Object Detection Performance According to Packet Loss of C-V2X

Eunmin Choi, Dongyoon Kwon, Kyungtae Kim, and Ji-Woong Choi
Dept. of Electrical Engineering and Computer Science
Daegu Gyungbuk Institute of Science & Technology
Daegu, South Korea
{eunminchoi, kwondy0131, knight27.kim, jwchoi}@dgist.ac.kr

***Abstract*—Studies on vehicle-to-everything (V2X) and edge computing to provide autonomous driving services mainly focus on low latency. However, in order to provide driving safety, it is also important to evaluate the performance considering the reliability of communication. In this paper, we evaluate the object detection performance according to the packet loss of V2X communication when transmitting video frames in consideration of the environment where camera sensor data is transmitted from the vehicle to the edge computer to recognize and judge.**

***Keywords*—Vehicle-to-everything, packet error rate, video data transmission, deep learning, object detection, detection performance**

I. Introduction

Research on key technologies of sensing, computation, control, and communication is being actively conducted in order to give advanced recognition, judgment, and control functions to autonomous vehicles. Vehicle-to-everything (V2X) communication technologies are focused on developing high-speed data transmission, low latency, and high reliability, and computing technologies are focused on developing efficient computing power and low processing time. Moreover, integrated research to provide autonomous driving services with the help of high-performance computing resources such as edge and cloud via V2X communication is also active [1]–[3]. Research on edge computing with V2X is mainly focused on providing low latency [1], [4], [5]. However, it is also important to analyze performance in terms of communication reliability. Packet loss may occur due to wireless communication when sensor data is transmitted to an edge computer through V2X to provide a sensor-sharing assisted driving service. As a result, recognition and judgment performance may suffer. When sending 3D point cloud data from a LiDAR sensor via a vehicle ad-hoc network (VANET), the effect of packet loss on object detection performance was investigated in paper [6]. The authors assume that packet loss has a minor influence on the visualization of 3D point cloud data where packet loss affects 3D point cloud data uniformly. In this paper, We reflect on the packet loss pattern of V2X communication to video frames using a system-level simulator and evaluate the object detection performance according to packet loss using Berkeley DeepDrive (BDD) dataset and you only look at once version 5 medium (Yolov5m) learning model [7]–[9]. The main contribution is that the object detection performance is evaluated considering packet loss of V2X communication. Based on the results, we provide the required V2X communication distance, modulation and coding scheme (MCS), and packet delivery rate (PDR) to reliably provide object detection performance. The abbreviations appeared in this paper are given in Table I.

TABLE I
List of abbreviations

Abbreviations	Definitions
V2X	Vehicle-to-everything
BDD	Berkeley DeepDrive
Yolo	You only look at once
Yolov5m	Yolo version 5 medium
BS	Base station
RSU	Road side unit
C-V2X	Cellular V2X
3GPP	3rd Generation Partnership Project
OFDM	Orthogonal frequency division multiplexing
PHY	Physical layer
SC-FDMA	Single carrier frequency division multiple access
MAC	Medium access control
TTI	Transmission time interval
DMRS	Demodulation reference symbols
ITS	Intelligent transport system
LTE	Long-term evolution
RB	Resource block
PRB	Physical resource block
MCS	Modulation and coding scheme
eNodeB	Evolved node B
R-CNN	Region-based convolutional neural network
RoI	Region of interest
SSD	Single-shot multi-box detector
AP	Average precision
mAP	Mean average precision
IoU	Intersection over union
AWGN	Additive white Gaussian noise
PDR	Packet delivery ratio
QAM	Quadrature amplitude modulation

II. Background

A. Scenario

Fig. 1 shows the sensor-sharing assisted driving service scenario. Through V2X communication, the vehicle transmits sensed data to an edge node with high processing capability, such as a base station (BS) or RSU, which performs perception and planning for driving. Autonomous driving is achieved by sending control commands back to the vehicle via V2X. We

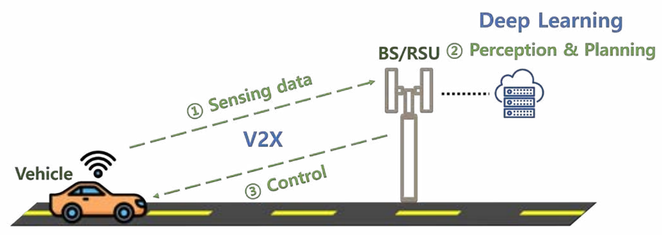


Fig. 1. Sensor-sharing assisted driving service scenario.

assume that the camera sensor data is transmitted to the RSU in this scenario, and we evaluate the effect of V2X packet loss on object detection accuracy.

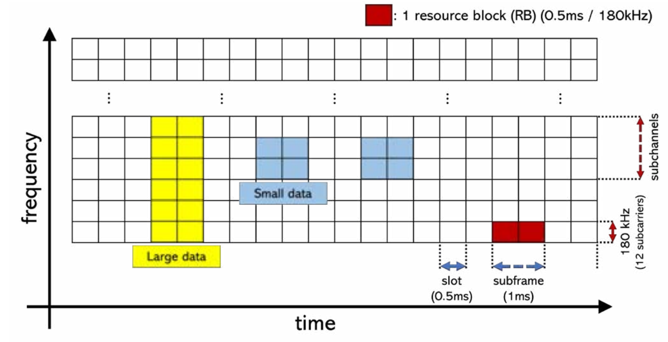


Fig. 2. Time-frequency resource grid for C-V2X.

B. V2X

LTE-V2X sidelink (C-V2X for short) defined by 3GPP Release 14 utilizes the orthogonal frequency division multiplexing (OFDM) at the phsycal (PHY) layer and single carrier frequency division multiple access (SC-FDMA) at the medium access control (MAC) layer, which is identical as the uplink PHY and MAC of the legacy LTE. Therefore, a time-frequency resource grid of Fig. 2 is used for resource allocation to transmit a packet with multiplexing.

In the time domain, the resource grid is divided in subframe units of 1 ms, which is the transmission time interval (TTI). One subframe is divided into two slots of 0.5 ms. In one subframe, 14 OFDM symbols are transmitted. There are 9 data symbols, 4 demodulation reference symbols (DMRS), and 1 guard band symbol, where the number of data symbols is less than the legacy LTE having 2 DMRS. This is due to adapt for high Doppler spread of the dynamically moving vehicle.

C-V2X utilizes 10 MHz or 20 MHz bandwidth at 5.9 GHz of intelligent transport system (ITS) band. In the frequency domain, the signal is divided into multiple subcarriers. The 12 subcarriers of each 15 kHz and one slot of 0.5 ms configures the resource block (RB). The two RB (RB pair) in one subframe of 1 ms, called a physical resource block (PRB), is the minimum unit for resource allocation. In case of 10 MHz bandwidth, there are 50 RBs in one slot (50 PRBs in one subframe). The group of RBs in one subframe is a subchannel. When the C-V2X vehicle transmits a packet, the vehicle first derives the number of RBs it should occupy per 1 subframe. Here, the number of required RBs is determined by the packet size and the MCS. As the MCS increases and the packet size decreases, the number of required RBs also decreases. Then, the vehicle allocates the subchannel configured as a group of RBs. In C-V2X, there are two methods for resource allocation. In mode 3, the base station (eNodeB) allocates the resource to the vehicle inside its coverage, while the vehicle autonomously allocates the resource in mode 4.

C. Object detection algorithm

The popular object detection algorithms for autonomous driving are classified into two types, two-stage object detection and one-stage object detection algorithms, according to the objection localization and object classification learning method. Two-stage object detection algorithm classifies objects after localization. There are faster region-based convolutional neural networks (R-CNN) and mask R-CNN. Faster R-CNN was proposed to improve the processing speed and accuracy of R-CNN and fast R-CNN. To extract the region of interest (RoI), a region proposal network, a learning layer, is added instead of the existing selective search to allow for end-to-end learning [10]. Mask R-CNN uses the RoIAlign pooling layer to map the object area more accurately, making it possible to recognize even smaller objects [11]. One-stage object detection algorithm performs object location and classification at once, and there are single-shot multi-box detector (SSD) and you only look at once (Yolo) algorithms. Yolo has been proposed to detect objects in real-time [12]. By dividing the image into grids, Yolo learns the confidence of the bounding boxes, which are the area where objects are likely, and the classes of objects. By multiplying the confidence of the bounding boxes by the probability of classes, the object location and classification are performed simultaneously. Since the bounding box is generated based on a grid, it is difficult to detect small objects. SSD has been proposed to detect objects of different sizes efficiently, and for this, they use multi-scale feature maps [13]. This paper uses the Yolov5m model for real-time object detection. Since Yolo is developed up to version 5, it has been improved to detect even small objects. We use the Yolov5m model, one of the most recent versions of the Yolov5 models, which has medium accuracy, processing speed, and complexity among Yolov5 models.

To evaluate the object detection performance, we use the mean average precision (mAP) when the intersection over union (IoU) threshold is 0.5, i.e., mAP@.5. The IoU is the area of overlapped region between prediction and ground truth region over the area of union. The detection criterion is whether the IoU exceeds a threshold. Based on IoU, precision and recall are calculated. The precision is the proportion of correctly detected results out of all detected results. The recall is the proportion of correctly detected results out of all ground truths. The precision-recall curve is obtained using the precision and recall, and the average precision (AP) is the area under precision-recall curve. The AP of each class can be calculated, and the mean of APs of all classes is mAP.

III. System model and simulation results

As explained in Section II-A, the required number of RBs for packet transmission is determined by MCS and packet size. In our paper, the sample driving video requires data rate of 5 Mbps. Fig. 3 shows the maximum data rate of C-V2X with the bandwidth of 10 MHz depending on the MCS variation. The maximum data rate is calculated assuming that the transmitting vehicle monopolizes the resource grid of Fig. 2 (i.e., a vehicle transmits a packet using all RBs in one subframe with a frequency of 1 kHz). According to Fig. 3, MCS must be at least 7 to satisfy 5 Mbps. Therefore, in this paper, MCS is controlled from 7 to 20. The remaining parameters for simulation are shown in Table II.

We use LTEV2Vsim, a system-level C-V2X simulation by MATLAB, and Yolov5m open-source and BDD dataset for object detection [7]–[9]. Our simulation environment is as follows. We set two nodes (one RSU and one vehicle) in the road, as shown in Fig. 1. The vehicle transmits a video streaming data at 1 kHz frequency to the RSU utilizing C-V2X, where the channel is additive white Gaussian noise (AWGN) with shadowing. It is assumed that the resource of the vehicle is allocated by the eNodeB as in the mode 3 method. Note that because we evaluate the object detection performance according to packet loss, the control data sent back from RSU to the vehicle does not consider in this simulation environment.

Fig. 4 shows the ratio of the subchannel size that must be occupied to satisfy the data rate of 5 Mbps compared to the total bandwidth of 9 MHz (the guard-band of 1 MHz is excluded) depending on the MCS variation. As shown, the ratio decreases as the MCS increases. This is because the modulation increases from QPSK (MCS 7∼10) to 16QAM (MCS 11∼20) as the MCS number increases, and the coding rate gradually increases in each modulation. Note that the higher MCS, the lower the subchannel occupancy rate, allowing more users.

Fig. 5a shows the packet delivery ratio (PDR) depending on the distance between the RSU and the vehicle. The PDR decreases as the distance and MCS increase. Fig. 5b shows the distances according to MCS that satisfy PDR of 0.95, 0.99, and 0.999 based on Fig. 5a. As the MCS increases, the distance to maintain the PDR decreases.

The object detection performance is evaluated when the 5 Mbps of video frames are received at each MCS and distance pair that satisfies the PDR of Fig. 5b. Fig. 6 represents the average mAP of the target PDR. The average mAP is 0.128 while transmitting at MCS and distance pairs corresponding to a PDR of 0.95. The average mAP rises as the PDR increases. Fig. 7 shows the received video frame according to the PDR. When the PDR is 0.999, it is visually similar to the original frame. The impairments such as color change and image distortion become more severe as the PDR drops to 0.99 and 0.95. As a result, the object detection performance is degraded.

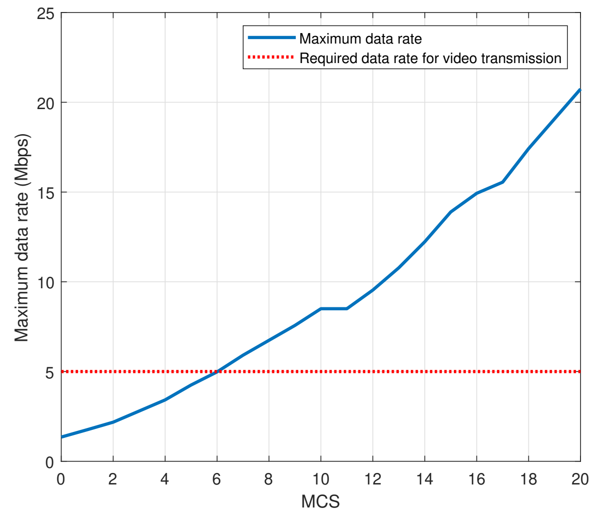


Fig. 3. Maximum data rate against MCS.

TABLE II
Parameters for simulation

Parameter	Value
Dataset	BDD dataset
# of training dataset	70,000
# of test dataset	10,000
Bit rate of video data	5 Mbps
Object detection model	Yolov5m
Channel model	AWGN channel
Shadowing	3 dB
Antenna gain	3 dB
Transmitter power	23 dBm
Transmit frequency	1 kHz
MCS	[7∼20]
Bandwidth	10 MHz

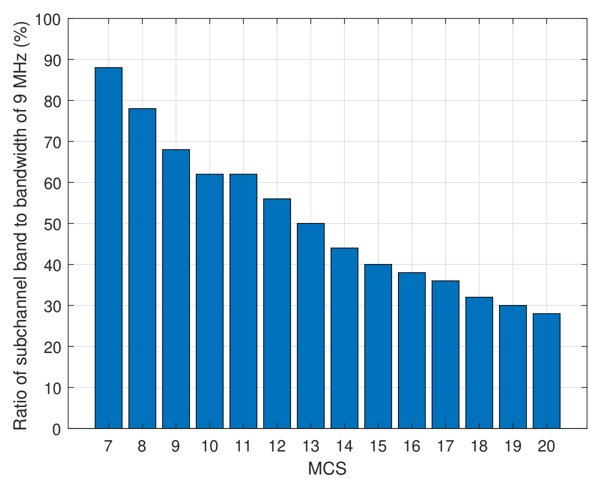
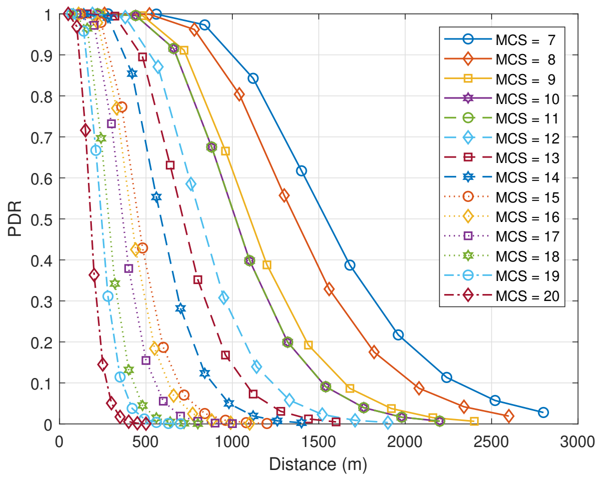
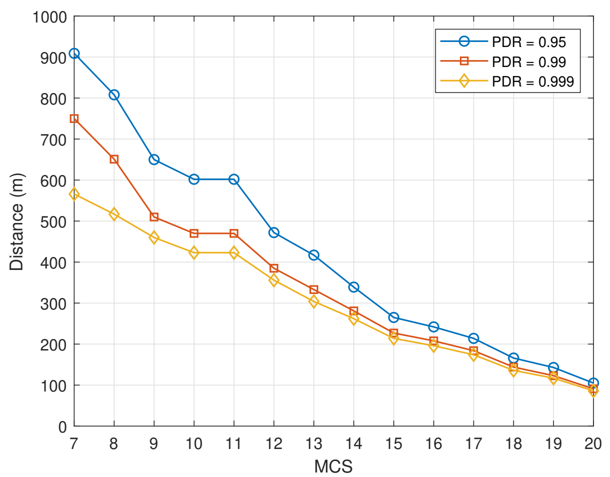


Fig. 4. Ratio of subchannel band to 9 MHz bandwidth depending on MCS.

According to the simulation results, when the PDR decreases to 0.95, the object detection performance (mAP)



(a)



(b)

Fig. 5. (a) PDR according to distance. (b) Distance according to MCS.

suffers greatly. Furthermore, in order to maintain a high PDR for mAP performance, it is necessary to select an appropriate MCS and distance. If the required distance to provide service is as long as 500 m, MCS 7 should be used to transmit 5 Mbps of video data. In this case, the subchannel occupancy rate of low MCS is high, limiting the utilization of multiple users. If the service can be supplied at a distance of 300 m, MCS can be used up to 13 which reduces the subchannel occupancy rate, allowing C-V2X service to be provided to more users.

IV. Conclusion

This paper evaluates object detection performance in the terms of C-V2X packet loss when sending video frames. The BDD dataset and the Yolov5m deep learning model are used. For object detection, a PDR of 0.99 or higher is required, and a lower PDR rapidly degrades object detection performance. The PDR depends on the MCS and the service distance between the vehicle and RSU. The MCS has an impact on

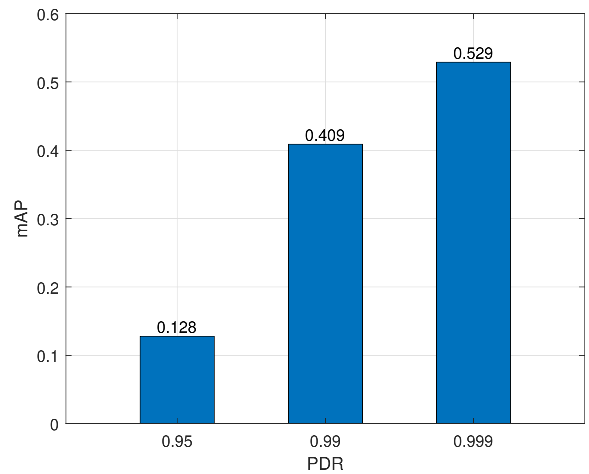


Fig. 6. mAP@.5 according to target PDR.

the subchannel occupancy rate affecting multi-user support. In the future, we will expand our research to analyze the trade-off between accuracy and communication error including experimental results, and to consider multiple users.

Acknowledgment

This research was partly supported by a grant from the Institute for Information and communications Technology Promotion (IITP) funded by the Korean government (MSIP) (No. 2021-0-01277) and by a grant from the National Research Foundation of Korea (NRF) funded by the Korean government (MSIT) (No. NRF-2021R1A2C2008415).

References

[1] A. Moubayed, A. Shami *et al.*, "Edge-enabled V2X service placement for intelligent transportation systems," *IEEE Transactions on Mobile Computing*, vol. 20, no. 4, pp. 1380–1392, 2020.

[2] H. Zhou, W. Xu *et al.*, "Evolutionary V2X technologies toward the internet of vehicles: Challenges and opportunities," *Proceedings of the IEEE*, vol. 108, no. 2, pp. 308–323, 2020.

[3] E. Coronado, G. Cebrian-Marquez, and R. Riggio, "Enabling computation offloading for autonomous and assisted driving in 5G networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[4] I. Shaer, A. Haque, and A. Shami, "Multi-component V2X applications placement in edge computing environment," in *2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[5] Y. Liu, H. Yu *et al.*, "Deep reinforcement learning for offloading and resource allocation in vehicle edge computing and networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 11 158–11 168, 2019.

[6] Y. Wang, V. Menkovski *et al.*, "VANET meets deep learning: The effect of packet loss on the object detection performance," in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, 2019, pp. 1–5.

[7] G. Cecchini, A. Bazzi *et al.*, "LTEV2Vsim: An LTE-V2V simulator for the investigation of resource allocation for cooperative awareness," in *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 2017, pp. 80–85.

[8] F. Yu, H. Chen *et al.*, "BDD100k: A diverse driving dataset for heterogeneous multitask learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2636–2645.

[9] G. Jocher, *YOLOv5*, 2020. [Online]. Available: https://github.com/ultralytics/yolov5

Fig. 7. Received video frames according to PDR.

[10] S. Ren, K. He *et al.*, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[11] K. He, G. Gkioxari *et al.*, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.

[12] J. Redmon, S. Divvala *et al.*, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

[13] W. Liu, D. Anguelov *et al.*, "SSD: Single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, 2016, pp. 21–37.