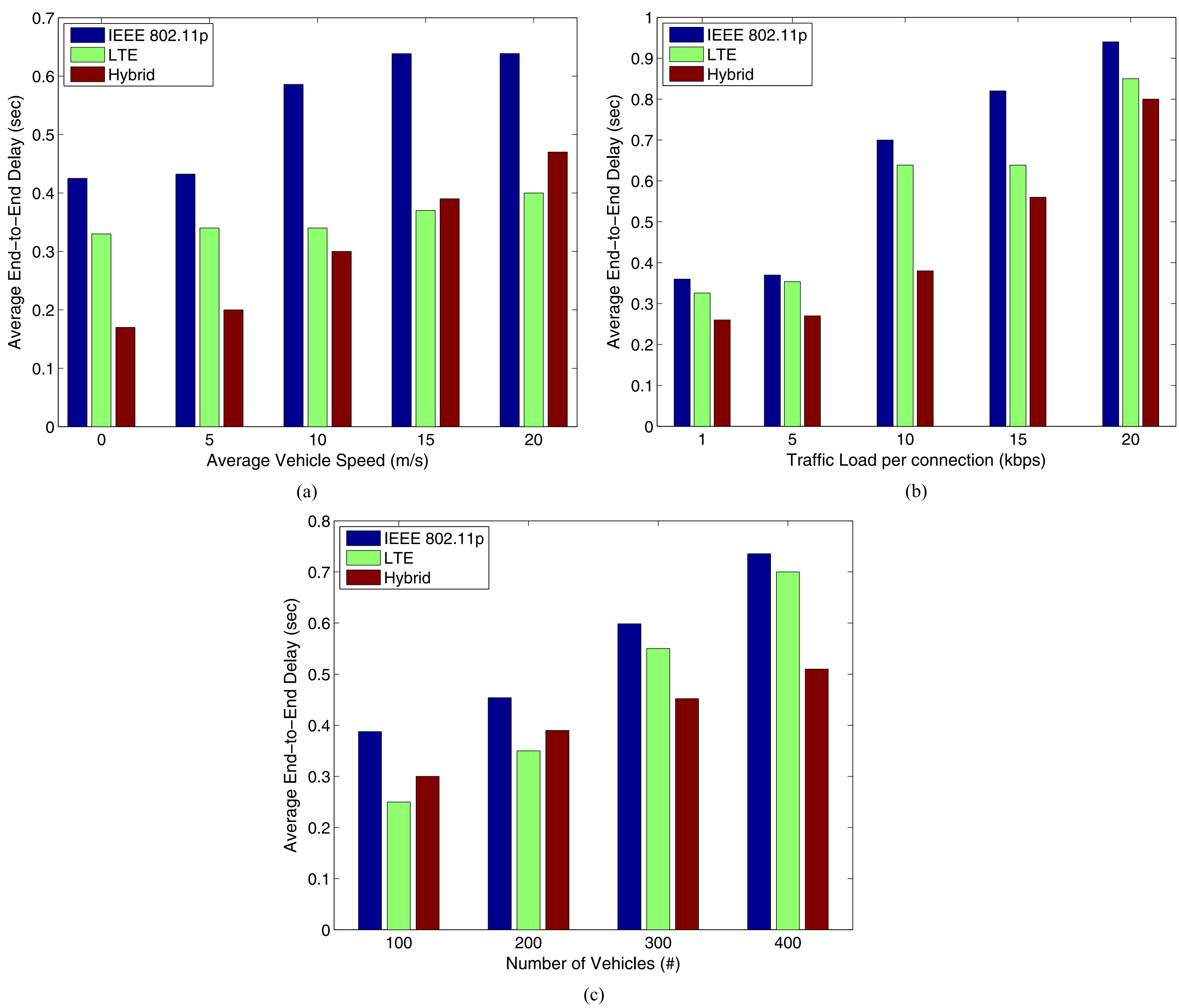


Fig. 15. Comparison of average E2ED in an urban scenario. (a) E2ED versus speed. (b) E2ED versus load. (c) E2ED versus vehicles.

by the average speed of the vehicles as much as the 802.11p-based networks due to the large area of the cells and better mobility support. The IEEE 802.11p-based wireless links are more prone to brakes, and the network graph changes more frequently as relative speed among nodes increases. This results in an increased delay both in pure IEEE 802.11p and hybrid network architectures. However, node density and traffic load influence LTE-based networks similarly [see Fig. 15(b) and (c)]. Notably, hybrid network architecture outperforms both pure IEEE 802.11p and LTE architectures in most cases, only to be defeated by LTE in very fast scenarios and low density where the network graph may become disconnected.

VI. Discussion and Conclusion

A hybrid network architecture with data and signaling traffic separation has been proposed in this paper. An analytical model for the calculation of the E2ED bounds for IEEE 802.11p, 3GPP LTE, and hybrid vehicular networks has been presented in this paper. We use an SNC approach to transform the original problem into a mathematically tractable problem. Using comprehensive simulations, it is demonstrated that the proposed approach provides relatively tight upper bounds for the E2ED for different networks that are considered in this paper.

The results presented in this paper augment the findings in [4], rather than contradicting them. The type of applications investigated here is unicast with medium data rates and multihop ad hoc connections. In such situations, IEEE 802.11p loses the significant benefit it has over LTE for broadcasting single-hop traffic. Our investigation suggests that hybrid networks can significantly help improve the performance of vehicular networks in terms of E2ED both for data traffic and signaling.

The benefits of hybrid networks come in the expense of installing two network interfaces on the vehicles. Apart for the capital expenditure required to equip vehicles with 11p and LTE interfaces, there is a significant cost required for infrastructure. The infrastructure cost related to the 11p is much higher compared with that of LTE, as reflected by the number of access points or BSs. This is another advantage of the proposed hybrid architecture as it utilizes already existing infrastructure and does not require the installation of new access points. In terms of operational expenditure reflected to the cost on the end user, LTE usage is much higher. Nevertheless, the signaling traffic that is carried over LTE will not overhaul the user's monthly bill; as for an average daily car use of 1 h, the amount of data sent by the LS does not exceed 3 MB/month.[3]

Future work will aim at LTE direct support from 3GPP LTE rel. 12, which at the moment is being standardized. The architecture of LTE-D is somehow similar to the proposed hybrid, where signaling is routed through the LTE core and data exchange is performed directly among the devices with the added benefit of a single interface requirement. In addition, multichannel operation on wireless channels, such as the IEEE 1609.4 standard, could be utilized for data and signaling separation on different channels.

[3]The average daily trip in the U.K. is $\sim$1 h [36]. LS_UPDATE packet size is 100 bytes every 5 s plus LS_REQUEST and LS_REPLY.

References

[1] "Vehicle safety and fuel economy rulemaking and research priority plan," Nat. Highway Traffic Safety Admin. (NHTSA), Washington, DC, USA, 2011–2013.

[2] Directive 2010/40/EU of the European Parliament and of the Council of July 7, 2010 on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport.

[3] "White Paper on land, infrastructure, transport and tourism in Japan," Ministry Land, Infrastructure, Transp. Tourism, Tokyo, Japan, Tech. Rep., 2012.

[4] "Intelligent Transport Systems—Framework for Public Mobile Networks in Cooperative ITS (C-ITS)," Eur. Telecommun. Standards Inst., Sophia Antipolis Cedex, France, Tech. Rep. ETSI TR 102 962, 2012.

[5] Z. H. Mir and F. Filali, "LTE and IEEE 802.11p for vehicular networking: A performance evaluation," *EURASIP J. Wireless Commun. Netw.*, vol. 2014, no. 89, pp. 1–5, 2014.

[6] X. Hou, P. Deshpande, and S. Das, "Moving bits from 3G to metro-scale WiFi for vehicular network access: An integrated transport layer solution," in *Proc. 19th IEEE ICNP*, 2011, pp. 353–362.

[7] C. Xu, T. Liu, J. Guan, H. Zhang, and G.-M. Muntean, "CMT-QA: Quality-aware adaptive concurrent multipath data transfer in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 11, pp. 2193–2205, Nov. 2013.

[8] J. Ahn, M. Sathiamoorthy, B. Krishnamachari, F. Bai, and L. Zhang, "Optimizing content dissemination in vehicular networks with radio heterogeneity," *IEEE Trans. Mobile Comput.*, vol. 13, no. 6, pp. 1312–1325, Jun. 2014.

[9] Y. Jiang and Y. Liu, *Stochastic Network Calculus*. New York, NY, USA: Springer-Verlag, 2008.

[10] "Intelligent transport systems—GeoNetworking—Media independent functionality," Eur. Telecommun. Standards Inst., Sophia-Antipolis Cedex, France, Tech. Rep. ETSI TS 102 636-4-1, 2011.

[11] K. Katsaros, M. Dianati, and L. Le, "Effective implementation of location services for VANETs in hybrid network infrastructures," in *Proc. IEEE ICC Workshop Emerging Veh.Netw.: V2V/V2I Railroad Commun.*, 2013, pp. 521–525.

[12] K. Katsaros, M. Dianati, Z. Sun, and R. Tafazolli, "An evaluation of routing in vehicular networks using analytic hierarchy process," *Wireless Commun. Mobile Comput.*, vol. 16, no. 8, pp. 895–911, Jun. 2016.

[13] C. Perkins, E. Belding, and S. Das, "Ad hoc on demand distance vector (AODV) routing," 2003.

[14] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[15] E. Felemban and E. Ekici, "Single hop IEEE 802.11 DCF analysis revisited: Accurate modeling of channel access delay and throughput for saturated and unsaturated traffic cases," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3256–3266, Oct. 2011.

[16] O. Tickoo and B. Sikdar, "Modeling queueing and channel access delay in unsaturated IEEE 802.11 random access MAC based wireless networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 4, pp. 878–891, Aug. 2008.

[17] X. Zhang, "A new method for analyzing nonsaturated IEEE 802.11 DCF networks," *IEEE Wireless Commun. Lett.*, vol. 2, no. 2, pp. 243–246, Apr. 2013.

[18] Y. Wang and T. Wang, "Applying stochastic network calculus to 802.11 backlog and delay analysis," in *Proc. IEEE IWQoS*, 2011, pp. 1–3.

[19] J. Xie and Y. Jiang, "A network calculus approach to delay evaluation of IEEE 802.11 DCF," in *Proc. IEEE Conf. LCN*, 2010, pp. 560–567.

[20] A. Kumar, E. Altman, D. Miorandi, and M. Goyal, "New insights from a fixed-point analysis of single cell IEEE 802.11 WLANs," *IEEE/ACM Trans. Netw.*, vol. 15, no. 3, pp. 588–601, Jun. 2007.

[21] G. Gupta and N. Shroff, "Delay analysis for multi-hop wireless networks," in *Proc. IEEE INFOCOM*, Apr. 2009, pp. 2356–2364.

[22] W. Jiao, M. Sheng, K.-S. Lui, and Y. Shi, "End-to-end delay distribution analysis for stochastic admission control in multi-hop wireless networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1308–1320, Mar. 2014.

[23] F. Ciucu, A. Burchard, and J. Liebeherr, "Scaling properties of statistical end-to-end bounds in the network calculus," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2300–2312, Jun. 2006.

[24] H. Graja, P. Perry, and J. Murphy, "A statistical estimation of average IP packet delay in cellular data networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2005, pp. 1273–1279.

[25] S. Louvros, A. Iossifides, K. Aggelis, A. Baltagiannis, and G. Economou, "A semi-analytical macroscopic MAC layer model for LTE uplink," in *Proc. Int. Conf. NTMS*, 2012, pp. 1–5.
[26] Y. Gao, X. Zhang, and Y. Chang, "Probabilistic delay analysis of Internet of Things based on LTE network," *China Commun.*, vol. 8, no. 8, pp. 80–86, 2011.
[27] L. Zhang, X. Chen, X. Xiang, and J. Wan, "A stochastic network calculus approach for the end-to-end delay analysis of LTE networks," in *Proc. iCOST*, 2011, pp. 30–35.
[28] R. Cruz, "A calculus for network delay. Parts I & II," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 114–141, Jan. 1991.
[29] J.-Y. Le Boudec and P. Thiran, *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet*. New York, NY, USA: Springer-Verlag, 2001.
[30] Q. Yin, Y. Jiang, S. Jiang, and P.-Y. Kong, "Analysis on generalized stochastically bounded bursty traffic for communication networks," in *Proc. IEEE Conf. Local Comput. Netw.*, 2002, pp. 141–149.
[31] R. Ramaswamy, N. Weng, and T. Wolf, "Characterizing network processing delay," in *Proc. IEEE GLOBECOM*, 2004, pp. 1629–1634.
[32] Y. Zhang and A. Arvidsson, "Understanding the characteristics of cellular data traffic," in *Proc. ACM SIGCOMM Workshop Cellular Netw.: Oper., Challenges, Future Design*, 2012, pp. 13–18.
[33] "Cisco visual networking index: Global mobile data traffic forecast update, 2014–2019," Cisco Syst. Inc., Tech. Rep., 2015.
[34] C.-S. Chang, *Performance Guarantees in Communication Networks*. New York, NY, USA: Springer-Verlag, 1999.
[35] Y. Gao, J. Yang, X. Zhang, and Y. Jiang, "Capacity limits for a cognitive radio network under fading channel," in *Proc. IFIP TC 6th Int. Conf. Netw.*, 2011, pp. 42–51.
[36] National Travel Survey statistics. U.K. Dept. Transp., London, U.K., 2002. [Online]. Available: https://www.gov.uk/government/publications/national-travel-survey-2012

Konstantinos Katsaros (S'10–M'15) received the Diploma in electrical and computer engineering from the University of Patras, Patras, Greece, in 2009 and the M.Sc. and Ph.D. degrees from the University of Surrey, Guildford, U.K., in 2010 and 2015, respectively.

He is currently a Research Fellow with the Institute for Communication Systems, University of Surrey. He has been involved in PRE-DRIVE C2X and DRIVE C2X FP7 projects and an industrial project with Huawei Technologies on Safe and Efficient Vehicular Communication. His research interests include, but are not limited, to vehicular communications, intelligent transportation systems, and mobile ad hoc networks.

Mehrdad Dianati (M'05–SM'14) received the B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran; the M.Sc. degree in electrical engineering from Khajeh Nasir Toosi University of Technology, Tehran; and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada.

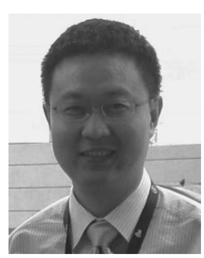
He is currently a Reader (Associate Professor) in communications and networking systems with the Institute for Communication Systems, University of Surrey, Guildford, U.K. He has been involved in a number of international projects in the area of communications and networking as a Technical Coordinator and Work Package Leader. In addition to his research experience, he has nine years of industrial experience as a Software/Hardware Developer and Technical Leader. His research area includes 1) analysis and design of advanced cross-layer optimized physical/medium access control layer techniques for Fifth-Generation wireless access networks and 2) communication network design for intelligent transport systems.

Dr. Dianati has been also actively involved in organizing committees of numerous international conferences in recent years. He is an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *IET Communications*, and Wiley's *Wireless Communications and Mobile*.

Rahim Tafazolli (SM'07) received the B.Sc., M.Sc., and Ph.D. degrees, all in electrical engineering.

He is currently the Director of the Institute for Communication Systems and the 5G Innovation Centre (5GIC), University of Surrey, Guildford, U.K. He has published more than 500 research papers in refereed journals, international conferences, and as an invited speaker. He is the editor of two books *Technologies for the Wireless Future* [Wiley, 2004 (Vol. 1) and 2006 (Vol. 2)].

Mr. Tafazolli is currently the Chairman of the EU Net!Works Technology Platform Expert Group and a Board Member of the U.K. Future Internet Strategy Group. He was appointed Fellow of the Wireless World Research Forum in April 2011 in recognition of his personal contribution to the wireless world and in heading one of Europe's leading research groups.



Xiaolong Guo received the Master's degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2005.

Since 2005, he has been with Huawei Technologies Company Ltd., Shenzhen, China, where he mainly worked on Fourth- and Fifth-Generation cellular systems, the Internet of Things, and connected vehicle topics. He is the author of numerous patents and has published more than 100 contributions on 3GPP.

Mr. Guo is the Chairman of International Mobile Telecommunications 2020 TDCT, the Vice Chairman of the Board of Cooperative Intelligent Transport Systems, and the Vice Chairman of TC5/WG3 of the China Communications Standards Association.