

Communication and Computation Offloading for 5G V2X: Modeling and Optimization

Ren-Hung Hwang¹, Md. Muktadirul Islam², Md. Asif Tanvir², Md. Shohrab Hossain² and Ying-Dar Lin³,

¹Department of Computer Science and Information Engineering, National Chung Cheng University, Chia-Yi, Taiwan

²Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Bangladesh

³Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan

Email: rhhwang@cs.ccu.edu.tw, mu.islam34@gmail.com, asifbrur005@gmail.com,

mshohrabhossain@cse.buet.ac.bd, ydlin@cs.nctu.edu.tw

Abstract—Vehicle-to-everything (V2X) is one of the generic services provided by emerging 5G wireless technology. Data traffic in V2X demands both communication and computation which is a challenge to ensure better user experience where communication and computation offloading are an effective scheme. Earlier works have only focused on communication offloading policies and algorithms. However, there have been a few analytical studies, especially on computation offloading and the role of Road-Side Units (RSUs). In this work, we propose an analytical model for communication and computation offloading to RSUs and gNB to minimize average packet delay by finding an optimal offloading probability through a sub-gradient based algorithm. Extensive simulations have been carried out to validate our model. Our results show that deploying RSUs with gNB increases efficiency by as much as 19% than deploying only gNB. Sensitivity analysis also shows that RSU service rate has 10 times more impact on reducing average packet delay than gNB service rate caused by relatively lower communication bandwidth of gNB.

Index Terms—V2X, RSU, gNB, Computation and Communication Offloading, Queueing model.

I. INTRODUCTION

The number of smart cars is increasing day by day with the rapid development of wireless networks and 5G. In several 5G scenarios, the number of smart cars will increase up to 2.8 billion in the near future [1]. In this scenario, every vehicle is a communication node and the vehicle network will generate data of different dimensions [2], [3]. There has also been a rapid increase in the number of requests for various content services from vehicular users, resulting in higher data traffic in vehicular communication. To ensure QoS of data, offloading becomes necessary. Data offloading means the distribution of data over different paths. By offloading the traffic between various infrastructures and base stations, we can minimize the delay and provide a more effective service to all clients.

Fig. 1 shows a generic scenario of communication between vehicles, RSUs and gNB. RSUs generally have relatively low computation power, but are located close to the vehicles, so that the communication cost for the RSUs is relatively low. Again, base stations, which are also known as gNBs, have much higher computation power. However, the number of gNBs is fewer than RSUs and communication costs for these is usually quite high. Data offloading between RSUs

and gNBs is thus essential to improve the performance of data communication and computation.

Data offloading can be done in many ways. Recent works [1], [4] and [5] have shown how queueing models can be used for data offloading, but have not shown the trade-off between the computation and communication cost between RSUs and gNBs. There are several methods [6]–[9] which have described offloading methods using VANETs and provide several algorithms for offloading. [10], [11] have provided some proposals for 5G-based networks and how offloading can be done in such networks. But, none of these works has proposed a methodology for offloading data between RSUs and gNBs. To the best of our knowledge, this is the *first work* that develops an analytical model for offloading data between the RSUs and gNBs in a way that optimizes communication and computation delay.

As communication and computation costs of the RSUs and gNBs are different, we have derived a probability with which data can be offloaded either to a gNB or an RSU. We have developed an analytical model and a sub-gradient method-based algorithm which determines an optimal offloading probability that minimizes average packet delay. Queueing theory has been used for the analytical analysis.

The *contributions* of this work are:

- (i) developing an analytical model for RSUs and gNBs using a queueing model.
- (ii) developing a sub-gradient method-based algorithm to determine the optimal offloading probability for minimum average packet delays.
- (iii) performing extensive simulations to validate our analytical model.
- (iv) investigating the impact of different system parameters on overall performance.

The main *application* of our work is that it can guide operators to deploy RSUs with gNB in such a way as to optimize delay of V2X communication and computation. Our work can also help to decide whether to offload a packet to the RSU or to gNB.

The rest of the paper is organized as follows. In Section II, we briefly describe the previous works that have been done in this area, and in Section III, we present our analytical model.

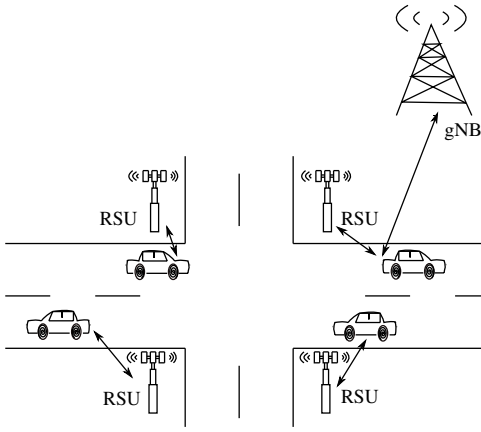


Fig. 1. V2X Communication and Offloading Scenario

In Section IV, we present analytical and simulation results, and Section V contains the concluding remarks.

II. BACKGROUND

There has been some work carried out on computation and communication offloading of V2X communication. Fowler et al. [4] developed a feasible analytical queueing model for determining the minimum number of channels needed for V2X communication. They showed how a queueing theory approach can provide a good estimation of costs and delays for routing problems. They used M/M/m queues and the First Come First Served (FCFS) queueing discipline. On the other hand, Wang et al. [5] modeled a traffic management system as a queueing system and regarded a cloudlet as a processing server based on an M/M/1 queue. They modeled parked and moving vehicles according to queueing theory. Using an offloading branch-and-bound algorithm they minimized the response time and delay of fog nodes. Hu et al. [1] also modeled the offloading optimization problem but as a minimum average total energy consumption problem by combining each member in platooning with an MEC server. Like [4], the offloading decision of whether the task is performed locally or on other platoon members is determined by queueing theory. But, in order to minimize average total energy consumption, an optimization algorithm based on Lyapunov was proposed.

There are also offloading methods other than queueing modeling. Zhou et al. [6] described various techniques of data offloading through VANETs under three categories: V2V, V2I, V2X. Otsuki and Miwa [9] mainly studied how to utilize V2X communications effectively by combining V2I and V2V communications to deliver content in real time. Their algorithm used route prediction information to enable vehicles to efficiently share contents through V2X communications. On the other hand, El Mouna Zhioua et al. [7] used ITSs to offload data using VANET capacity and V2V and I2V link capacity. They described the problem as a multi-constraints optimization problem that aimed to select a maximum target set of flows to offload the cellular data through VANETs. Malandrino et al. [8] also studied the effectiveness of using

VANETs to offload cellular data but for content downloading in presence of inaccurate mobility prediction and proposed that RSUs could decide which data they should fetch from the cellular network and to schedule transmission to vehicles. They proposed a fog-of-war model to predict the uncertainty of traffic manager.

There have been recent works on 5G based computations in V2X. Aujla et al. [10] proposed a 5G SDN-based data offloading method to offload 5G cellular data through VANETs. The authors designed an SDN-based controller that controls data offloading by using a priority manager and a load balancer. Ning et al. [11] described the method for computation offloading, power allocation and channel assignment scheme for 5G-enabled traffic management systems, with the purpose of maximizing the achievable sum rate. Instead of using a load balancer like [10], they used a position-based clustering scheme and for offloading decisions, used a utility function for each cluster head. In Table I, we have summarized some of these works stating their offloading problem, methodology, and performance metrics. Here, we have emphasized only the entities between which data is offloaded and the methods for solving this offloading decision.

TABLE I
RELATED WORKS ON DATA OFFLOADING IN V2X

| Paper | Offloading Problem | Solution | Performance metric |
|-----------|--|---|--------------------------|
| [5] | Fog node, Cloudlet in IoV | M/M/1 modeling, Branch-and-bound algorithm | Avg. response time |
| [1] | MEC server, Platoon members in 5G | M/M/k modeling, Lyapunov algorithm | Avg. energy consumption |
| [7] | RSU, eNodeB in LTE | I2V link estimation | Avg. data flow |
| [8] | RSU, Base station in LTE-Advanced | Fog-of-war model, Linear programming | Offloading efficiency |
| [9] | Vehicle, Content server in DTN | Approx. algo using route prediction | Avg. content transmitted |
| [10] | SDN controller, Priority manager in 5G | Load balancing, Priority manager | Avg. data offloaded |
| [11] | RSU, Macro-cell in 5G | Stochastic theory | Network offloaded ratio |
| This work | RSU, gNB in 5G | M/M/1 modeling, Sub-gradient search algorithm | Avg. packet delay |

However, none of these works describes a procedure of offloading data between RSUs and gNBs in such a way that the overall communication and computation costs are minimized. To the best of our knowledge, this is the first work that uses a queueing model for offloading 5G V2X data between RSUs and gNBs so as to minimize the data traffic between them.

III. SYSTEM MODEL

The system architecture of the V2X scenario consists of vehicles, RSUs, and gNB. In this scenario, we assume that

RSU and gNB have both computation and communication capacity and a gNB is associated with N RSUs. A data packet is generated at a vehicle and can be offloaded either to an RSU or a gNB. The packet will be offloaded to gNB with probability p_L . This is a probabilistic offloading strategy where a packet is offloaded to a computation resource based on a given probability.

We have used queueing theory to develop mathematical models for the above V2X scenario. The notations used in the analysis are listed in Table II. To denote different service rates, we have used superscript X , and subscript X and superscript Y to denote different delays. The queueing model is shown in Fig. 2 where a data packet can go to gNB with offloading probability, p_L . Hence, the offloading probability to go to RSU is $1 - p_L$. There are N RSUs in the system. Vehicles arrive and depart from the RSU coverage at rate λ_v and μ_v , respectively. Packet arrival rate from each vehicle is λ_d . When a packet is sent to the RSU or gNB, it is sent through uplink V2I or V2N, respectively. Depending on the bandwidth of the communication link, a communication delay is incurred. Similarly, after the computation is finished by RSU or gNB, the result is sent back through the downlink I2V or N2V, respectively.

TABLE II
NOTATIONS USED IN THE ANALYSIS

| Symbol | Parameter Name |
|---|---|
| λ_v | Vehicle arrival rate to an RSU |
| μ_v | Vehicle departure rate to an RSU |
| λ_d | Packet arrival rate from each vehicle |
| μ^X ($\mu^R, \mu^g, \mu^{V2I}, \mu^{I2V}, \mu^{V2N}, \mu^{N2V}$) | Service rate at an RSU, gNB or communication link where, X is either R (RSU), g (gNB), V2I (uplink from vehicle to RSU), I2V (downlink from RSU to vehicle), V2N (uplink from vehicle to gNB), N2V (downlink from gNB to vehicle) |
| p_L | Offloading Probability to send to gNB |
| D_X^Y | Average packet delay of Y for X where Y is either V2I, R, I2V, V2N, g, N2V and X is either RSU (packet offloaded to RSU) or gNB (packet offloaded to gNB) |
| D_X | Average packet delay for X where X is either RSU, gNB, or RSU+gNB (overall average packet delay with offloading probability p_L) |
| N | Number of RSUs under one gNB |

A. Assumptions

The following are the assumptions of the model:

- Data arrival process from vehicle to RSU link and from vehicle to gNB link is a Poisson process;
- The service time of packets is assumed to follow exponential distributions;
- The queue size is assumed to be infinite, and
- This scenario is assumed to be homogeneous, where all the RSUs have the same vehicle arrival and departure

rates. RSUs have same service rate and communication bandwidth.

For the analytical model, we consider each RSU, gNB and communication link as an M/M/1 queue. We derive the average packet delay experienced by a data packet in two cases: the packet is served either by the RSU or the gNB. We derive the average packet delay with offloading probability p_L by combining the two cases.

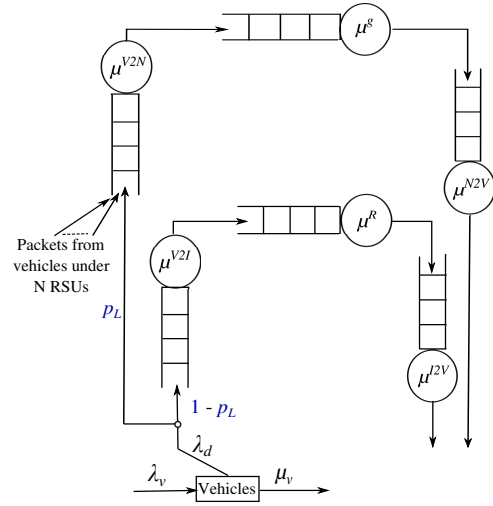


Fig. 2. Queueing model of V2X offloading

B. Case 1: The packet is served by RSU

In this case, we will calculate the average delay of a packet served by an RSU. For the average packet delay, we will calculate the communication (uplink and downlink) and computation delays experienced by a packet. Assuming that the arrival of vehicles follows a Poisson process with rate, λ_v and using vehicle departure rate, μ_v we can obtain the average number of vehicles in the network, which is $\frac{\lambda_v}{\mu_v - \lambda_v}$. As the packet arrival rate from each vehicle is λ_d , we can obtain the total arrival rate of packet to link queue V2I as $\frac{\lambda_v \lambda_d (1 - p_L)}{\mu_v - \lambda_v}$ where $(1 - p_L)$ is the offloading probability of a packet sent to RSU. As these packets are served at RSU and sent back through communication downlink I2V, the packet arrival rate of RSU and I2V queue is similar to the arrival rate of the V2I queue. Therefore, the average packet delay at V2I uplink queue can be expressed as

$$D_{RSU}^{V2I} = \frac{1}{\mu^{V2I} - \frac{\lambda_v \lambda_d (1 - p_L)}{(\mu_v - \lambda_v)}} \quad (1)$$

Similarly, we can derive the average packet delay at RSU and I2V communication downlink using their respective service rates. Therefore, the average packet delay at RSU is

$$D_{RSU}^R = \frac{1}{\mu^R - \frac{\lambda_v \lambda_d (1 - p_L)}{(\mu_v - \lambda_v)}} \quad (2)$$

and average packet delay at I2V downlink queue is

$$D_{RSU}^{I2V} = \frac{1}{\mu^{I2V} - \frac{\lambda_v \lambda_d (1 - p_L)}{(\mu_v - \lambda_v)}} \quad (3)$$

In this case, each packet has to visit V2I, RSU and I2V queue once. We can therefore calculate the average packet delay for only the RSU case from Equations (1), (2) and (3) as

$$D_{RSU} = D_{RSU}^{V2I} + D_{RSU}^R + D_{RSU}^{I2V}. \quad (4)$$

C. Case 2: The packet is served by gNB

The calculation of the average packet delay is similar to Case 1. There can be N RSUs under one gNB. As we have assumed this as a homogeneous scenario, packets will arrive at gNB from vehicles under N RSUs at the same rate. The packet arrival rate from vehicles under one RSU to the gNB communication uplink queue V2N is $\frac{\lambda_v \lambda_d p_L}{\mu_v - \lambda_v}$, where p_L is the offloading probability of a packet going to gNB. The packet arrival rate at V2N queue is then $\frac{\lambda_v \lambda_d p_L N}{\mu_v - \lambda_v}$. This packet arrival rate will be the same for gNB as well as the N2V communication downlink queue, because packets will be sent to gNB and N2V queues after the V2N queue. We can calculate the average packet delays by using the service capacity of each queue. The average packet delay at V2N queue is

$$D_{gNB}^{V2N} = \frac{1}{\mu^{V2N} - \frac{\lambda_v \lambda_d p_L N}{\mu_v - \lambda_v}}. \quad (5)$$

Similarly, the average packet delay at gNB queue with service rate μ^g is

$$D_{gNB}^g = \frac{1}{\mu^g - \frac{\lambda_v \lambda_d p_L N}{\mu_v - \lambda_v}} \quad (6)$$

and the average packet delay at N2V queue is

$$D_{gNB}^{N2V} = \frac{1}{\mu^{N2V} - \frac{\lambda_v \lambda_d p_L N}{\mu_v - \lambda_v}}. \quad (7)$$

Using the individual queue delays of V2N, gNB and N2V queue, we can calculate the average packet delay for Case 2 as

$$D_{gNB} = D_{gNB}^{V2N} + D_{gNB}^g + D_{gNB}^{N2V}. \quad (8)$$

D. Average packet delay with offloading probability p_L

We can obtain the average packet delay by combining the derived equations of Case 1 and Case 2, which can then be expressed as

$$D_{RSU+gNB} = (1 - p_L)D_{RSU} + p_L D_{gNB}. \quad (9)$$

E. Sub-gradient Search Algorithm

A subgradient method is an iterative method for solving a convex minimization problem. As we see the convexity of the objective function, we can find the optimal p_L so that the overall delay, $D_{RSU+gNB}$ is minimized. We developed a sub-gradient search algorithm similar to [13] to find out the optimal p_L as shown in Algorithm 1. In [13], it optimizes objective function for two variables but this algorithm optimizes only one variable, p_L . We adopted nonsummable diminishing step size for achieving convergence.

The algorithm first initializes a step size and two different p_L s. For each iteration, average packet delay is calculated for

each p_L by using Equation (9). p_L is set towards the lower delay by using the step size. At the end of each iteration, step size is recalculated. This process is repeated until step size and difference between delays are below than respective threshold.

Algorithm 1 Sub-Gradient Search Algorithm

Input: $N, \lambda_v, \mu_v, \lambda_d, \mu^R, \mu^g, \mu^{V2I}, \mu^{I2V}, \mu^{V2N}, \mu^{N2V}$

Output: optimal p_L

```

1:  $p_L = 0.5$ ; //initialize
2: step = 0.25; k = 1; //iteration count
3:  $new\_p_L = p_L + step$ ;
4: while step > 0.001 do
5:   old_delay = Calculate average packet delay using Equation (9)
   for  $p_L$ ;
6:   new_delay = Calculate average packet delay using Equa-
   tion (9) for  $new\_p_L$ ;
7:   if |new_delay - old_delay| <  $10^{-15}$  then
8:     break; //exit loop
9:   else if new_delay > old_delay then
10:    if  $new\_p_L > p_L$  then
11:       $new\_p_L = p_L - step$ ;
12:    if  $new\_p_L < 0$  then  $new\_p_L = 0.0$ ;
13:    else
14:       $new\_p_L = p_L + step$ ;
15:    if  $new\_p_L > 1$  then  $new\_p_L = 1.0$ ;
16:    end if
17:  else
18:     $p_L\_old = p_L$ ;  $p_L = new\_p_L$ ;
19:    if  $new\_p_L < p_L\_old$  then
20:       $new\_p_L = p_L - step$ ;
21:    if  $new\_p_L < 0$  then  $new\_p_L = 0.0$ ;
22:    else
23:       $new\_p_L = p_L + step$ ;
24:    if  $new\_p_L > 1$  then  $new\_p_L = 1.0$ ;
25:    end if
26:  end if
27:  k = k + 1;
28:  step =  $0.25/\sqrt{k}$ ; //nonsummable diminishing step size
29: end while

```

IV. RESULTS

In order to validate our analytical methodology, we carried out extensive simulations. We used Ciw [14], which is a discrete event simulation library, for open queueing networks. For each simulation run, the time was set at 180 sec, with data being recorded after 10 sec had lapsed and was stopped for the last 10 sec. For each measurement of interest, the simulation value was taken as the average value obtained over 1000 simulation runs.

The baseline parameters used for the analysis and simulation are listed in Table III. Vehicle arrival rate, λ_v is used in a similar way to which it was used in [12]. Departure rates were set greater than the arrival rates as per M/M/1 queue requirement. We set the packet arrival rate of each vehicle, λ_d similar to that in [15]. We found service rates up to 1000 Mbps and set the service rates of different queues according to 3GPP standards [16], choosing rates that did not violate QoS requirements. For converting Mbps to packets/sec, we used the packet size according to [12].

TABLE III
BASELINE PARAMETERS FOR THE ANALYSIS AND SIMULATION

| Parameter | Value | Parameter | Value |
|-------------|-------------------|-------------|-------------------|
| λ_v | 2 vehicles/sec | μ_v | 3 vehicles/sec |
| λ_d | 600 packets/sec | μ^R | 10000 packets/sec |
| μ^g | 70000 packets/sec | μ^{V2I} | 50000 packets/sec |
| μ^{I2V} | 50000 packets/sec | μ^{V2N} | 20000 packets/sec |
| μ^{N2V} | 20000 packets/sec | N | 5 |

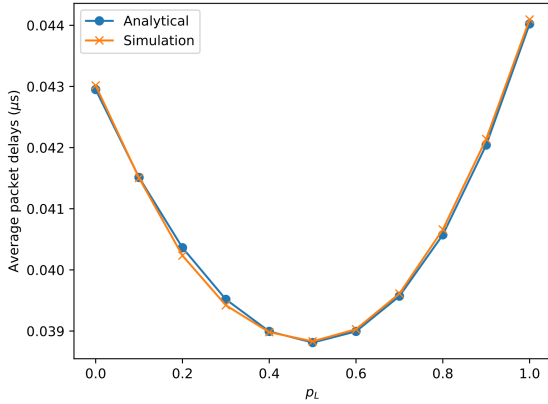


Fig. 3. Average packet delays vs. offloading probability, p_L .

Fig. 3 validates our analytical results, which closely match the simulation results. This shows that queueing theory can successfully estimate the average packet delay of the V2X scenario. It also shows that the optimal p_L for a minimum average packet delay is 0.5. Our sub-gradient search algorithm yielded the optimal p_L of 0.502. These results confirm that the sub-gradient search algorithm can determine an optimal p_L .

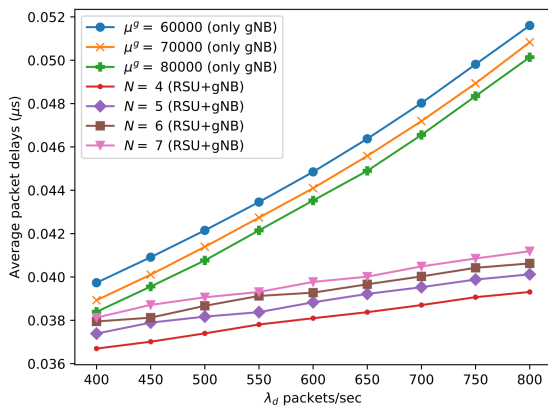


Fig. 4. Average packet delays vs. λ_d on different μ^g and N

A. Only gNB vs. RSU+gNB

We wished to ascertain whether deploying RSUs with gNB performed better than deploying only gNB, and also whether one could add additional RSUs to a single gNB. The results of the comparison between an only gNB and RSU+gNB V2X scenario is shown in Fig. 4. In the case of RSUs+gNB, we varied the number of RSUs (N) while keeping the value of μ^g constant at 70000 packets/sec. We found that the average packet delay increased with an increase in λ_d . However, we also found that a higher value of a gNB service rate (μ^g) did not significantly reduce the average packet delay. The slope of the gNB-only case is approximately three times higher than that of RSU+gNB, because the communication delay for gNB is relatively greater. Although the computation capacity of gNB is greater, the communication delay increases the overall delay. As RSUs have a relatively lower communication delay, and our sub-gradient search algorithm makes an optimal choice of p_L , the delay experienced with RSU+gNB is significantly lower and optimized. This shows that deployment of RSUs with gNB can boost the performance much greater than deploying only gNB.

We also observed that a greater number of RSUs increased the average packet delay, because a higher value of N increased the data arrival rate at gNB. Interestingly, the slope of the average packet delay is relatively low because our sub-gradient search algorithm had chosen the optimal p_L to minimize the delay. Therefore, more RSUs can be added while keeping the average packet delay below the QoS requirements.

B. Effect of μ^g vs. μ^R

We next wished to determine how RSU and gNB service rates affected average packet delays and whether an increase of RSU or gNB service rates improved performance. Fig. 5 shows the comparison of different RSU service rates and gNB service rates. To determine average packet delays, the optimal p_L was calculated by using the sub-gradient search algorithm. Intuitively, a higher packet arrival rate from each vehicle, λ_d would increase the average packet delay. Increasing the service rate of RSU by 10% reduced the average packet delay by approximately 16%. On the other hand, increasing the service rate of gNB by 10% reduced the average packet delay only by approximately 1.6%. This shows that a higher service rate of RSU could provide a 10-times greater reduction in delay than that of gNB. As the communication delay of RSU was already lower than gNB, increasing the service rate of the RSU decreased the computation delay, resulting in overall delay reduction. A higher gNB service rate also decreased the computation delay, but the overall delay did not decrease as significantly as RSU because of much higher communication delay. It is therefore much better to increase the service rate of RSU to obtain better performance.

C. Effect of λ_v and μ_v

Fig. 6 shows the impact of vehicle density on optimal p_L and when more packets are sent to RSUs by comparing different vehicle arrival (λ_v) and departure rates (μ_v). We

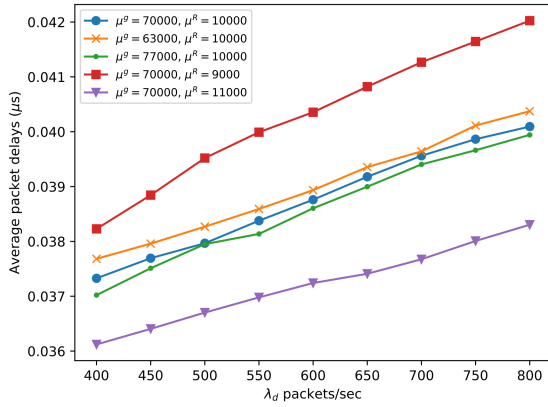


Fig. 5. Average packet delays vs. λ_d on different μ^g and μ^R

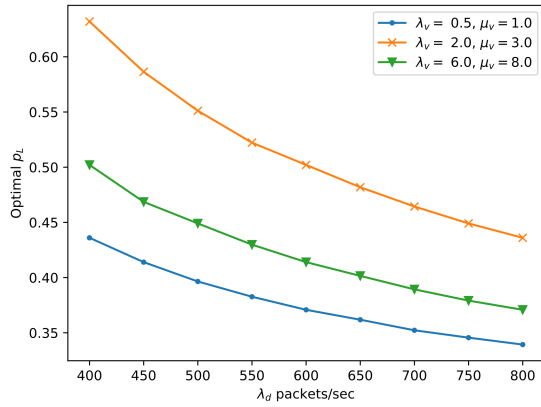


Fig. 6. Optimal offloading probability, p_L vs. λ_d on different λ_v and μ_v

considered three scenarios where the vehicle arrival and departure rates were high, low and similar to baseline values. For each scenario, the packet arrival rate from each vehicle was varied and the optimal p_L plotted. It is found that the optimal p_L tended to decrease with the increase in overall packet arrival rate. It was also found that the optimal p_L was relatively lower when vehicle's arrival and departure rates were higher for the same packet arrival rate from each vehicle. As the average number of vehicles within an RSU coverage increased, the packet arrival rate from all the vehicles to RSUs or gNB also increased. Packets experienced relatively higher uplink and downlink delays at gNB at lower communication bandwidths. For this reason, fewer packets were sent to gNB to optimize delays and optimal p_L decreased with an increase in the density of vehicles.

V. CONCLUSION AND FUTURE WORKS

In this paper, we have presented an analytical model for communication and computation offloading of 5G V2X using an M/M/1 model. We developed a sub-gradient method-based algorithm to determine the optimal offloading probability that

would minimize the average packet delay. Extensive simulations have been conducted to verify our analysis. Analytical results very closely match the simulation results, verifying the validity of our model.

Our results yielded the following findings.

- Deploying RSUs with gNB can increase efficiency by as much as 19%.
- The RSU service rate has a greater impact on improving performance than the gNB service rate.
- The optimal offloading probability decreased with the increase in packet arrival rate and vehicle density.

This modeling can be extended to minimize the probability of average packet delay greater than QoS requirements. This scenario will be investigated in the future.

REFERENCES

- [1] Y. Hu, T. Cui, X. Huang, and Q. Chen, "Task Offloading Based on Lyapunov Optimization for MEC-assisted Platooning," in *11th International Conference on Wireless Communications and Signal Processing (WCSP)*, Xian, China, Oct. 2019.
- [2] F. Ding, Z. Ma, Z. Li, R. Su, D. Zhang, and H. Zhu, "A terminal-oriented distributed traffic flow splitting strategy for multi-service of V2X networks", *Electronics*, vol. 8, no. 6, pp. 1-16, 2019.
- [3] H. H. R. Sherazi, Z. A. Khan, R. Iqbal, S. Rizwan, M. A. Imran, and K. Awan, "A heterogeneous IoV architecture for data forwarding in vehicle to infrastructure communication," *Mobile Information Systems*, 2019.
- [4] S. Fowler, C. Haan, D. Yuan, G. Baradish, and A. Mellouk, "Analysis of vehicular wireless channel communication via queueing theory model," in *IEEE International Conference on Communications*, Sydney, Australia, pp. 1736-1741, 2014.
- [5] X. Wang, Z. Ning, and L. Wang, "Offloading in Internet of vehicles: A fog enabled real-time traffic management system," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4568-4578, Oct. 2018.
- [6] H. Zhou, H. Wang, X. Chen, X. Li, S. Xu, "Data offloading techniques through vehicular ad hoc networks: a survey," *IEEE Access*, vol. 6, no. 1, pp. 65250 - 65259, Nov. 2018.
- [7] G. el Mouna Zhioua, H. Labiod, N. Tabbane, and S. Tabbane, "Cellular content download through a vehicular network: I2V link estimation," in *IEEE Vehicular Technology Conference*, Glasgow, Scotland, May 2015.
- [8] F. Malandrino, C. Casetti, C.-F. Chiasserini, and M. Fiore, "Content download in vehicular networks in presence of noisy mobility prediction," *IEEE Transaction of Mobile Computing*, vol. 13, no. 5, May 2014.
- [9] S. Otsuki and H. Miwa, "Contents delivery method using route prediction in traffic offloading by V2X," in *International Conference Intelligent Network Collaborative Systems*, Taipei, Taiwan, Sep. 2015.
- [10] G. S. Aujla, R. Chaudhary, N. Kumar, J. J. P. C. Rodrigues, and A. Vinel, "Data offloading in 5G-enabled software-defined vehicular networks: A Stackelberg-game-based approach," *IEEE Communication Magazine*, vol. 55, no. 8, Aug. 2017.
- [11] Z. Ning, X. Wang, J. J. P. C. Rodrigues, F. Xia, "Joint computation Offloading, power allocation, and channel assignment for 5G-enabled traffic management systems," *IEEE Transaction Industrial Informatics*, vol. 15, no. 5, May 2019.
- [12] H. D. R. Albonda and J. Pérez-Romero, "An efficient RAN slicing strategy for a heterogeneous network With eMBB and V2X services," in *IEEE Access*, vol. 7, 2019.
- [13] Ren-Hung Hwang, Yung-Cheng Lai, and Ying-Dar Lin, "Offloading Optimization with Delay Distribution in the 3-tier Federated Cloud, Edge, and Fog Systems," in *IEEE International Symposium on Cloud and Services Computing (IEEE SC2)*, Kao-Hsiung, Taiwan, Nov. 18-21, 2019.
- [14] Palmer G. I., Knight V. A., Harper P. R., Hawa, A. L., "Ciw: An open-source discrete event simulation library," *Journal of Simulation*, vol. 13, no.1, pp. 68-82, 2018.
- [15] M. Boban, A. Kousaridas, K. Manolakis, J. Eichinger and W. Xu, "Connected roads of the future: use cases, requirements, and design considerations for Vehicle-to-Everything communications," in *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 110-123, Sept. 2018.
- [16] 3GPP TS 22.186 V15.3.0: *5G; Service requirements for enhanced V2X scenarios (Release 15)*, 3GPP Std., ETSI, Sept. 2018.