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

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

***Abstract*—There is an ongoing debate in research and industry communities whether IEEE 802.11p or 3GPP LTE should be used for vehicular communications. In this work, 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 networks architecture, by data and signalling traffic separation on independent wireless networks. We then develop analytical models for calculating the stochastic upper bound of the end-to-end delay 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 Stochastic Network Calculus theory, which provides a solid and uniform framework for analysis of the upper bound of the end-to-end delay in communication networks. It is demonstrated that the proposed hybrid network provides a lower end-to-end delay compared to the other two alternatives. Comparisons of realistic simulation results, carried out in NS3, and analytical results show that the proposed delay bounds provide relatively tight approximations for the end-to-end delay in the three alternative architectures for vehicular networks investigated in this paper.**

***Index Terms*—Hybrid Vehicular Networks, End-to-End Delay, Stochastic Network Calculus, Location Services**

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 standardising and regulating the different technologies that will facilitate inter-vehicle and vehicle-to-infrastructure communications. Several communication technologies have been proposed and investigated for vehicular network applications. Each of these technologies have certain properties that make them suitable for particular type of applications, mostly dictated by the end-to-end delay 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 WAVE, ETSI ITS G5, and cellular, e.g., 3GPP 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,that require unicast multi-hop communications, are challenging. Connection to road-side units (RSUs) and from there to the internet is intermittent with an average duration of a few seconds depending on vehicle speed. Handing-off between RSUs introduces significant delay for address re-configuration 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 hand-offs in the network layer. To this end, it seems like a hybrid networking solution that exploits advantages of both technologies can provide more effective networking solution for vehicular networks. However, designing and analysing such a hybrid network architecture is a non-trivial issue and is the focus of this work.

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

K.Katsaros, M.Dianati and R.Tafazolli are with the Institute for Communication Research (ICS) at the University of Surrey, UK. Email: {K.Katsaros, M.Dianati, R.Tafazolli}@surrey.ac.uk.

Xiaolong Guo is with Huawei Technologies Co. Ltd. Email: guoxiaolong@huawei.com

on stripping application data on the two networks for content distribution only from a remote host towards a vehicle. Their performance evaluation is mainly limited to simulation based analysis of achievable throughput without considering the effect on delay. Particularly, such systems may suffer from increased jitter in data dissemination due to inequalities in path latencies.

End-to-end delay is a significant key performance indicator for communication networks, and its modelling is not a trivial issue. Different delay models have been developed in the literature in order to evaluate a communication network in a systematic approach predominantly using three methodologies: (a) Markov model, (b) Queueing 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. End-to-end delay 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 signalling traffic required by location services. In the proposed architecture signalling traffic for location service is carried by a LTE-based network and application data is served by an ad-hoc DSRC system. By splitting data and signalling 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 signalling, we provide lower end-to-end delay 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 end-to-end delay of location-based routing in hybrid network architectures for vehicular communications. For analytical modelling, we use *Stochastic Network Calculus (SNC)* [9] to obtain the upper bounds of the end-to-end delay 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 summarised as follows:

- A mechanism for location-based routing in an ad-hoc/cellular hybrid network architecture is proposed to improve the end-to-end delay performance of communication in vehicular networks, based on the separation of data and signalling on different networks.
- A novel approach based on *SNC* theory is introduced to obtain the stochastic upper bounds on end-to-end delay for the three aforementioned network architectures.
- The proposed analytical models are validated using realistic simulation scenarios in NS3 environment.
- In addition, comprehensive performance analysis is carried out to compare the aforementioned architectures in terms of end-to-end delay of data and signalling 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 modelling 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 summarises the main conclusions of this paper.

II. Hybrid Networking Architecture

In this section, we introduce the proposed concept of hybrid networking architecture by firstly 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 location server as depicted in Fig. 1. Vehicles are equipped with two network interfaces, 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, while 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 multi-hop communications, as proposed by ETSI for unicast communications in vehicular networks [10]. The actual next-hop selection strategy of the routing protocol

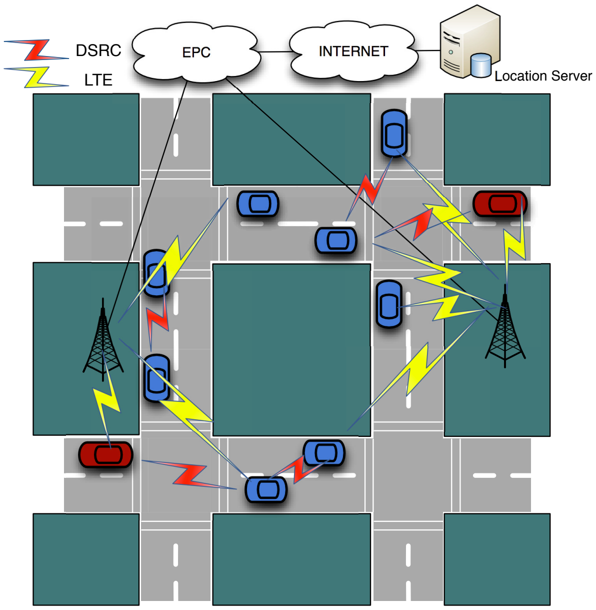


Fig. 1. Network model with hybrid architecture

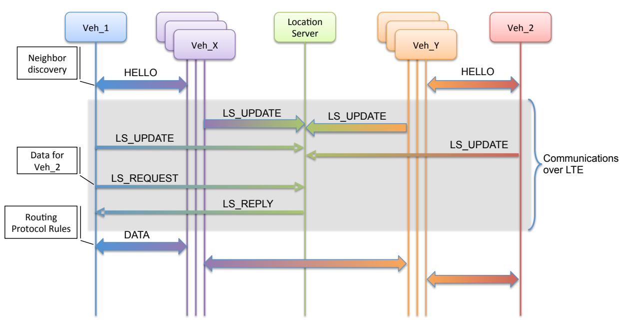


Fig. 2. Message sequence diagram for location-based routing

is out of the scope of this work, thus we consider a simple greedy location-based forwarding, where the closest-to-destination neighbour is selected as next-hop. All location-based routing protocols require a location service (*LS*) to provide information related to the location of the destination. There are infrastructure-based and infrastructure-less location services as we describe in [11]. We consider a centralised 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 signalling), we propose an out-of-band approach where LTE network connectivity is exploited for *LS* traffic.

A message sequence diagram for location-based routing in vehicular networks is depicted in Fig. 2, highlighting those exchanged over the LTE network. Vehicles exchange periodic 1-hop broadcast HELLO messages as a means of neighbour discovery mechanism, whose frequency can be dynamically adjusted in order to control the signalling 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 2x the broadcast interval as specified in similar approaches in HELLO-based protocols, e.g. AODV [13]. The vehicle should assume that the link to a neighbour is currently lost if it does not receive any message (HELLO or otherwise) for that time. In this simple scenario, *Veh_1* has several neighbours represented with *Veh_X* and *Veh_2* has neighbouring 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 travelled by a vehicle; in this work, 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 of the destination vehicle, to start the forwarding process. If the required information is not locally available, a vehicle sends a LS_REQUEST message to the *LS* server requesting the location information of 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 a LS_REPLY message. LS_REQUEST messages contain the identifier of the destination node and LS_REPLY 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 2x times of it. This ensures the freshness of the location information on the vehicles.

III. End-to-End Delay Modelling

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

A. State-of-the-Art in E2ED Modelling

As stated earlier in the introduction, there are three main methodologies in the literature in order to model the delay of a communication system, namely, (a) Markov model, (b) Queueing Theory (*QT*), and (c) Network Calculus (*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 non-saturated scenarios, which are more realistic, there are a number of related works in the literature as follows. Felemban and Ekici in [15] have introduced a tight and accurate model for IEEE 802.11 DCF. Using an iterative algorithm to compute the binomial distribution for the contenting nodes, the authors extend the saturated model to un-saturated cases. On the other hand, Tickoo and Sikdar in [16] 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 IEEE 802.11e standard. One of the most recently developed models for IEEE 802.11 DCF is introduced in [17], which combines the Markov modelling with *QT*. The backoff transitions are considered a Markov renewal process and the service is characterised by a $M/G/1$ 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 aim to model IEEE 802.11 using *NC* methodology [18], [19]. The first work does not provide an analytical form for the calculations of the upper bound of the service curve, which is evaluated numerically according to a heuristic algorithm. Moreover, both of them are restricted to saturated single hop scenario deriving their *service curves* from Kumar et al. IEEE 802.11 model [20]. Gupta and Shroff in [21] presented a model for lower bound delay in multi-hop scenarios. The authors' aim is to develop a delay-efficient scheduler. They claim that 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 a packet experiences at each hop along a path. Then, end-to-end delay 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 analysed hop-by-hop, compared to $\mathcal{O}(n \log n)$ when analysed as one system as we present in next section.

Mathematical models of end-to-end delay in cellular networks have not been adequately investigated. A model for 3G cellular technology in [24] analyses the delays contributed from RLC and PHY layers on IP packets based on stochastic models. A semi-analytical Markov model of 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 works, [26] and [27], have used *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, [27] presents a more generic LTE architecture, which considers a MIMO air interface as well as an *EPC* with multiple routers and strict priority scheduling. Each component is modelled 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.

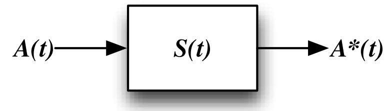


Fig. 3. Basic Input-Output System

B. *Stochastic Network Calculus Overview*

Network Calculus (*NC*) is a theory to analyse queueing/flow systems used for modelling communication networks. It originated from the work of Cruz [28], which introduced an alternative to the classical queueing theory for analysing backlog and delay in communication networks and has split into Deterministic Network Calculus [29] and Stochastic Network Calculus (*SNC*) [9]. *SNC* uses more relaxed characterisation 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 queueing theory, which are not tractable for complex real systems. *NC* employes min+/max+ algebra, which can transform non-linear queueing systems into analytically tractable linear 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. There are the following definitions and notations in the Network Calculus framework [9], [29]:

- 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, $A(s,t) \equiv A(t) - A(s), \forall s < t$.
- *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 $A \sim_{vb} \langle f, \alpha \rangle$, 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 have *SAC* 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 stochastic service curve $\beta(t)$ with bounding function $g(x)$, denoted by $S \sim_{ws} \langle g, \beta \rangle$, 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 a *stochastic service curve* (SSC) $\beta(t)$ with bounding function $g_t(x)$, denoted by $S \sim_{sc} \langle g_t, \beta \rangle$, if for all $t \geq 0$ and all $x \geq 0$

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

If a server provides to the input a *weak stochastic* service $S \ \sim_{ws} \ \langle g, \beta \rangle$, it provides a *stochastic service* $S \ \sim_{sc} \ \left\langle g_t^{\theta}, \beta_{-\theta} \right\rangle$ 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 $\theta > 0$, $[z]_1 \equiv \min\{z, 1\}$.

- 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

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

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 (**End-to-End Delay Bound**). *Consider a system with an arrival flow characterised by the arrival curve $\alpha(t)$ with bounding function $f(x)$, and the service has a stochastic service curve $\beta(t)$ with bounding function $g(x)$, then the* virtual delay $d(t)$ *satisfies the inequality*

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

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

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

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

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

and if $\forall i$ $A_i \sim_{vb} \langle f_i, \alpha_i \rangle$, then $A \sim_{vb} \langle f, \alpha \rangle$ 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 a stochastic service curve $S \sim_{sc} \langle g_i, \beta_i \rangle$ 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 a stochastic service curve $S \sim_{sc} \langle g, \beta \rangle$. If flow $A_2(t)$ has a v.b.c. SAC, $A_2 \sim_{vb} \langle f_2, \alpha_2 \rangle$, then the system guarantees to flow $A_1(t)$ a stochastic service curve characterised by*

$$\beta_1'(t) = \beta(t) - \alpha_{2,\theta}(t), \tag{11}$$

$$g_1'(x) = (g \otimes f_{2,t}^{\theta})(x). \tag{12}$$

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

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

IV. Proposed End-to-End Delay Bounds

In this section, we develop upper bound models for the end-to-end delay of location-based traffic including both data and signalling 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 long range cellular network is used for signalling.

End-to-end delay 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 queueing 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 as well as the delay in Evolved Packet Core (*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 analyse the end-to-end delay; thus, we first describe the arrival processes with stochastic curves in subsection IV-A. Subsequently, calculations of the service curves and delay bounds according to *Theorem 1* for the three aforementioned network architectures are given in subsections IV-B, IV-C, and IV-D, respectively.

A. Modelling of Arrival Processes

Data and signalling traffic are characterised by different arrival processes. These processes can be modelled by one of the generic traffic types with the corresponding $v.b.c.$ *SAC* $A \sim_{vb} \langle f, \alpha \rangle$ 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 characterised by a generalised 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 in order to control the network overhead. In this paper, we consider fixed interval; thus, the HELLO message traffic can be characterised by a constant inter-arrival time process. The packet size for HELLOs is 100Bytes.
- *Location Service 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 characterised by a constant inter-arrival 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 characterised by a Poisson process. The packet sizes are fixed to 100Bytes for all *LS* traffic.
- The packet arrival in cellular networks has bursty

characteristics with small packet sizes based on the analysis in [32]. Further, as more than 60% of mobile traffic is multimedia content as reported by Cisco [33], we can therefore characterise *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. End-to-End Delay Bound: Ad-Hoc Network Scenario

In this subsection, we obtain the upper bound of the end-to-end delay 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 modelled by a stochastic process $S(t)$, which comprises a FIFO buffer and the second stochastic process $\hat{S}(t)$ to model the access to the shared channel as shown in Fig. 4. The *stochastic service curve* 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]. $\hat{S}(t)$ characterises the service experienced by a packet that is at *head-of-line* (HOL) until it is successfully transmitted, otherwise known as access model. In IEEE 802.11p access model, access delay is dictated by a multi-stage 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, $\bar{t}_{serv}$, as follows[1].

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

where p is the collision probability and $\bar{t}_j$ is the mean time 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 queueing 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 queueing theory 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 $\rho = \lambda \bar{t}_{serv}$ and λ is the average arrival rate of the packets. $E(\Re)$ is the mean residual processing time, which can be calculated by:

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

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

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

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

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

where

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

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.

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

$$\alpha^i(t) = \alpha^i_D(t) + \alpha^i_{LS}(t) + \alpha^i_H(t), \tag{20}$$

$$f^i(x) = (f^i_D \otimes f^i_{LS} \otimes f^i_H)(x), \tag{21}$$

where (α_D, f_D) , (α_{LS}, f_{LS}), and (α_H, f_H) are the arrival curve and bounding function for data flow, location service 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, $S^i_D \sim_{sc} \langle g^i_D, \beta^i_D \rangle$, is given by

$$\beta^i_D(t) = \beta^i(t) - [\alpha^i_{LS,\theta}(t) + \alpha^i_{H,\theta}(t)], \tag{22}$$

$$g^i_D(x) = (g^i_t \otimes f^{i,\theta}_{LS,t} \otimes f^{i,\theta}_{H,t})(x), \tag{23}$$

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

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

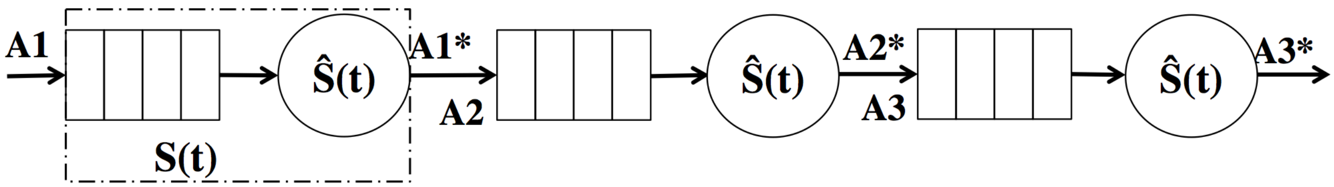


Fig. 4. Abstract Scenario for multi-hop wireless network

Further, based on the *Theorem 3* for systems in tandem, the service that a flow will experience after n nodes is $S_D^{net} \sim_{sc} \langle g_D^{net}, \beta_D^{net} \rangle$, where

$$\beta_D^{net}(t) = (\beta_D^1 \otimes \beta_D^2 \otimes \cdots \otimes \beta_D^n)(t), \tag{24}$$

$$g_D^{net}(x) = (g_D^1 \otimes g_D^2 \otimes \cdots \otimes g_D^n)(x). \tag{25}$$

Finally, the end-to-end delay bound for the data flow is given by *Theorem 1* as

$$P\{D_D > h(\alpha_D(t) + x, \beta_D^{net}(t))\} \leq (f_D \otimes g_D^{net})(x). \tag{26}$$

The Location Service traffic is routed from a vehicle, through a *RSU* towards the internet in order 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, $S_{LS}^{net} \sim_{sc} \langle g_{LS}^{net}, \beta_{LS}^{net} \rangle$, as follows.

$$\beta_{LS}^{net}(t) = (\beta_{LS}^1 \otimes \beta_{LS}^{int})(t), \tag{27}$$

$$g_{LS}^{net}(x) = (g_{LS}^1 \otimes g_{LS}^{int})(x), \tag{28}$$

where $S_{LS}^1 \sim_{sc} \langle g_{LS}^1, \beta_{LS}^1 \rangle$ is the stochastic service curve of wireless node for one hop to reach the RSU, and $S^{int} \sim_{sc} \langle g_{LS}^{int}, \beta_{LS}^{int} \rangle$ is the stochastic service curve provided by the internet. Note that the internet is considered as a set of routers in tandem providing a deterministic ($g^{int}(x) = 0$) latency-rate service with capacity C, maximum packet size L^{max} and a strict priority scheduling modelled by the service curve $\beta^{int}(t) = \frac{L^{max}}{C} t + C$. Thus, the end-to-end delay bound for the Location Service flow is given by

$$P\{D_{LS} > h(\alpha_{LS}(t) + x, \beta_{LS}^{net}(t))\} \leq (f_{LS} \otimes g_{LS}^{net})(x). \tag{29}$$

C. End-to-End Delay Bound: Cellular Network Scenario

In this subsection, we examine the end-to-end delay 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 of Fig. 5.

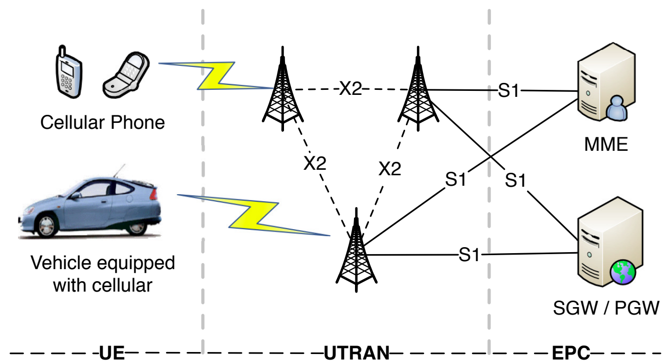


Fig. 5. LTE reference scenario

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 characterised a $v.b.c.$ *SAC* $A \sim_{vb} \langle f, \alpha \rangle$ which is the aggregation of background and vehicle flows calculated using *Theorem 2* as follows.

$$\alpha(t) = \alpha_{bg}(t) + \alpha_{veh}(t), \tag{30}$$

$$f(x) = (f_{bg} \otimes f_{veh})(x). \tag{31}$$

To characterise 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: uplink, *EPC* and downlink. The uplink and downlink are governed by the channel characteristics and the scheduler, while the *EPC* by the underlying network capabilities of the core servers. This can be seen as three systems in tandem, therefore the total service curve, $S \sim_{sc} \langle \beta_{net}, g_{net} \rangle$, provided by the LTE network

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

can be calculated using *Theorem 3* as follows.

$$\beta_{net}(t) = (\beta_{uplink} \otimes \beta_{EPC} \otimes \beta_{downlink})(t), \tag{32}$$

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

In this paper, we consider a SISO air interface for uplink and downlink with Round Robin scheduler, where the channel can be modelled as a two state Markov model; (i) ON state where transmission succeeds with probability of 1; (ii) OFF 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 a stochastic service curve $S \sim_{sc} \langle \beta(t), g(x) \rangle$ [34], [35] with

$$\begin{gathered} \beta(t) = -\tfrac{1}{\theta_c} \log \tfrac{\omega(\theta_c)}{2} t, \\ \omega(\theta_c) = q_{00} + q_{11} e^{-c\theta_c} \\ + \sqrt{(q_{00} + +q_{11} e^{-c\theta_c})^2 - 4(q_{00} + q_{11} - 1)e^{-c\theta_c}}, \end{gathered} \tag{34}$$

$$g(x) = e^{-\theta_c x}, \tag{35}$$

where $\theta_c > 0$ is optimisation parameter, c is the number of arrivals at the state ON. Selection of θ_c depends on the constraints for each specific traffic type. In this system model, a re-transmission until success policy is employed, which means 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, modelled similar to the service curve of 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.

$$\beta_{veh}(t) = \beta_{net}(t) - \alpha_{bg,\theta}(t), \tag{36}$$

$$g_{veh}(x) = (g_{net} \otimes f^{\theta}_{bg,t})(x). \tag{37}$$

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

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

D. End-to-End Delay Bound: Hybrid Network Scenario

In the case of hybrid communications, each vehicle is equipped with two network interfaces; one for $11p$ and another for LTE. Data and HELLO messages traffic is served by the $11p$ network, whereas Location Service 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

$$\alpha^{ah}(t) = \alpha_D(t) + \alpha_H(t), \tag{39}$$

$$f^{ah}(x) = (f_D \otimes f_H)(x), \tag{40}$$

and for cellular by

$$\alpha^{cell}(t) = \alpha_{bg}(t) + \alpha_{LS}(t), \tag{41}$$

$$f^{cell}(x) = (f_{bg} \otimes f_{LS})(x). \tag{42}$$

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

$$\beta^{ah}(t) = (\beta^1 \otimes \beta^2 \otimes \cdots \otimes \beta^n)(t), \tag{43}$$

$$g^{ah}(x) = (f^1 \otimes f^2 \otimes \cdots \otimes f^n)(x), \tag{44}$$

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 give by

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

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

$$\beta^{LS}(t) = (\beta^{cell}_{LS} \otimes \beta^{int})(t), \tag{46}$$

$$g^{LS}(x) = (g^{cell}_{LS} \otimes g^{int})(x), \tag{47}$$

where the $\langle g^{cell}_{LS}, \beta^{cell}_{LS} \rangle$ is the characteristics of the service provided to the Location Service flow from the LTE network and $\langle g^{int}, \beta^{int} \rangle$ is the characteristics 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

$$\beta^{cell}_{LS}(t) = \beta^{cell}(t) - \alpha_{bg,\theta}(t), \tag{48}$$

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

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

$$\beta^{cell}(t) = (\beta_{uplink} \otimes \beta_{EPC})(t), \tag{50}$$

$$g^{cell}(x) = (g_{uplink} \otimes g_{EPC})(x). \tag{51}$$

and for the downlink as

$$\beta^{cell}(t) = (\beta_{downlink} \otimes \beta_{EPC})(t), \tag{52}$$

$$g^{cell}(x) = (g_{downlink} \otimes g_{EPC})(x). \tag{53}$$

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

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

V. Model Validation and Performance Evaluation

In this section, we numerically evaluate the models for the three networking architectures and validate them through simulations using NS-3 simulator (Subsection V-A). In addition, we perform extensive performance evaluation of the aforementioned network architectures in terms of end-to-end delay (upper bounds and statistical characteristics) for data and signalling traffic, and throughput (Subsection V-B). For the validation of our model we have used a highway segment as a reference scenario. It is depicted in Fig. 6, where vehicles, ordinary cellular users, *RSUs*, *BS* and backhaul network are represented. This road segment is 1km long, and vehicles travel uniformly on two directions with speeds 50-80km/h, while ordinary users are randomly distributed in the area with fixed positions. Each *RSU* serves only one group of vehicles (graphically represented with a circle), while only one *BS* serves all vehicles and all ordinary users of the LTE network. Further, for the evaluation we have used an urban 5x5 Manhattan Grid area, where vehicles travel with speeds 20-50km/h. In this scenario, IEEE 802.11p capable RSUs are located at each intersection and 9 BS cover the area. Vehicles are equipped with IEEE 802.11p and/or LTE communication modules. The distance between each *RSU* is 300m, 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, towards corresponding vehicles in the right-hand side. The configuration parameters for the simulations and the analysis are summarised in Table II.

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

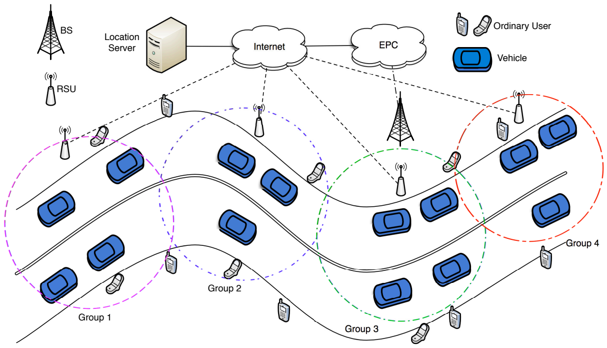


Fig. 6. Reference area for simulations and validation of model

TABLE II
Configuration Parameters

Parameter	**Value**
Number of vehicles (highway)	30 / 60 veh/mile/direction [3]
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

A. Model Validation

Our first scenario consists of 30veh/mile and 10 vehicles from group #1 send data to vehicles from group #4, forming a 3-hop communication at the $11p$ network, using 6Mbps data rate. Figures 7 and 8 show the numerical evaluation of the models and simulation results, and as it can be observed there is a relative tight approximation of the delay bound. For this scenario, the $11p$ and LTE networks are closely competing with each other, while the proposed hybrid shows significantly lower bounds. In terms of location service traffic (Fig. 8), the hybrid network can provide very low bounds compared to the $11p$ network. The LTE scheduler can provide stricter QoS as opposed to the enhanced distributed channel access (EDCA) mechanism of IEEE 802.11p. We can observe, the $11p$ curve has a very long tail resulted by the blocking of certain packets.

The second scenario evaluates $11p$ and hybrid networks when different number 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, is not analysed in this scenario. For 2-hop communications, vehicles from group #1 send data

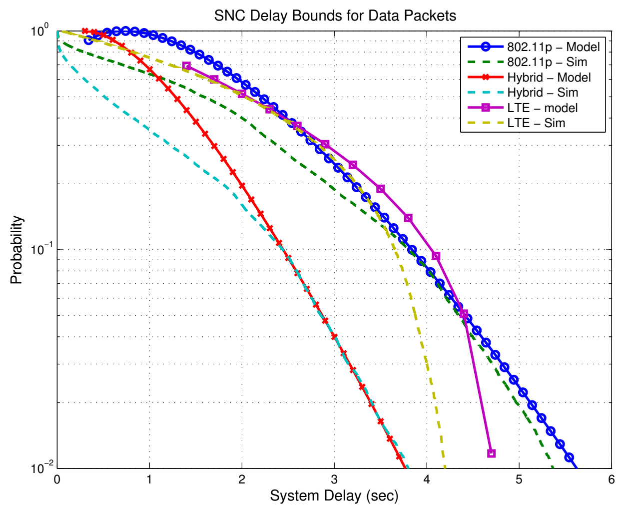


Fig. 7. Comparison of model bound and simulation results for data packets (30veh/mile & 10 con. scenario)

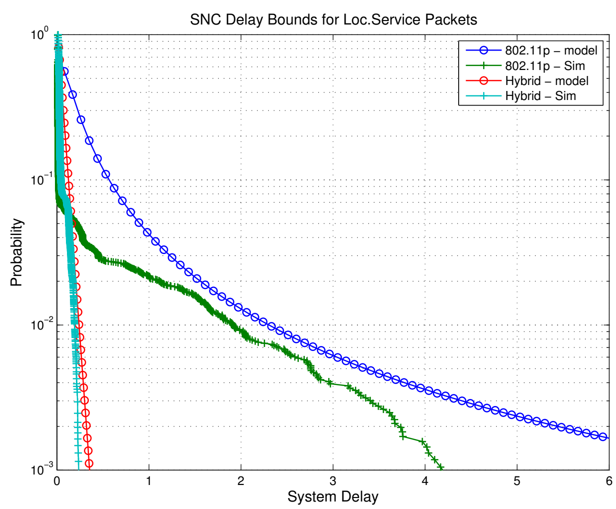


Fig. 8. Comparison of model bound and simulation results for Loc. Service packets (30veh/mile & 10 con. scenario)

to group #3 and for 3-hops to group #4. Figure 9 presents the delay distribution over the $11p$ and hybrid network, respectively. It is obvious that more hops result in increased end-to-end delay, but even with only two hops, the proposed hybrid scenario can provide lower bounds compared to the $11p$ network. This is resulted by the reduced contention levels on the medium due to the separation of signalling 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 6Mbps to 27Mbps has a significant impact in reducing the end-to-end delays. Yet, we have to point out that in order to achieve such high data rates, higher

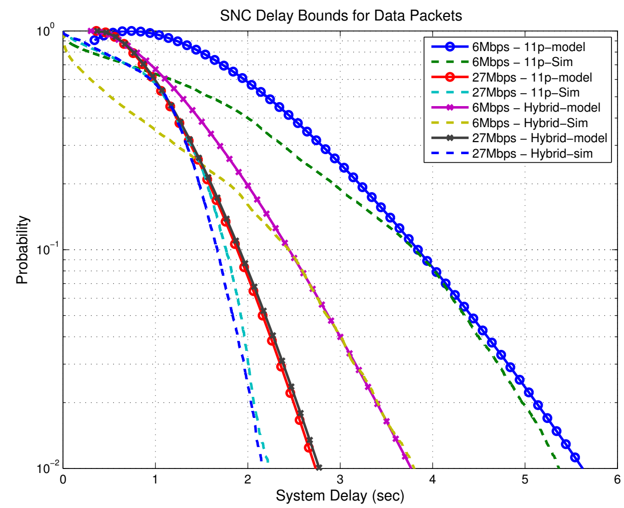


Fig. 10. Comparison of model bound and simulation results for different data rates

modulation and coding schemes (MCS) 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 are fixed to three for both data rates. The impact of increased data rate on the end-to-end delay bounds in the $11p$ and Hybrid scenario are presented in Fig. 10. We observe that with 27Mbps data rate, the two architectures provide relatively similar delay bounds. The available data rate is adequate to carry the data and location service traffic (for the $11p$ network) without increasing the collision probability dramatically, thus keeping the access delay low. However, in the 6Mbps scenario, the bound of $11p$ network is significantly higher than that of the hybrid. This reflects the increased contention levels on the shared medium from both data and signalling traffic.

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 subsection we provide further performance evaluations of the proposed hybrid architecture in terms of end-to-end delay and throughput.

B. Performance Evaluation

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

We increase the number of vehicles to 60veh/mile and the number of connections to 20, and evaluate the 3-hop scenario, with the results presented in Fig. 11. It is clear that the increase in number of vehicles and connections affects the contention on the shared channel of IEEE

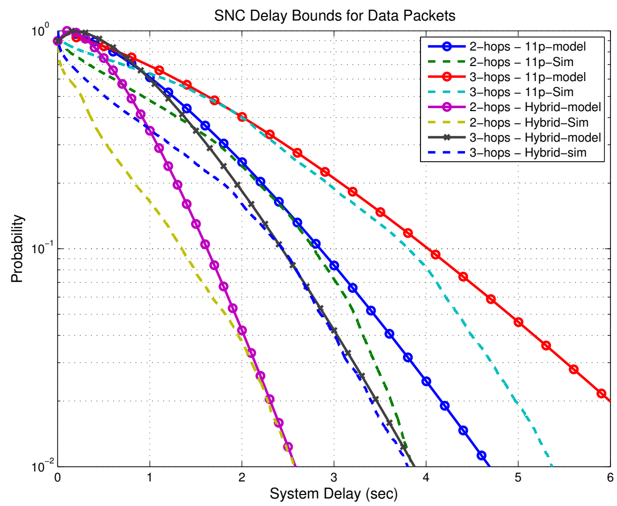
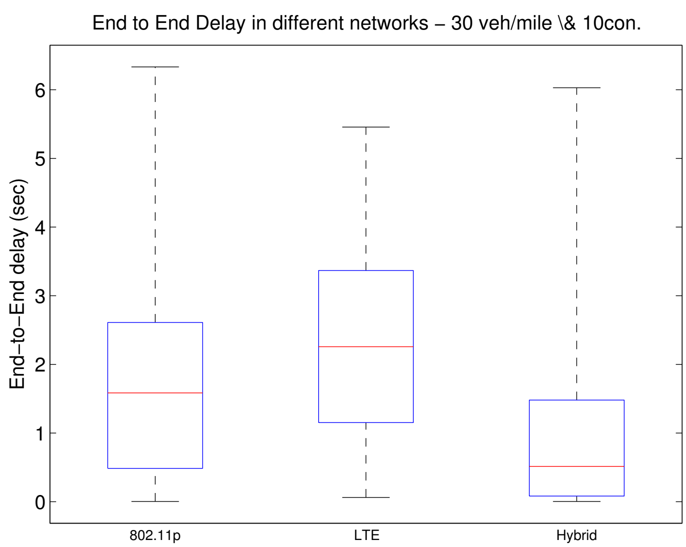
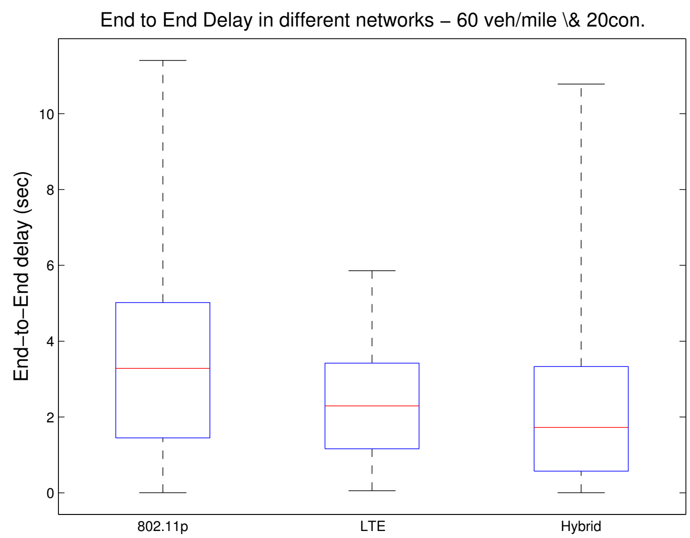


Fig. 9. Comparison of model bound and simulation results for different hop count (30veh/mile & 10 con. scenario)



(a) 30veh/mile & 10 con.



(b) 60veh/mile & 20 con.

Fig. 12. Comparison of end-to-end delay distribution for data traffic

802.11p and the hybrid network, while 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 increased significant and now the hybrid network is closely competing with the LTE network. However, the results suggest that the hybrid network still can deliver better end-to-end delay (average and 75^{th}-percentile) than the other two network architectures (Fig. 12a and 12b). On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points not considered outliers.

In addition to the data traffic, signalling is also important. We evaluate the average end-to-end delay of Location Service traffic in IEEE 802.11p and the hybrid networks, accounting both for uplink and downlink flows. In the previous section we analysed how the delay bounds for $11p$ and hybrid with respect to signalling are formed (Fig. 8) and argued that the hybrid architecture provides lower bounds. Now we measure the average end-to-end delay 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 hybrid network is affected less by the increase in the number vehicles and can deliver lower end-to-end delay when the number of vehicles increases.

Further, we evaluate the aforementioned network architectures in terms of average normalised throughput on the data flows (Fig. 14). In scenarios with low contention

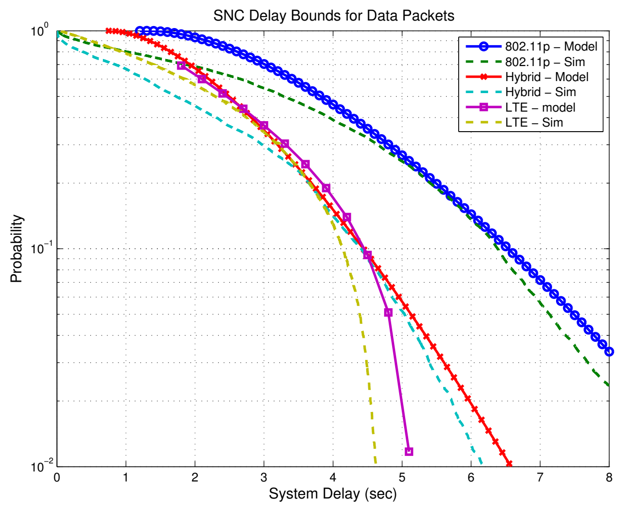


Fig. 11. Comparison of model and simulation results for data packets (60veh/mile & 20 con. scenario)

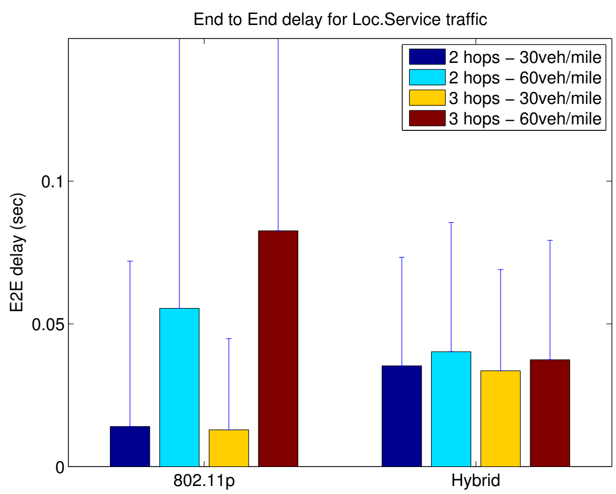


Fig. 13. Comparison of average end-to-end delay for signalling traffic

on the shared channel (30 veh./mile) and when IEEE 802.11p data rate is 27Mbps, all three network architectures provide relatively similar results, with $11p$ and hybrid slightly higher ($< 5\%$) than LTE. Nevertheless, in scenarios with low data rates (6Mbps), when the number of contenting nodes increase, both $11p$ and hybrid show low performance. This degradation of throughput is caused by the increased collision probability in these scenarios.

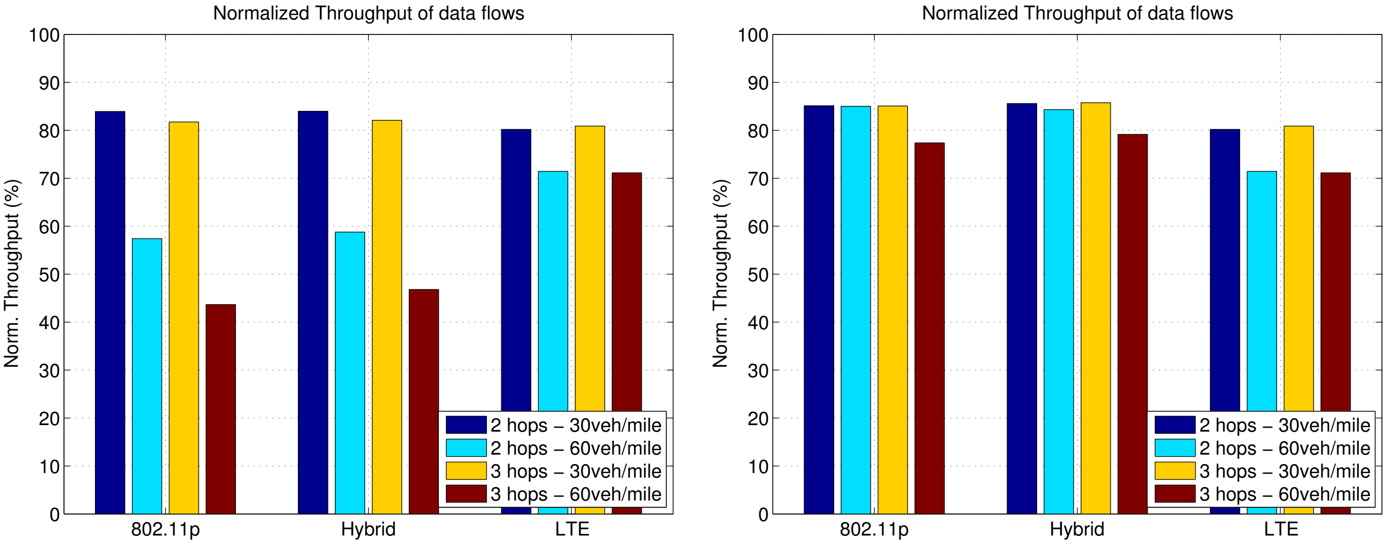
Finally, we evaluate the aforementioned network architectures in an urban environment in terms of average end-to-end delay and investigate the effect of average vehicle speed, traffic load and number of vehicles in the reference area. The results presented in Fig. 15a suggest that LTE-based networks are not affected by the average speed of the vehicles as much as 802.11p-based ones, 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 (Fig. 15b, 15c). 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 signalling traffic separation is proposed in this paper. An analytical model for the calculation of the end-to-end delay bounds for IEEE 802.11p, 3GPP LTE and hybrid vehicular networks is presented in this paper. We use a Stochastic Network Calculus 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 end-to-end delay for different networks that are considered in this paper.

The results presented in this work augment the findings of [4], rather than contradicting them. The type of applications investigated here is unicast with medium data rates and multi-hop ad-hoc connections. In such situations, IEEE 802.11p losses 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 end-to-end delay both for data traffic and signalling.

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 to that of LTE, as reflected by the number of access points or base stations. This is another advantage of the proposed hybrid architecture as it utilises 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 signalling traffic that is carried over LTE will not overhaul the user's monthly bill as for an average daily car use of 1hr, the amount of data sent



(a) IEEE 802.11p @ 6Mbps

(b) IEEE 802.11p @ 27Mbps

Fig. 14. Comparison of throughput for data rate

by the location service does not exceed 3MBytes/month[4].

Future work will aim at LTE-Direct support from 3GPP LTE rel.12, which at the moment is being standardised. The architecture of LTE-D is somehow similar to the proposed hybrid, where signalling 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, multi-channel operation on wireless channels, such as the IEEE 1609.4 standard, could be utilised for data and signalling separation on different channels.

[4]The average daily trip in UK is ∼1hr [36], LS_UPDATE packet size is 100B every 5sec plus LS_REQUEST and LS_REPLY.

References

[1] NHTSA Vehicle Safety and Fuel Economy Rulemaking and Research Priority Plan 2011-2013.

[2] Directive 2010/40/EU of the European Parliament and of the Council of 7 July 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 of Land, Infrastructure, Transport and Tourism, Tech. Rep., 2012.

[4] ETSI TR 102 962, "Intelligent Transport Systems - Framework for Public Mobile Networks in Cooperative ITS (C-ITS)," ETSI, Tech. Rep., 2012.

[5] Z. H. Mir and F. Filali, "LTE and IEEE 802.11p for vehicular networking: a performance evaluation," *EURASIP Journal on Wireless Communications and Networking*, vol. 2014, no. 89, 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 *Network Protocols (ICNP), 2011 19th IEEE International Conference on*, 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," *Mobile Computing, IEEE Transactions on*, vol. 12, no. 11, pp. 2193–2205, 2013.

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

[9] Y. Jiang and Y. Liu, *Stochastic Network Calculus*. Springer, 2008.

[10] ETSI TS 102 636-4-1, "Intelligent Transport Systems - GeoNetworking - Media Independent Functionality," ETSI, Tech. Rep., 2011.

[11] K. Katsaros, M. Dianati, and L. Le, "Effective Implementation of Location Services for VANETs in Hybrid Network Infrastructures," in *IEEE ICC Workshop on Emerging Vehicular Networks: V2V/V2I and Railroad Communications*, 2013.

[12] K. Katsaros, M. Dianati, Z. Sun, and R. Tafazolli, "An evaluation of routing in vehicular networks using analytic hierarchy process," *Wireless Communications and Mobile Computing*, 2015.

[13] C. Perkins, E. Belding, and S. Das, "Ad hoc On Demand Distance Vector (AODV) Routing," 2003.

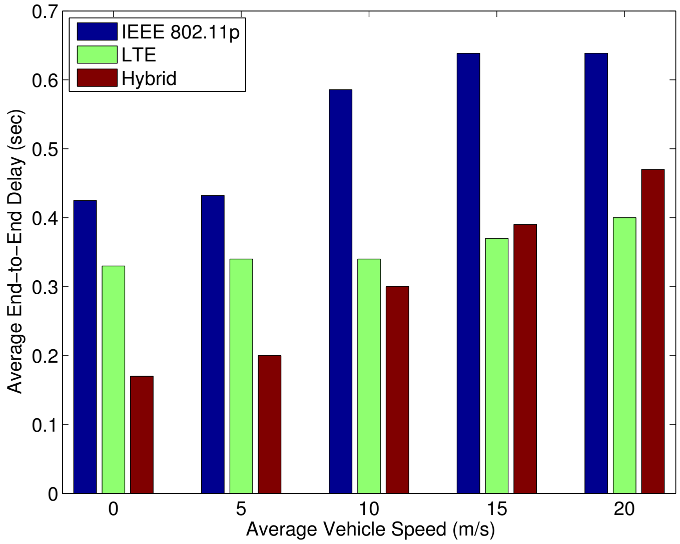
[14] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *Selected Areas in Communications, IEEE Journal on*, vol. 18, no. 3, pp. 535 –547, 2000.

[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," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 10, pp. 3256–3266, 2011.

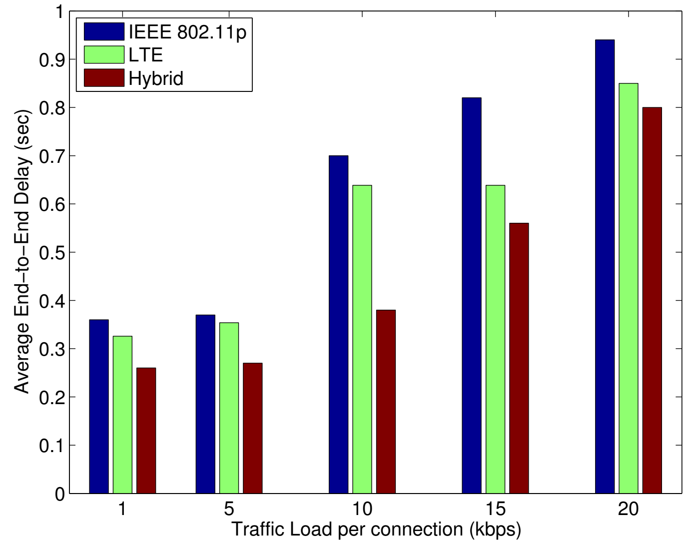
[16] O. Tickoo and B. Sikdar, "Modeling Queueing and Channel Access Delay in Unsaturated IEEE 802.11 Random Access MAC Based Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 16, no. 4, pp. 878–891, 2008.

[17] X. Zhang, "A New Method for Analyzing Nonsaturated IEEE 802.11 DCF Networks," *Wireless Communications Letters, IEEE*, vol. 2, no. 2, pp. 243–246, 2013.

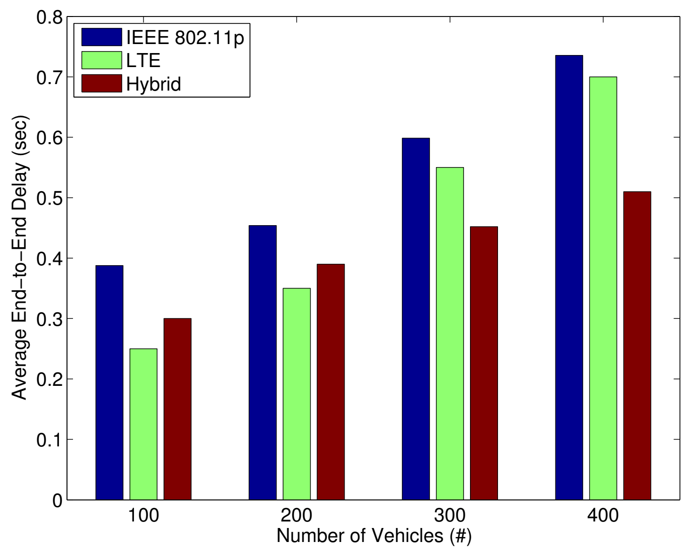
[18] Y. Wang and T. Wang, "Applying stochastic network calculus to 802.11 backlog and delay analysis," in *IEEE Intern. Workshop on Quality of Service (IWQoS)*, 2011.



(a) E2ED Vs Speed



(b) E2ED Vs Load



(c) E2ED Vs Vehicles

Fig. 15. Comparison of average End-to-End delay in urban scenario

[19] J. Xie and Y. Jiang, "A network calculus approach to delay evaluation of IEEE 802.11 DCF," in *IEEE Conference on Local Computer Networks (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," *Networking, IEEE/ACM Transactions on*, vol. 15, no. 3, pp. 588 –601, june 2007.

[21] G. Gupta and N. Shroff, "Delay analysis for multi-hop wireless networks," in *INFOCOM 2009, IEEE*, April 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," *Wireless Communication, IEEE Trans. on*, 2014.

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

[24] H. Graja, P. Perry, and J. Murphy, "A statistical estimation of average IP packet delay in cellular data networks," in *IEEE Wireless Communications and Networking Conference*, 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 *Int. Conf. on New Technologies, Mobility and Security (NTMS)*, 2012.

[26] Y. Gao, X. Zhang, and Y. Chang, "Probabilistic Delay Analysis of Internet of Things Based on LTE Network," *China Communications*, vol. 8, no. 8, 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 *Int. Conf. on Selected Topics in Mobile and Wireless Networking (iCOST)*, 2011.

[28] R. Cruz, "A calculus for network delay. Parts I & II," *Information Theory, IEEE Transactions on*, vol. 37, no. 1, pp. 114 –141, 1991.

[29] J.-Y. Le Boudec and P. Thiran, *Network calculus: a theory of deterministic queuing systems for the internet*. 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 *IEEE Conference on Local Computer Networks*, 2002.

[31] R. Ramaswamy, N. Weng, and T. Wolf, “Characterizing network processing delay,” in *IEEE GLOBECOM*, 2004, pp. 1629–1634.

[32] Y. Zhang and A. Arvidsson, “Understanding the characteristics of cellular data traffic,” in *Proceedings of the 2012 ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design*, 2012, pp. 13–18.

[33] Cisco Systems Inc., “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014–2019,” Tech. Rep., 2015.

[34] C.-S. Chang, *Performance Guarantees in Communication Networks*. Springer, 1999.

[35] Y. Gao, J. Yang, X. Zhang, and Y. Jiang, “Capacity limits for a cognitive radio network under fading channel,” in *Proceedings of the IFIP TC 6th international conference on Networking*, 2011, pp. 42–51.

[36] UK Department of Transport. (2012) National Travel Survey statistics. [online] https://www.gov.uk/government/publications/national-travel-survey-2012.