Analysis of vehicular wireless channel communication via queueing theory model

Scott Fowler, Carl Henrik Häll, Di Yuan, George Baravdish and A. Mellouk

Original Publication:

Scott Fowler, Carl Henrik Häll, Di Yuan, George Baravdish and A. Mellouk, Analysis of vehicular wireless channel communication via queueing theory model, 2014, Communications (ICC) 2014 IEEE International Conference on Communications, 1736-1741.
http://dx.doi.org/10.1109/ICC.2014.6883573

Analysis of Vehicular Wireless Channel Communication via Queueing Theory Model

Scott Fowler*, Carl H. Häll†, Di Yuan*, George Baravdish‡ and Abdelhamid Mellouk§
*Mobile Telecoms., †Traffic Systems ‡Mathematics, Dept. of Science & Tech., Linköping University, Norrköping, Sweden
§LiSSi/TINC laboratory, Dept. of Networks and Telecoms, University of Paris-Est Créteil VdM (UPEC), France

***Abstract*— The 4G standard Long Term Evolution (LTE) has been developed for high-bandwidth mobile access for today's data-heavy applications, consequently, a better experience for the end user. Since cellular communication is ready available, LTE communication has been designed to work at high speeds for vehicular communication. The challenge is that the protocols in LTE/LTE-Advanced should not only provide good packet delivery but also adapt to changes in the network topology due to vehicle volume and vehicular mobility. It is a critical requirement to ensure a seamless quality of experience ranging from safety to relieving congestion as deployment of LTE/LTE-Advanced become common. This requires learning how to improve the LTE/LTE-Advanced model to better appeal to a wider base and move toward additional solutions. In this paper we present a feasibility analysis for performing vehicular communication via a queueing theory approach based on a multi-server queue using real LTE traffic. A M/M/m model is employed to evaluate the probability that a vehicle finds all channels busy, as well as to derive the expected waiting times and the expected number of channel switches. Also, when a base station (eNB) becomes overloaded with a single-hop, a multi-hop rerouting optimization approach is presented.**

I. Introduction

Vehicle-to-infrastructure and vehicle-to-vehicle communication (V2X) has acquired considerable attention from both academia and industry because it holds the potential to improve road safety, decrease vehicle energy usage (e.g. fuel) and to reduce accidents. This is due to the fact that electronic devices are becoming smaller and more powerful, allowing for new features to be introduced into V2X. Consequently, we can expect vehicular communications to become even more popular in the future, resulting in a significant increase in the spectral bandwidth required by these applications. For this reason V2X communication has been considered a key technology of future Intelligent Transportation System (ITS) applications.

With the accessibility of cellular wireless communications, exploring the usage of cellular technology in V2X is contributing to the development of ITS. One cellular communication solution in Fourth Generation (4G) is LTE/LTE-Advanced. This wireless network has been designed to work at high speeds for vehicular communication and high-bandwidth mobile access for today's data-heavy applications. This is important since ITS has to deal with many types of telecommunication traffic ranging from entertainment (e.g. online games, real-time movies) to vehicular safety (e.g. information about vehicle accidents, congestion etc.).

LTE/LTE-Advanced (hence forth referred to as LTE) is able to provide a major contribution to ITS because of it's key features such as higher data rate, higher spectral efficiency, scalable bandwidth, multi-antenna configuration, low user-plane latency, full mobility, co-existence with legacy standards, support for TDD/FDD mode of operations, better power-saving mechanism, seamless integration of unicast and enhanced broadcast transmission, and reasonable system and terminal complexity.

While LTE provides both practical benefits to safety, transport efficiency and service to drivers, there is little consideration to vehicle volume and LTE network having to deal with different services. The challenge is that the protocols in LTE should not only provide good packet delivery but adapt to changes in the network topology due to vehicle volume and vehicular mobility. It is a critical requirement to ensure a seamless quality of experience ranging from safety to relieving congestion as deployment of LTE becomes common. This requires learning how to improve the LTE model to better appeal to a wider base and move toward additional solutions. It will require investigating the amount of available resources that can be assessed using some quantitative performance measures. One way to perform this analysis is via queueing theory, which has been known to permit the derivation of several performance measures including the average waiting time in the queue or the system and the probability of encountering the system in certain states, such as empty and full. In this paper, we present a feasibility analysis for performing vehicular dynamic spectrum access across LTE via a queueing theory approach that is primarily based on a multi-server queue. We provide estimations on performance measures of vehicular communications using M/M/m model.

II. Related Work

Queueing theory has been used for modelling of simple telecommunication transmission to complex distributed telecommunication systems [1].

Some work has been conducted for queueing analysis in cognitive radio systems [2]–[5]. The general model [2] introduced analyzes the performance of a cognitive radio link subject to recurrent failures and interruptions. Also the model was able to work with network models with single or multiple channels. An M/M/1 queue with service interruption from primary users is considered [3] and a pricing policy is used to charge each secondary user that enters a queue. An access strategy for cognitive radio overlay on top of a primary network that takes into account the primary user link adaptation and other primary link information [4]. By doing

this the scheme approximates the optimum method without the need of global channel information. The work of [5] considered four different priority queueing disciplines for its dynamic optimization model. By considering a dynamically priority queueing discipline it allowed for optimize a joint performance utility function of two classes of cognitive radio.

The performance of vehicular network communication in transportation made use of queueing models [6]–[10]. [6] proposed an analytical model showing the performance of periodic broadcast in vehicular ad hoc networks (VANET) in terms of packet collision probability and average packet delay. A M/M/∞ comprehensive vehicular traffic dynamics over a roadway, with intermittently connected networks is presented in [7]. [8], presented analytical models to assess how queue length estimation at intersection is influenced by the percentage of probe vehicles in the traffic stream. A discrete time D/M/1 model for analyzing the performance of periodic broadcast in VANETS is present in [9]. The model show numerical results of packet collision probability and average packet delay.

Wireless LTE network [11] by means of queueing theory, develop a method for 3GPP LTE wireless networks serving a variable bit-rate calls in order to calculate the quality of service perceived by the users. Other queueing theory wireless network models have been considered. [12] presented an analytic model by using packet delay, buffer overflow rates and energy consumption to assess the polling behaviour on a MAC. In their work [13] they develop models to characterize the delay performance of block-based random linear for multicast transmission over a communication channel model wherein errors occur. In paper [14], they develop mixed queueing network models of various mobility users at various access points. This allowed to accurately predicting number of network-level performance and user-level performance in a wireless network.

Cloud computing has become an important paradigm, in which queueing theory models are being presented. [15] models the cloud center as an M/G/m/m+r queueing system with a single task arrivals and a buffer of finite capacity. Cloud using an M/M/m queueing model [16] which takes many factors into considerations such as requirement of service and workload. By doing this the authors were able to formulate an optimal multiserver configuration for profit maximization.

Despite these models there has been no consideration of real LTE data with queueing models. We present a feasible analytical queueing model for determining the minimum number of channels needed, while maintaining real time data quality for V2X communication. Also we show how the queueing model can be used to analyze the effects of rerouting packets.

III. System Model

Figure 1 depicts the basic architecture of LTE. The radio network architecture proposed by the 3GPP LTE consists of evolved NodeB (eNB). The eNB provides a link between the UE and the core network. As shown in Figure 1, eNB is connected to the core network (by means of MME/S-GW). The eNB is responsible for the majority of the radio resource management (RRM) functions such as packet scheduling. Both the Mobility Management Entity (MME) and the Serving Gateway (S-GW) are part of the core networks. The MME is responsible for paging and the UE mobility in idle mode within the network, while the Serving Gateway (S-GW) node is responsible to route user data packets and handle other user requests e.g. handover.

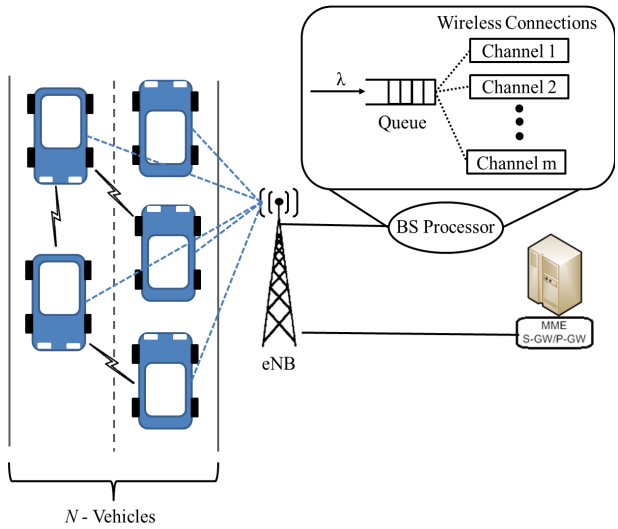

Fig. 1: *Basic LTE network Queueing System Model.*

Figure 1 also shows vehicles that are travelling along a highway and communicating with a LTE wireless network. A bandwidth request from each communication is a request for service. This means each vehicle is not considered a single customer entity in the queueing model. This is represented with the dash line from the vehicle to the eNB. No matter whether it is one-to-one, one-to-many or multiple-to-multiple, the wireless communication is defined as the set of sender and receivers performing wireless transmission. The wireless transmission may be either vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I). It is assumed the setup connection between the sender and receiver is handled before they start looking for available bandwidth for transmission. The available bandwidth resources are within the transmission range of the vehicle and communication channels are set as servers. The transmission time of each communication link is modeled as the service time of each customer. Also included in the waiting time of the wireless channels is the response time of delivering each packet. Plus, the total time cost of each packet also includes the inquiry time to check channel availability database and channel sensing time.

We collected real LTE data and applied to the model. We used a Samsung™ GT-B3730 USB Modem launched by TeliaSonera®. This USB Modem supports LTE (2600MHz) with the peak speedup to 100/50Mbps DL/UL for LTE. We collected the measurements for LTE connected to a Dell Desktop computer having Intel Core 2 Duo 1.6GHz processor and 1GB RAM, and running different kinds of applications using the internet. Wireshark Network Protocol Analyzer version

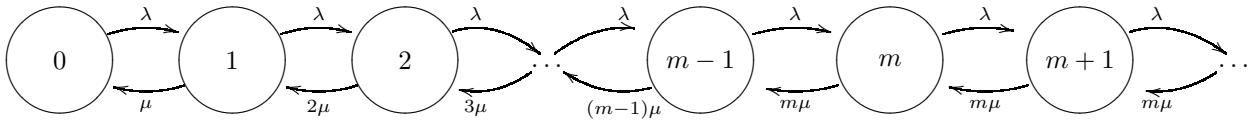

Fig. 2: *State-diagram for a M/M/m system*

1.6.7 was used to capture network traces for the Skype voice. For the Skype data; version 5.10.0.114 was installed on the Desktop computer and the other party having Dell Core i3 laptop with 2GB RAM and running the same Skype version, was called and the data was recorded.

IV. Queueing Theory Model

In this paper the wireless channel connections are treated as an M/M/m queueing system due to the stochastic nature of our traffic source. From each vehicle in the system, we have a Poisson stream of service requests with arrival rate λ, giving inter-arrival times that are independent and identically distribute exponential random variables with mean $1/\lambda$. The service rate of a server in the system is μ, giving service times that are exponentially distributed with mean $1/\mu$. We will let the number of wireless channel connections/servers be *m*, providing service independently of each other. The first-come first-served (FCFS) queueing discipline is adopted. Due to limited space in the paper, we did not focus on a dynamic vehicle environment.

The total traffic intensity ρ also depends on the number of vehicles, *N*, within the transmission range of the eNB, and is defined as:

$$\rho = \frac{N\lambda}{m\mu} \tag{1}$$

The probability that there is no jobs in service in such a system, is given by [1]:

$$P_0(m,\rho) = \left[\sum_{n=0}^{m-1} \frac{(m\rho)^n}{n!} + \frac{(m\rho)^m}{m!(1-\rho)}\right]^{-1} \tag{2}$$

The probability that a newly submitted service request must wait because of all servers being busy (refered to as Erlang-C) is given by:

$$P_q(m,\rho) = \frac{(m\rho)^m}{m!\,(1-\rho)} P_0 \tag{3}$$

Therefore we can say $1 - P_q$ is the probability that at least one wireless channel is available. The expected (mean) response time *T* is the time a job (on average) spends in the system, and is given as:

$$T = \frac{1}{\mu}\left(1 + \frac{P_q}{m\,(1-\rho)}\right) \tag{4}$$

V. Optimal Number of Channels

Figure 2 shows the birth-death-process (state diagram) of an M/M/m system. Finding the optimal number of channels (*m*) in such a system, plays an important role. This is important, because if all channels are busy, the user equipments will lose their connections or will be unable to maintain the connections as the number of vehicles increases.

The number of channels/servers needed to serve N number of vehicles, each requesting data with a packet-rate of λ, in such away that $1-\beta_i$ of the requested packets will be served within time t_i, is given by:

$$P_q e^{-(m\mu - N\lambda)t_i} < \beta_i \tag{5}$$

which gives us:

$$m > \frac{log\left(\frac{1}{\beta_i}P_q\right) + N\lambda t_i}{\mu t_i} \tag{6}$$

Equations (5) and (6), in which the specific values of β_i and t_i depend on what service (i) we are considering, describe the number of channels needed to serve a specific combination of number of vehicles and the amount of traffic they request. To illustrate this, Figure 3 shows the number of channels (m) as a function of λ and N, where λ represents the average demand (packet rate) of every user belonging to the N number of vehicles. Since (as described in Equation (6)) m is a function

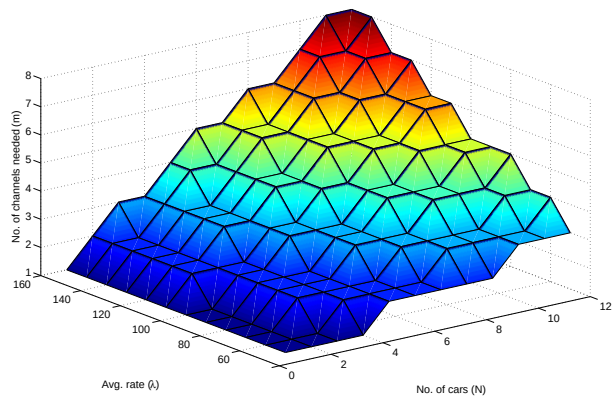

Fig. 3: *Optimal Number of Wireless Channel Connections*

of μ, β_i and t_i, these values have been fixed. t_i and β_i are set to 100 ms and 0.01 respectively, representing the packet delay budget and the packet error loss rate acceptable in the case of voice communication (see Table I) [17], and μ is set to 250 (representing the service rate).

The QoS Class Identifier (QCI) is used to provide information on how to perform packet forwarding treatment (e.g.

QCI	Resource Type	Priority	Delay Budget	Packet Error Loss Rate	Example Service
1	GBR	2	100 ms	10^{-2}	Conv. Voice
2	GBR	4	150 ms	10^{-3}	Conv. Video
3	GBR	5	300 ms	10^{-6}	Buf. Streaming
4	GBR	3	50 ms	10^{-3}	RT gaming
5	Non-GBR	1	100 ms	10^{-4}	IMS signaling
6	Non-GBR	7	100 ms	10^{-3}	Live streaming
7	Non-GBR	6	300 ms	10^{-6}	Buf. streaming, TCP, Email, Web, P2P file sharing, etc.
8	Non-GBR	8	300 ms	10^{-6}	
9	Non-GBR	9	300 ms	10^{-6}	

TABLE I: *QCI characteristics for the bearer QoS profile*

scheduling weights, admission thresholds, queue management thresholds, link layer protocol configuration, etc.). This ensures multivendor deployments and roaming.

In a situation where the total demand within a cell exceeds the capacity of the eNB, packets can either be dropped or rerouted to another eNB. If rerouting packets, the first decision to make is what packets to reroute. Assume one single user equipment trying to send information at an instance in time when the eNB is already used to a large part of its capacity, i.e. the number of channels the user requests cannot be established between user equipment and eNB. In such a situation one can choose to reroute all information from the user (on the number of channels needed) or only reroute part of the information (on the number of channels that could not be established to the original eNB).

The second decision to make is how to reroute the packets. Packets could be rerouted directly to another eNB (single-hop) or relayed over other vehicles (multi-hop). Considering both single-hop and multi-hop, the rerouting can be made according to one out of four possible ways; *all on one path*, *all on many paths*, *part on one path* and *part on many paths*. Meaning whether all traffic requested by a customer (or only part of the requested traffic), is rerouted on one path (or on many paths) to another eNB. The resulting optimization problem is a *Minimum cost flow problem* (MCFP), where the objective is to send the packets that are to be rerouted to the eNB to the lowest possible cost.

Assuming a directed network $G(V, A)$, where V is the set of all vertices (vehicles and eNBs in the area of interest), and A is the set of all arcs (representing the wireless transmission possibilities between the vertices) in the network, where each arc (i, j) in A has a cost c_{ij} of sending a packet along the arc and a capacity u_{ij}, and where each vertex i in V has a value b_i describing the supply or demand of that vertex. Then the problem can be formulated as (e.g. described in [18]):

$$\min \sum_{(i,j)\in A} c_{ij} x_{ij} \tag{7}$$

subject to:

$$\sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = b_i \quad \forall\, i \in V \tag{8}$$

$$0 \le x_{ij} \le u_{ij} \quad \forall\, (i,j) \in A \tag{9}$$

where x_{ij} represents the flow on arc $(i, j) \in A$.

In our case, the capacity u_{ij} of each arc is defined by the number of available channels on that arc, and b_i describes the total capacity needed to serve the demand of sender i. The cost c_{ij} can be a combination of several actual costs. We consider it only to be the time delay caused by using the arc at a specific time.

Under some specific assumptions the problem of finding the best routing of packets from one specific sender can also be treated as a *Shortest path problem*. If we consider a routing strategy where only one path is used in the network (all to one or part to one) and we do not consider any capacity constraints we have a shortest path problem. In this case all that remains is to find the shortest (cheapest) way to send the information from the sender to an eNB, (using single- or multi-hop). However, this is under the assumption that all arcs considered in the network have enough free capacity to be able to serve the amount of information the sender wants to transmit, i.e. to be able to establish a sufficient number of channels. If this assumption does not hold, the minimum cost flow formulation is kept, (but many algorithms for this problem solve several instances of shortest path problems). For examples of algorithms for these problems, see e.g. [18].

Since we consider the cost of the arcs only to be the delay, and we make it even more simplified by considering the delay to be static (for the short time period we consider) the total delay of all the flow of a solution to the problem is the same as the objective function value of that solution, described in Equation (7). However, if we let S_k be the set of arcs belonging to path k, the total delay of the flow on an individual path k (TD_k) can be expressed as:

$$TD_k = \sum_{(i,j)\in S_k} c_{ij} x_{ij} \tag{10}$$

and the total delay of an individual packet on path k (PD_k) is:

$$PD_k = \sum_{(i,j)\in S_k} c_{ij} \tag{11}$$

If the capacity restrictions (u_{ij}) on the arcs forces the rerouted packets to be divided on several paths, the problem is still only to find k shortest paths (as long as we consider only one additional UE at a time to serve). In such a case one is to find the shortest path and put as much flow (as many packets) as possible (under the limitations of u_{ij}) on that path and then search for the second shortest path. This is done until all packets to reroute have been served by some path or until no more paths with capacity left can be found.

Now, if the total delay of a packet on any path that is part of the solution is greater than the acceptable packet delay budget β_i then the packets will be dropped, meaning that it was not possible to find a way to reroute all the data without loss of packets, i.e. if $PD_k > \beta_i$, packets sent on path k are dropped. If less than t_i of all the packets are dropped, the service is considered to be "good enough", and the solution of moving

some of the demand away from the overloaded cell to another eNB was still successful even though it was not done without loss of packets.

VI. Results

To illustrate the use of the queueing theory model and the rerouting approach, we have created a small example described in Figure 4. From the LTE equipment experiments, we have used $\lambda = 97$ and the average packet size was about 120 Bytes. To make an illustrative example, we have chosen $\mu = 250$ and $m = 5$ to describe the service capability of the LTE network. We also assume that all vehicles request voice communication, which means that the index i for β_i and t_i is redundant. Therefore we have used $\beta = 0.01$ and $t = 0.1$, according to [17]. These numbers give us (by fulfilling Equation (6)) that 12 users can be served at the same time by the eNB (x-axis in Figure 3). When an additional user requests service in the area of that eNB, the request (or at least part of the 97 packets requested per second) must be rerouted or dropped.

The cost c_{ij} of using arc (ij) is for vehicle-to-vehicle communication considered only to be a fixed cost (τ_{ij}), while it for vehicle-to-eNB communication is a combination of the fixed cost and the delay caused by queueing for service and time spent in service. This delay d_{ij} is the expected response time T, that was given in Equation (4). However, since T depends on the amount of traffic being requested from/to the eNB in question (eNB j), so does d_{ij}.

In our small example based on Figure 4, let λ_j be the sum of previously assigned traffic to j, called AT_j, and the amount of traffic we reroute to j. Then, λ_j can be expressed as:

$$\lambda_j = AT_j + \sum_{i \in V} x_{ij} \tag{12}$$

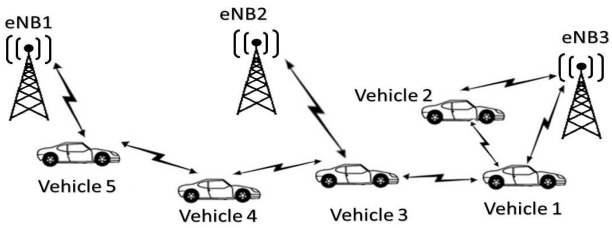

Fig. 4: *Wireless Network Topology.*

This means that c_{ij} will be:

$$c_{ij} = \begin{cases} \tau_{ij}, & \text{for V2V} \\ \tau_{ij} + \frac{1}{\mu}\left(1 + \frac{P_q}{m(1-(\lambda_j/\mu))}\right), & \text{for V2I} \end{cases} \tag{13}$$

and the values of τ_{ij} for our example are given as:

$$\tau_{ij} = \begin{cases} 20 \text{ ms}, & \text{if } j \in \text{set of vehicles} \\ 30 \text{ ms}, & \text{if } j \in \text{set of eNB} \end{cases} \tag{14}$$

This formulation of the cost for using a V2I arc will not affect what route that will be used to a specific eNB (other than choosing the route with least hops), since all arcs leading to the same eNB station are affected in the same way. However, it can affect what eNB the packet is routed to.

In Figure 4, assume that eNB2 is not used at all, and vehicles (such as vehicle 3) are starting to request service from it, by transmitting packets to it. The total (and average) throughput will then increase as described in Figure 5 (and Figure 6) according to the values of "Throughput without Rerouting (RR)".

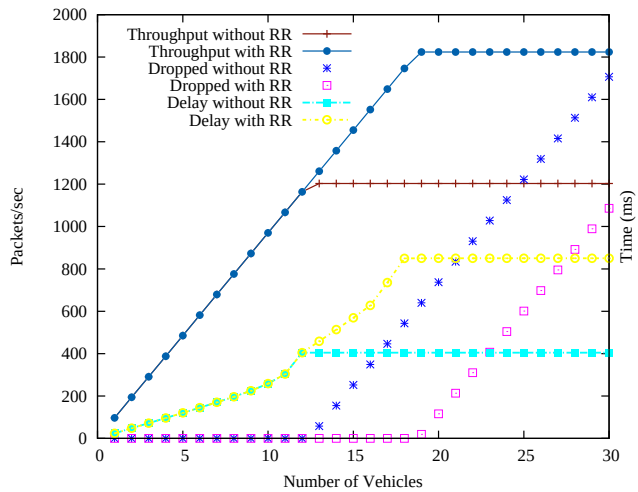

Fig. 5: *Total performance in relation to no. of Vehicles.*

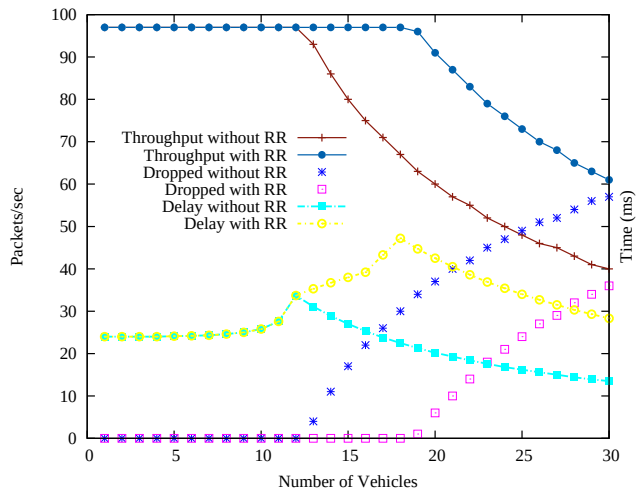

Fig. 6: *Average performance in relation to no. of Vehicles.*

From the queueing theory analysis, and as can be seen in Figure 5, we see that the throughput of the vehicles using eNB2 is not affected until the number of vehicles increase to 13, since Equation (5) is still fulfilled for twelve vehicles. However, while the number of vehicles increases, so does the delay of the packets that are sent. When 13 or more vehicles are requesting service from eNB2, the throughput decreases

since packets are starting to get dropped. This is of course due to that the wireless channels to eNB2 have reached full capacity resulting in packets being dropped as the number of vehicles increases (labeled Dropped without RR). A packet is considered lost if the UE (vehicle) does not acknowledge it within the delay threshold. Delay threshold is in accordance to LTE QCI characteristic as shown in Table I.

What happens in the case with rerouting, of course depends on the available amount of resources in the eNBs that packets can be rerouted to (eNB1 and eNB3). In the test-case, that the results in Figure 5 and Figure 6 are based on, it is assumed that eNB1 is serving 10 vehicles and eNB3 is serving 8 vehicles. So, when eNB is fully used, packets from 2 vehicles can be rerouted to eNB1 and packets from 4 vehicles can be rerouted to eNB3. Since the cost of rerouting them to eNB3 is lower than to reroute them to eNB1 (see Equations (13) - (14)), eNB3 is chosen first. By means of searching for other eNBs than the first option, it is in this case possible to serve 18 of the vehicles that initially wanted to be served by eNB2.

By allowing rerouting, the total throughput increases (labeled Throughput with RR) since it is now possible to serve additionally 6 requests using other vehicles to reroute the data traffic to other eNBs (eNB1 and eNB3) making use of these eNBs free capacity and the routes acceptable low delay. In the results the packet delay is larger when using rerouting (labeled Delay with RR) since the number of packets increase but it does not decrease performance since the number of packets dropped do not occur until the number of vehicles increase to 19. At the same time we can see from Figures 5 - 6 that the delay increases since the rerouted packets take longer time to deliver.

In this paper we have showed how queueing theory can be incorporated in the routing of packets, i.e. in solving a MCFP for wireless link communication with an extra set of constraints saying that the total flow of packets to each eNB may not exceed the total amount of traffic the eNB can handle and still maintain the level of service required.

The new set of constraints can be formulated as:

$$\sum_{i:(i,j)\in A} x_{ij} \leq \frac{m\mu t_k - log(\frac{P_q}{\beta_k})}{t_k} \quad \forall\, j \in E \qquad (15)$$

where E is the set of eNBs, k is the assumed service according to Table I, P_q is defined according to Equation (3) and the costs c_{ij}, used in the original constraints of the MCFP, are defined according to Equation (13).

VII. Conclusion

In this paper we present a multi-server queueing model that considers the optimal number of channels for the analysis. In addition to the M/M/m queueing model, when an eNB becomes overloaded with a single-hop, a multi-hop rerouting optimization approach is presented. The paper shows how the queueing theory approach can provide good estimations of what costs, in form of time delay, to use when solving the routing problem. This work is useful for showing the best channel parameters for LTE so as to achieve an efficient network to an acceptable QoS level depending on the QCI and desired outcome of the network traffic. By not selecting the best parameters or optimal route will affect the network traffic performance.

Acknowledgment

Scott Fowler was partially supported by EC-FP7 Marie Curie CIG grant, Proposal number: 294182.

References

[1] Leonard Kleinrock. *Queuing Systems, Volume II: Computer Applications*, volume 2. Wiley, 1976.

[2] A. Azarfar, J. Frigo, and B. Sanso. Analysis of cognitive radio networks based on a queueing model with server interruptions. In *IEEE International Conference on Communications (ICC)*, pages 1703–1708, June 2012.

[3] Husheng Li and Zhu Han. Socially optimal queuing control in cognitive radio networks subject to service interruptions: To queue or not to queue? *IEEE Transactions on Wireless Communications*, 10(5):1656–1666, May 2011.

[4] F.E. Lapiccirella, Xin Liu, and Zhi Ding. Distributed control of multiple cognitive radio overlay for primary queue stability. *IEEE Transactions on Wireless Communications*, 12(1):112–122, January 2013.

[5] A. Azarfar, J.-F. Frigon, and B. Sansó. Dynamic selection of priority queueing discipline in cognitive radio networks. In *IEEE Vehicular Technology Conference (VTC-Fall)*, pages 1–5, September 2012.

[6] Qiong Yang, Jun Zheng, and Lianfeng Shen. Modeling and performance analysis of periodic broadcast in vehicular ad hoc networks. In *IEEE Global Telecommunications Conference (GLOBECOM)*, pages 1–5, December 2011.

[7] M.J. Khabbaz, W.F. Fawaz, and C.M. Assi. A simple free-flow traffic model for vehicular intermittently connected networks. *IEEE Transactions on Intelligent Transportation Systems*, 13(3):1312–1326, September 2012.

[8] G. Comert and M. Cetin. Analytical evaluation of the error in queue length estimation at traffic signals from probe vehicle data. *IEEE Transactions on Intelligent Transportation Systems*, 12(2):563–573, June 2011.

[9] M.J. Khabbaz, W.F. Fawaz, and C.M. Assi. Modeling and delay analysis of intermittently connected roadside communication networks. *IEEE Transactions on Vehicular Technology*, 61(6):2698–2706, July 2012.

[10] Shengbo Yang, Chai Kiat Yeo, and Bu Sung Lee. Predictive scheduling in drive-thru networks with flow-level dynamics and deadlines. In *IEEE International Conference on Communications (ICC)*, pages 1–5, June 2011.

[11] M. Karray and M. Jovanovic. A queueing theoretic approach to the dimensioning of wireless cellular networks serving variable bit-rate calls. *IEEE Transactions on Vehicular Technology*, 62(6), July 2013.

[12] Haiming Yang and B. Sikdar. Queueing analysis of polling based wireless mac protocols with sleep-wake cycles. *IEEE Transactions on Communications*, 60(9):2427–2433, September 2012.

[13] B. Shrader and A. Ephremides. Queueing delay analysis for multicast with random linear coding. *IEEE Transactions on Information Theory*, 58(1):421–429, January 2012.

[14] Yung-Chih Chen, J. Kurose, and D. Towsley. A mixed queueing network model of mobility in a campus wireless network. In *IEEE INFOCOM*, pages 2656–2660, March 2012.

[15] H. Khazaei, J. Misic, and V.B. Misic. Performance analysis of cloud computing centers using m/g/m/m+r queuing systems. *IEEE Transactions on Parallel and Distributed Systems*, 23(5):936–943, May 2012.

[16] J. Cao, K. Hwang, K. Li, and A. Zomaya. Optimal multiserver configuration for profit maximization in cloud computing. *IEEE Transactions on Parallel and Distributed Systems*, 24(6):1087–1096, June 2012.

[17] K.I. Pedersen, T.E. Kolding, F. Frederiksen, I.Z. Kovacs, D. Laselva, and P.E. Mogensen. An overview of downlink radio resource management for utran long-term evolution. *IEEE Communications Magazine*, 47(7):86–93, July 2009.

[18] Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network flows: theory, algorithms, and applications*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.