Estimating the Required Resources in a Vehicular Ad Hoc Network Using Queueing Theory Models

Irene Keramidi
Dept. Informatics & Telecommuications
University of Peloponnese
Tripolis, Greece
ekeramidi@uop.gr

Ioannis Moscholios
Dept. Informatics & Telecommuications
University of Peloponnese
Tripolis, Greece
idm@uop.gr

Michael Logothetis
Dept. Electrical and Computer Engineering
University of Patras
Patras, Greece
mlogo@upatras.gr

Panagiotis Sarigiannidis
Dept. Electrical & Computer Engineering
University of W. Macedonia
Kozani, Greece
psarigiannidis@uowm.gr

Abstract— **In this paper, we evaluate the performance of a Vehicular Ad Hoc Network (VANET) exploiting the *M*/*M*/*c* and the *M*(*N*)/*M*/*c* queueing models in order to determine the number of the resources that are required to satisfy the Quality of Service (QoS). To evaluate the system's performance we examine various quantities, such as the system's waiting time, the throughput and the number of users being served. These performance metrics are important for a VANET as they directly affect the QoS. The queueing models are employed in the following cases: a) when a vehicle communicates directly with a gNodeB (vehicle-to-infrastructure communication) and b) when a vehicle communicates with other vehicles on the road (vehicle-to-vehicle communication) in order to transmit indirectly its request to the destination gNodeB. Based on our study, we show that when the minimum possible resources are exploited, the number of users being served may be highly reduced and consequently the system's throughput can be decreased. Therefore, this condition should be adopted only when it is guaranteed that the system's overall performance will not be dangerously affected.**

Keywords—VANET, Poisson, finite, QoS, delay, throughput

I. Introduction

In the concept of Intelligent Transportation Systems (ITS) [1] a vehicular ad-hoc network (VANET) consists of fixed infrastructure units and moving vehicles and offers ITS services to users [2]. In a VANET, vehicle's communication is performed mainly using two methods: i) two vehicles communicate with each other, thus forming a vehicle-to-vehicle (V2V) communication or ii) a vehicle communicates with a fixed road side unit, thus forming a vehicle-to-infrastructure (V2I) communication. The combination of V2V and V2I technologies is described using the term Vehicle-to-Everything (V2X) (see Fig. 1).

V2X can provide increased road safety, accident reduction, travelling time minimization as well as it is a contributing technology to reduce the energy footprint since it provides mechanisms for fuel economy and traffic congestion avoidance. However, V2X can offer different applications that have diverse performance requirements, which in many cases are very strict, e.g. requiring extremely low latency. In this context, a VANET constitutes a dynamically varying network environment due to the high speeds of moving vehicles [3], [4]. In this highly demanding environment, the accurate system's performance estimation is a very challenging task.

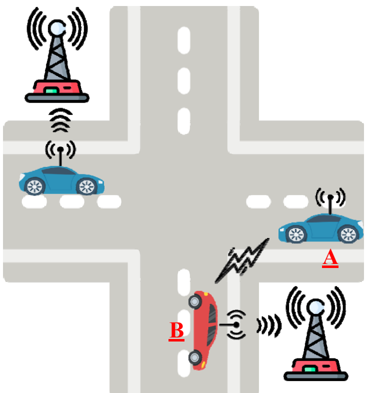

Fig. 1: Basic system model for a VANET

In this work, we examine the performance of a VANET by calculating critical metrics, e.g. throughput and delay, that affect the quality of service (QoS). To this end, we examine two queueing models for the determination of the number of the necessary resources that guarantee the QoS. In particular, the springboard for our analysis is [5], [6] where the classical *M*/*M*/*c* model was adopted. The main drawback of this model is that the arrival process is Poisson, an assumption that implies an infinite number of vehicles and leads to a VANET over-dimensioning. Note that recently, in [7], the $M/E_k/c$ queueing model has been also considered, in which the arrival process is Poisson but the service time follows the Erlang type-*k* distribution. Herein, we circumvent the drawback of the Poisson arrival process, by applying the *M*(*N*)/*M*/*c* model to estimate critical performance metrics, e.g. throughput and delay, as in [5]. The main advantage of the *M*(*N*)/*M*/*c* model is that it considers a finite number of vehicles, denoted as *N* and serviced by the VANET. Both the *M*/*M*/*c* and the *M*(*N*)/*M*/*c* models have been studied for two scenarios: a) when a vehicle communicates with a gNodeB (gNB) directly (V2I communication) and b) when a vehicle communicates with other vehicles on the road (V2V communication) in order to transmit indirectly its request to the destination gNB. Under both scenarios, we show that the *M*(*N*)/*M*/*c* model provides better results both in terms of

979-8-3503-9958-5/22/$31.00 ©2022 IEEE

throughput and delay when the number of their resources is equal. However, when the minimum resources condition is adopted i.e. when we use the minimum possible resources for satisfying the QoS, the number of required resources in the *M*(*N*)/*M*/*c* model is significantly smaller compared to the corresponding in *M*/*M*/*c* model. As a consequence, the number of vehicles being served is highly reduced which consequently affects the system's performance.

The remainder of this paper is the following: In Section II, we review the model of [5]. In Section III, we initially present the *M*/*M*/*c* and the *M*(*N*)/*M*/*c* models, next we provide the corresponding formulas for the determination of the main performance measures and finally we present a tutorial example. Numerical results of this example are presented and discussed in Section IV. Finally, conclusions can be found in Section V.

II. SYSTEM DESCRIPTION

To portray a VANET, we form a directed and connected graph $G(V,E)$ where the vertices V represent both the moving vehicles and the gNBs while the wireless interconnections between them are denoted by the edges E [5]. Each edge (i, j) in E is identified by: a) the edge's capacity v_{ij} and b) the cost c_{ij} of transmitting a packet along this arc. In what follows, we assume that the cost c_{ij} corresponds to the time delay spent using the edge (i, j) and we consider that the delay is static. In a VANET, the traffic generated by vehicles (or vertices) is transmitted over the wireless links (or edges) to the destination gNB in order the request to be served. The packet transmission can be either single-hop or multi-hop. According to Fig.1, a vehicle can communicate either directly with a gNB forming a single-hop transmission (e.g., vehicle B) or indirectly, that is when vehicle A, transmits its packets indirectly to the gNB exploiting vehicle B, conducting in this case a multi-hop transmission. Note that during the multi-hop transmission, the intervening vehicles that may exist between the sender and the receiver do not have an active role on the transmission.

In general, each vehicle transmits its packets to the nearest gNB in order to be served. However, if the gNB is fully occupied or it cannot be accessed, the request for service can be rerouted to another gNB. If all gNBs' resources are either fully utilized or not enough vehicles on the road do exist to provide the required connections for the multi-hop transmission, the corresponding packets are dropped.

Moreover, the receiving gNB can be almost fully utilized, which means that its resources may be sufficient for serving only a part of the vehicle's request. On this condition, there exist two cases: a) some packets of the vehicle's request will be transmitted and served by the initially receiving gNB while the rest will be sent for service to another gNB and b) the whole request will be rerouted to another gNB. Therefore, the rerouting process includes four options: a) all packets are transmitted over a single path, b) all packets are transmitted over multiple paths, c) part of the packets is transmitted over a single path or d) part of the packets is transmitted over multiple paths.

In addition, the necessity arises to define the suitable path to route the packets i.e. the path with the minimum total cost [5]. Since the cost c_{ij} of using a link (i, j) refers to the time spent to transmit the packets on this link, the path with the minimum total cost is the path with the minimum total delay. When the optimal path is considered, we should also examine whether the link capacities are sufficient to transfer the traffic to the destination. Under these conditions, it is assumed that the links have the necessary capacity to deliver the transmitted packets and the routing problem can be handled as a shortest path problem (SPP) whose target is to find the path with the minimum cost, which in our case is equal to the minimum total delay. If the packets are transmitted over multiple paths, the main target is to find the k shortest paths. In particular, this process can be described as follows: First, we estimate the shortest path to transfer as many packets as possible. Then we execute the same procedure until all packets are sent to their destination.

The packet delay of a single packet on path k is given by [5]:

$$PD_k = \sum_{(i,j)\in U_k} c_{ij} \tag{1}$$

where U_k is the set of links establishing the individual path k.

The total delay of the traffic on path k is given by [5]:

$$TD_k = \sum_{(i,j)\in U_k} x_{ij} c_{ij} \tag{2}$$

where x_{ij} denotes the traffic transmitted through a link $i \rightarrow j$.

III. QUEUEING MODELS BASED PERFORMANCE EVALUATION

A. The M/M/c model

We consider the 5G-based system of Fig. 1 that consists of user equipments (UE) formed by vehicles moving across the highway and gNBs, which are responsible for establishing a wireless connection between the core network and the UEs. All gNBs are identical and each of them provides c wireless channels for connectivity. Each wireless connection is treated as an *M*/*M*/*c* system with a first-in, first-out (FIFO) queueing discipline. The vehicles communicate with each other forming a V2V communication and as such a wireless communication may be either V2I or V2V. Packets arrive at gNBs via a Poisson process with arrival rate λ. The transmission time of a packet over the communication link is considered as the service time and it is an exponential random variable with mean $1/\mu$. However, in case that all servers are fully occupied, the arriving request should wait to be served. The probability that all servers are busy, is given by the classical Erlang-C formula [8]:

$$P_{busy} = \frac{a^c}{c!}\frac{c}{c-a}P(0) \tag{3}$$

where a is the offered traffic load equal to $\alpha = \lambda/\mu$ and $P(0)$ is the probability that the system is empty defined as:

$$P(0) = \frac{1}{\sum_{n=0}^{c-1}\frac{a^n}{n!} + \frac{a^c c}{c!(c-a)}} \tag{4}$$

In addition, the total time T of a packet in the system is given by:

$$T = W_q + \frac{1}{\mu} = \frac{a^c}{c!c\,\mu(1-p)^2}P(0) + \frac{1}{\mu} \tag{5}$$

where W_q refers to the time that a packet remains in the queue and p is the server utilization.

By denoting with N the number of vehicles within the coverage area of a gNB, the server utilization p is determined by:

$$p = \frac{N\lambda}{c\mu},\ p < 1 \tag{6}$$

By assuming that the probability that a packet will wait above time t, $\Pr\{T_q > t\}$, should not exceed a predefined threshold β, we have that [5]:

$$\beta > P_{busy} e^{-(c\mu - N\lambda)t} \tag{7}$$

Given the values of N, μ, λ, t and β, the optimal number of system's resources can be determined via (7) by recursively increasing the value of c.

B. The M(N)/M/c model

We consider the aforementioned 5G-based system which consists of gNBs and vehicles that move across the highway and communicate with a 5G wireless network. In this case, each wireless connection is modeled as an $M(N)/M/c$ system with an infinite FIFO queue. We assume that all servers are identical and their service times are exponential random variables with mean $1/\mu$. In the $M(N)/M/c$ model, the population of sources is of size N and each idle source has a call arrival rate equal to v. Therefore, the offered traffic load per idle source, a_{fin}, is defined as [8]:

$$a_{fin} = v / \mu \tag{8}$$

The probability that the system is empty is given by:

$$P_{fin}(0) = \left[\sum_{n=0}^{c-1} \binom{N}{n} a_{fin}^{n} + \sum_{n=c}^{N} \binom{N}{n} \frac{n!}{c!c^{n-c}} a_{fin}^{n} \right]^{-1} \tag{9}$$

Similarly, the probability that n users are being served, $P_{fin}(n)$, is calculated via:

$$P_{fin}(n) = \begin{cases} \binom{N}{n} a_{fin}^{n} P_{fin}(0) & (1 \le n \le c-1) \\ \binom{N}{n} \frac{n!}{c!c^{n-c}} a_{fin}^{n} P_{fin}(0) & (c \le n \le N) \end{cases} \tag{10}$$

In order to calculate the probability of waiting in the queue the following equation can be used:

$$P_{fin,busy} = \sum_{n=c}^{N} P_{fin}(n) \tag{11}$$

The total time, T, that a packet remains in the system is determined by:

$$T = \frac{L}{v(N-L)} \tag{12}$$

where L is the mean number of packets in the system given by:

$$L = \sum_{n=1}^{N} n P_{fin}(n) \tag{13}$$

The server utilization p_{fin} in the $M(N)/M/c$ model is calculated by:

$$p_{fin} = \frac{v(N-L)}{c\mu} \tag{14}$$

Based on (13) and (14), a change on the number of vehicles N modifies the values of L and p_{fin}. Thus, to determine the number of vehicles N whose offered load results in a certain p_{fin}, we can increase N until the value of p_{fin} is met.

The probability that a packet will wait above t is defined as:

$$P_r\{T_q > t\} = 1 - W_q(t) = \sum_{n=c}^{N-1} Q(n) \sum_{i=0}^{n-c} \frac{(c\mu t)^i e^{-c\mu t}}{i!} \tag{15}$$

where $Q(n)$ is the probability that n packets are being served when an arriving packet requests for service and it is determined by [8]:

$$Q(n) = \frac{(N-n)P_{fin}(n)}{N-L} \tag{16}$$

As a consequence, the number of servers c can be determined by assuming that $P_r\{T_q > t\}$ should not exceed a predefined threshold β, i.e.:

$$\beta > \sum_{n=c}^{N-1} Q(n) \sum_{i=0}^{n-c} \frac{(c\mu t)^i e^{-c\mu t}}{i!} \tag{17}$$

C. A tutorial example

In the direction of exploiting the aforementioned queueing models for estimating the number of the required resources in a 5G-based V2X network, we present a short example where both the $M/M/c$ and $M(N)/M/c$ models are exploited. In more detail, as it is presented on Fig. 2, our system consists of 3 gNBs and some vehicles crossing the road where both V2I and V2V communication options are available. Each gNB provides c = 7 wireless channels while the service rate per channel is an exponential distributed random variable with mean μ = 100 packets/s. Apropos of the application parameters β and t, we consider that their values are equal to 0.01 and 0.1 s, respectively, which are representative values of common applications, e.g. voice. By exploiting these values and applying them in (7) and (17), we can determine the minimum number of servers per gNB needed to fulfill this demand in both the queueing models. In more detail, in the $M/M/c$ model the number of minimum necessary servers is equal to c = 6 while in the $M(N)/M/c$ model is c = 2. Moreover, the cost c_{ij} of using arc (i,j) refers to both V2V and V2I communication and it is given by:

$$c_{ij} = \begin{cases} N_j \tau_{ij}, & \text{for V2V} \\ N_j (\tau_{ij} + T), & \text{for V2I} \end{cases} \tag{18}$$

where τ_{ij} is a fixed link cost whose value depends on the type of communication and it is defined by (19). In particular, in the V2V communication, the cost is considered to be fixed (τ_{ij}) while for the V2I communication the cost is the sum of a fixed cost (τ_{ij}) and the total time that a packet spends in the system T given by (5) and (12) for the *M*/*M*/*c* and *M*(*N*)/*M*/*c* models respectively:

$$\tau_{ij} = \begin{cases} 30\,\text{ms}, & \text{if } j \text{ is vehicle} \\ 20\,\text{ms}, & \text{if } j \text{ is a gNB} \end{cases} \tag{19}$$

In the *M*/*M*/*c* model, packets arrive at the system according to a Poisson process which implies that the number of sources is infinite. By denoting that the arrival rate is λ = 90 packets/s, the total traffic λ_j that is served at a specific time instance in a gNB is:

$$\lambda_j = N_j \lambda + \lambda_{rer,j} \tag{20}$$

where N_j denotes the number of vehicles communicating with gNB j and $\lambda_{rer,j}$ is the traffic that is rerouted to this gNB.

In accordance, the total throughput at a specific time instance in the system is defined as:

$$Throughput = \sum_{j \in V} \lambda_j \tag{21}$$

In the *M*(*N*)/*M*/*c* model the number of sources is considered to be finite and each idle source has a packet arrival equal to v = 90 packets/s. Let v_j be the total traffic that is served at a specific moment in gNB j, determined via:

$$v_j = (N_j v + v_{rer,j}) / N_{tot,j} \tag{22}$$

where $v_{rer,j}$ is the traffic that is rerouted to gNB j and N_j is the number of vehicles whose offered load was totally served by this gNB. The term $N_{tot,j}$ includes both the number of vehicles being totally served by this gNB and the vehicles whose offered load was partially served, i.e. a part of the offered packets was served.

In the same way, when the *M*(*N*)/*M*/*c* model is applied, the total throughput at a specific time instance in the system is:

$$Throughput = \sum_{j \in V} v_j \tag{23}$$

The packets are being served by the gNBs under a specific policy: the gNBs serve the whole request until their resources are almost fully consumed. When the resources utilization reaches a predefined threshold as such that a vehicle's whole request cannot be served, only a proportion of the request can be served. This proportion of packets depends exclusively on the system's capacity limitation. On the other hand, regarding the remaining packets, two options are available: a) the "no rerouting case" where the packets are dropped since there is inactive any rerouting functionality to transmit traffic to gNB or b) "the rerouting case" where the packets are rerouted to other gNBs in order to be served.

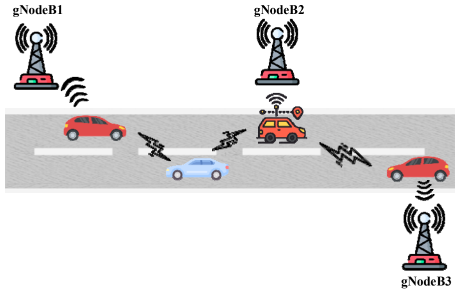

Fig. 2: System under study

1) "No rerouting" case

In the "no rerouting" scheme, a vehicle requires service directly from the destination gNB, which in our example is gNB2. It is assumed that gNB2's channels are idle when the first request for service arrives at the gNB2. Each gNB can serve packets until its resource utilization reaches a specific limit, which equals to 80% in our system, thus the maximum number of packets that can be served depends on the utilization of its resources at the moment a request arrives. If the system's utilization reach this limit, the system cannot guarantee to the users the agreed QoS and the packets are dropped since there isn't any rerouting option.

By solving (6) and (14), the maximum number of packets that can be served per gNB in both queueing models can be calculated. However, this value strongly depends on the adopted queueing model i.e. the *M*/*M*/*c* and *M*(*N*)/*M*/*c* in our study. Moreover, the total number of packets that can be served also depends on the number of the available servers. By utilizing (7) and (17), the minimum number of servers for the QoS satisfaction is calculated and it is equal to c = 6 in the *M*/*M*/*c* model and c = 2 in the *M*(*N*)/*M*/*c* model. However, this value does not guarantee that the number of users being served will not be affected. In more detail, as little is the number of available resources, so earlier the resources' utilization reaches its limit which means that the total number of served packets is decreased. In particular, when the *M*/*M*/*c* model is employed and the number of servers equals to c = 7, the maximum number of packets that can be served corresponds to the load of N = 6 vehicles and the 22% of the packets of a 7th vehicle. On the opposite, for c = 6 and for the same utilization limit, the maximum load that can be served corresponds to N = 5 and 33% of the packets of the 6th vehicle. In accordance, when the *M*(*N*)/*M*/*c* model is applied and the number of servers equals to c = 7, the maximum number of packets that can be served is increased compared to the *M*/*M*/*c* model and it is equal to N = 12. However, for c = 2 this number diminishes to N = 3 and roughly the 15.5% of a 4th vehicle, which seems reasonable due to the decreased number of servers. Regarding the number of packets being served at a specific time, it is calculated via (21) and (23) in the *M*/*M*/*c* and the *M*(*N*)/*M*/*c* models, respectively, and it is illustrated in Fig. 3 for the "no rerouting" scheme. In addition, the delay of a packet transmitted on this link is given by (1) where c_{ij} refers to the cost of using the link between the vehicle and the gNB2 and it is determined by (18) while N_j refers to the number of vehicles served by gNB2. Moreover, in case that the gNB2 serves only a part of the entire request of a vehicle, then it is assumed that N_j = 1. It is significant to clarify that the delay per vehicle must always be smaller than the application's QoS requirements (i.e. $\sum_{l} c_{ij} / N_j < t$), otherwise the packets are dropped. The delay for all the aforementioned schemes in the "no rerouting" case is presented in Fig.5.

2) "Rerouting" case

In the same manner with the "no rerouting" scheme, in the "rerouting" scheme the vehicles communicate directly with the gNB2 requesting for service. However, in contrast to the previous case, when the resources' utilization reaches its limit, the vehicle's request has the alternative to be rerouted to another gNB i.e. gNB1 and gNB3 in order to be served instead of being dropped. As it is illustrated in Fig.2, in our system there should exist at least 4 vehicles on the road during the rerouting process in order to maintain the necessary connections between the vehicles and the gNBs. In order to determine the most cost-effective route to transmit the packets, the SPP algorithm is utilized. As it is inferred from Fig.2, when the utilization limit of gNB2 is reached, packets transmitted from vehicles to gNB2 will be rerouted to gNB3 since the route to gNB3 is shorter compared to that to gNB1. In accordance, when gNB3 resources are also fully utilized, then the packets requesting for service from gNB2 will be rerouted to gNB1 in order to be served since it is the only available gNB. However, when the rerouting procedure begins, both gNB1 and gNB3 are already partially occupied by serving other vehicles. In particular, gNB1 serves 2 vehicles and gNB3 serves 3 vehicles. Therefore, when the *M*/*M*/*c* model is employed and the number of servers equals to $c = 6$, the gNB1 is capable of serving 3 additional vehicles and the 33% of the packets of a 4th vehicle. In accordance, when the *M*/*M*/*c* model is applied and $c = 7$, this number increases to $N = 4$ vehicles and the 22% of packets of a 5th vehicle. On the other hand, when the *M*(*N*)/*M*/*c* model is applied and the number of servers is equal to $c = 2$, the gNB1 can additionally serve 1 vehicle and the 15.5% of a 2nd vehicle while this number rises to $N = 10$ for $c = 7$. Apropos to the gNB3, when the *M*/*M*/*c* model is applied and $c = 6$ then it is capable of additionally serving 2 vehicles and the 33% of the packets of a 3rd vehicle while when $c = 7$, this number is adequate to serve $N = 3$ vehicles and the 22% of packets of a 4th vehicle. On the opposite, when the *M*(*N*)/*M*/*c* model is employed and $c = 2$ the gNB3 can additionally serve only the 15.5% of the packets of one vehicle while for $c = 7$ it is capable of serving the whole packet request of 9 vehicles. Similar to the "no rerouting" case, the total number of packets being served at a specific time is determined via (21) and (23) in the *M*/*M*/*c* and the *M*(*N*)/*M*/*c* models, respectively, and it is illustrated in Fig. 4. In the same way, the delay of a packet transmitted on this link is determined similarly to the "no rerouting" case and it is presented on Fig.6. It is significant to mention that the delay per vehicle must always be smaller than the application's QoS requirements (i.e. $\sum_{l} c_{ij} / N_j < t$).

IV. Numerical Results

The performance of the proposed V2X system presented in subsection III.C is illustrated in Figs. 3-6. In particular, a quantitative comparison between the *M*/*M*/*c* and the *M*(*N*)/*M*/*c* queueing models in terms of throughput and delay is presented for a) the "no rerouting" and b) the "rerouting" scheme. In the first set of results illustrated in Figs.3-4, we can observe the system's throughput in the cases where there do not exist or do exist rerouting alternatives, respectively. As it is evident, the number of served packets is increased when the rerouting procedure is activated. In addition, in the cases where the number of servers is equal to $c = 7$, the number of vehicles being served when the *M*(*N*)/*M*/*c* model is applied (i.e. $N = 12$ in the "no rerouting" case and $N = 31$ in the "rerouting" case) is significantly higher. This remark stands on the fact that, in contrast to the *M*/*M*/*c* model which considers Poisson arrivals, in the *M*(*N*)/*M*/*c* model packet arrivals are generated from a finite number of sources. Consequently, this feature makes the *M*(*N*)/*M*/*c* model more suitable for evaluating systems with a small number of vehicles-sources. However, in the cases where the minimum number of servers is utilized i.e. when $c = 6$ in the *M*/*M*/*c* and $c = 2$ in the *M*(*N*)/*M*/*c*, it is shown that in the *M*/*M*/*c* model a higher throughput is achieved both in the "no rerouting" and "rerouting" cases. The aforementioned observation is attributed to the fact that the utilization limit is reached quite early when $c = 2$. Consequently, when deploying the minimum required number of servers given by (7) and (17), it must be also taken into consideration that the number of vehicles being served is reduced so the system's throughput may also be severely affected.

In respect to the second set of results illustrated in Figs.5-6, we can observe the packet delay experienced in the system when either the rerouting option is available or not. From Fig. 5, it can be seen that in the "no rerouting" case the packet delay is higher when the *M/M/c* model is applied. As it is evident from (18), the packet delay depends on the total time *T* that a packet spends in the system which in general was observed to be smaller when the *M*(*N*)/*M*/*c* was employed. This occurs because the *M/M/c* model considers Poisson arrivals, so the system behaves as such it serves an infinite number of vehicles which does not apply in our system under study [9]. In addition, in both queueing models the packet delay experienced by the users was higher when the minimum number of servers were considered compared to the corresponding cases where $c = 7$. This tendency was also observed when the rerouting mechanism was activated as it is illustrated in Fig. 6. Therefore, as little is the number of servers in the system, so stressed the system's available resources are. Moreover, as it is depicted in Fig.6, in the cases where $c = 7$ it is observed that for the same number of vehicles, the delay observed in the *M*(*N*)/*M*/*c* model was significantly smaller compared to the *M/M/c* model which is justified by the fact that the value of *T* is smaller too. However, it is remarked that in the *M*(*N*)/*M*/*c* model the packet delay is highly affected by the link's cost which was not observed in the *M/M/c* model in a similar situation.

As a general remark, the utilization of the minimum possible resources to satisfy the QoS in terms of the application parameters β and t, does not guarantee that the system's overall performance will remain unaffected. In more detail, by employing the *M*(*N*)/*M*/*c* model for the performance evaluation of our system it is shown that it provides better results compared to the *M*/*M*/*c* model when the number of their resources is the same. However, when the minimum resources condition is adopted, the number of required resources in the *M*(*N*)/*M*/*c* model shows a rapid decrease compared to the corresponding value in the *M*/*M*/*c* model which severely affects the system's performance both in terms of throughput and delay. As a consequence, applying the minimum possible resources in a VANET may be beneficial to the extent of the implementation cost. However, this solution can be adopted only as long as the number of vehicles being served and consequently the system's overall performance is not severely affected.

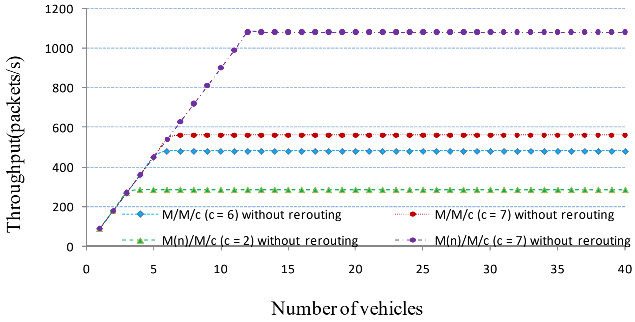

Fig. 3: System's throughput versus number of vehicles in the "no rerouting" case

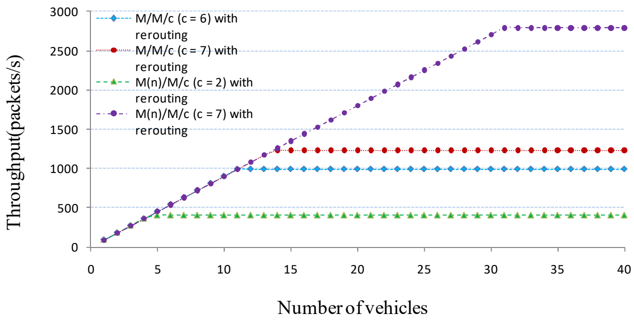

Fig. 4: System's throughput versus number of vehicles in the "rerouting" case

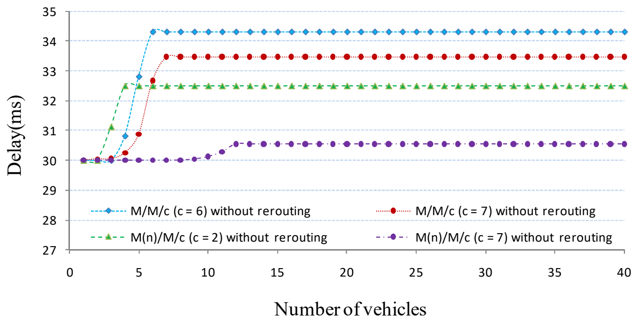

Fig. 5: System's delay versus number of vehicles in the "no rerouting" case

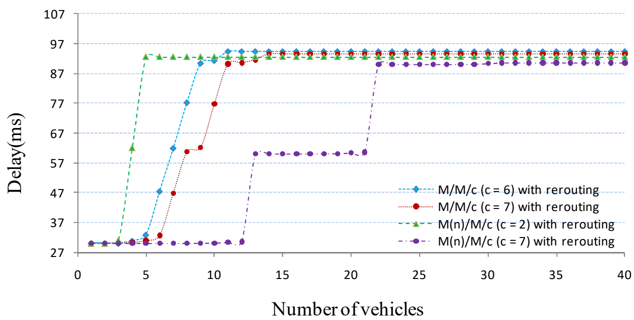

Fig. 6: System's delay versus number of vehicles in the "rerouting" case

V. Conclusion

In this work, we utilized two queueing theory models, namely the *M*/*M*/*c* and the *M*(*N*)/*M*/*c*, in order to estimate the required resources that satisfy the agreed QoS in a VANET. In our system, two types of communications are available: a) a vehicle transmits its packets to another vehicle inducing a V2V communication and b) a vehicle transmits its packets to a gNB forming V2I communication. A vehicle can also take advantage of the other vehicles on the road in order to indirectly transmit its request to a gNB. It is shown that the number of needed resources for satisfying the QoS requirements of an application strongly depends on the adopted queueing theory model. When the *M*(*N*)/*M*/*c* model is adopted, the number of required resources is significantly smaller compared to the corresponding *M*/*M*/*c* model. However, this outcome has a huge impact on the system's overall performance since the number of vehicles being served is rapidly reduced. The same outcome is also observed when the *M*/*M*/*c* model is employed but the performance degradation isn't as huge as that in the *M*(*N*)/*M*/*c* model. As a result, adopting the minimum possible resources for satisfying the application's requirements isn't the only factor that should be considered during the design process of a VANET. The minimum possible resources case should be adopted only when it is guaranteed that the system's overall performance will not be severely affected. As a future work, we intend to consider the application of multi-rate loss models in a VANET [10], [11].

Acknowledgment

The research work was supported by the Hellenic Foundation for Research and Innovation (HFRI) under the 3rd Call for HFRI PhD Fellowships (Fellowship Number: 6681).

References

[1] M. Z. Chowdhury, M. Shahjalal, S. Ahmed and Y. M. Jang, "6G wireless communication systems: applications, requirements, technologies, challenges, and research directions," in IEEE Open Journal of the Communications Society, vol. 1, pp. 957-975, 2020.

[2] S. Malik and P. K. Sahu, "A comparative study on routing protocols for VANETs", Heliyon, vol. 5, no. 8, 2019.

[3] M. H. C. Garcia et al., "A tutorial on 5G NR V2X communications," in IEEE Communications Surveys & Tutorials, vol. 23, no. 3, pp. 1972-2026, 2021.

[4] Z. MacHardy, A. Khan, K. Obana and S. Iwashina, "V2X access technologies: regulation, research, and remaining challenges," in IEEE Communications Surveys & Tutorials, vol. 20, no. 3, pp. 1858-1877, 2018.

[5] S. Fowler, C. Häll, D. Yuan, G. Baravdish and A. Mellouk, "Analysis of vehicular wireless channel communication via queueing theory model", Proc. IEEE ICC, Sydney, Australia, June 2014.

[6] K. Gill, K. Heath, R. Gegear, E. Ryder and. A. Wyglinski, "On the capacity bounds for bumblebee-inspired connected vehicle networks via queueing theory", Proc. IEEE 87th Veh. Tech. Conf. (VTC spring), Porto, Portugal, June 2018.

[7] I. Keramidi, P. Kardaras, I. Moscholios, P. Sarigiannidis and M. Logothetis, "'A study on the impact of service time distributions in a Vehicular Ad Hoc Network", Proc. Int. Mediterranean Conf. Commun. and Networking, Athens, Greece, Sept. 2021.

[8] J. Shortle, J. Thompson, D. Gross and C. Harris, Fundamentals of queueing theory, 5th edition, John Wiley, 2018.

[9] I. Keramidi, D. Uzunidis, I. Moscholios, P. Sarigiannidis and M. Logothetis, "On Queueing Models for the Performance Analysis of a Vehicular Ad Hoc Network," Proc. International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, Sept. 2022.

[10] M. Stasiak, M. Glabowski, A.Wisniewski, and P. Zwierzykowski, Modeling and Dimensioning of Mobile Networks, John Wiley, 2011.

[11] I. Moscholios and M. Logothetis, Efficient multirate teletraffic loss models beyond Erlang, John Wiley & IEEE Press, 2019.