Computers, Materials & Continua CMC, vol.65, no.1, pp.243-261, 2020

Network-Aided Intelligent Traffic Steering in 5G Mobile Networks

Dae-Young Kim[1] and Seokhoon Kim[2, *]

Abstract: Recently, the fifth generation (5G) of mobile networks has been deployed and various ranges of mobile services have been provided. The 5G mobile network supports improved mobile broadband, ultra-low latency and densely deployed massive devices. It allows multiple radio access technologies and interworks them for services. 5G mobile systems employ traffic steering techniques to efficiently use multiple radio access technologies. However, conventional traffic steering techniques do not consider dynamic network conditions efficiently. In this paper, we propose a network aided traffic steering technique in 5G mobile network architecture. 5G mobile systems monitor network conditions and learn with network data. Through a machine learning algorithm such as a feed-forward neural network, it recognizes dynamic network conditions and then performs traffic steering. The proposed scheme controls traffic for multiple radio access according to the ratio of measured throughput. Thus, it can be expected to improve traffic steering efficiency. The performance of the proposed traffic steering scheme is evaluated using extensive computer simulations.

Keywords: Mobile network, 5G, traffic steering, machine learning, MEC.

1 Introduction

Recently, mobile traffic is exploding with the spread of smart phones. In 2024, mobile data generated by smartphones is expected to reach 95% of total mobile data. The expansion of Over-The-Top (OTT) services increases the demand for mobile broadband and causes an increase in mobile traffic. Massive devices for Internet of Things (IoT) services also increase mobile traffic. That is, the demand for data traffic for service subscribers of mobile networks will increase exponentially [Eriksson, Forsman, Ronkainen et al. (2019)]. There is an increasing demand for new communication to provide stable services for explosive data traffic. The fifth generation (5G) mobile network system has been proposed as a solution. The wireless and core networks specifications are defined by ITU and 3GPP. The 5G mobile communication provides high mobility in fully connected services. It covers enhanced mobile broadband (eMBB)

[1] School of Computer Software, Daegu Catholic University, Geyongsan, 38430, Korea.

[2] Department of Computer Software Engineering, Soonchunhyang University, Asan, 31538, Korea.

[*] Corresponding Author: Seokhoon Kim. Email: seokhoon@sch.ac.kr.

Received: 29 April 2020; Accepted: 25 May 2020.

and ultra-reliable and low latency communications (URLLC). It also supports massive machine type communication (mMTC). Through network virtualization, 5G mobile systems enable logical network slices. In addition, 5G mobile systems allows multiple radio access networks for 3GPP and Non-3GPP. Traffic steering among them is applied to the 5G mobile networks [Rost, Banchs, Berberana et al. (2016); 3GPP TS 23.501 (2019); 5G PPP (2019); Rommer, Hedman, Olsson et al. (2019)].

The 5G mobile network is composed of a fronthaul and backhaul network. Radio access occurs in a fronthaul network, while a backhaul network performs network core operations. For 3GPP communication, a terminal device (user equipment: UE) is connected to a next generation node B (gNB). The gNBs constitute the fronthaul network as a radio access network. Data traffic from the fronthaul network is transmitted to a user plane function (UPF) in a backhaul network. In non-3GPP communication, radio access is connected to UPF in a backhaul network through a 5G communication element. The UPF is connected to the Internet via a data core of 5G mobile networks and delivers user data traffic [3GPP TS 23.501 (2019); 5G PPP (2019); Rommer, Hedman, Olsson et al. (2019); Agyapong, Iwamura, Staehle et al. (2014)]. 5G mobile networks are completely divided into control plane and user plane in the network. The UPF enables the connection of multiple radio access technologies, and it handles user data transmission. In addition, the mobile edge computing (MEC) server can be integrated to the UPF. The MEC server provides cloud computing services to a radio access network in network edge [Kekki, Featherstone, Fang et al. (2018); Hu, Patel, Sabella et al. (2015); Mach and Becvar (2017); Frank, Fuhrmann and Ghita (2016)]; it monitors user traffic of radio access technologies and provides virtualization services for radio access technologies. Thus, network aided traffic steering can be applicable using the MEC system in the 5G mobile networks. In addition, if we employs machine learning in the MEC system, we can provide intelligent traffic steering to 5G mobile services.

The traffic steering methods can efficiently provide mobile services for users. Keeping multiple data sessions, traffic routes in radio access can be controlled: traffic steering can choose a particular radio access technology for data transmission, change particular radio access during data transmission, and employ multiple radio access technologies for mobile services. Thus, traffic steering can improve the quality of user experiences (QoE). In 5G mobile networks, because 3GPP and Non-3GPP connections are integrated to UPF, traffic steering can be operated in the UPF. In 5G standards, there are several traffic steering methods however they do not use network data statistics for monitoring. In general, applying machine learning to mobile networks and systems gives them intelligence [Alsheikh, Lin, Niyato et al. (2014); Zhang, Patras and Haddadi (2019); Rostami, Sangaiah, Wang et al. (2019)]. It can improve the accuracy of decision-making in network controls. Kim et al. [Kim, Kim and Park (2018)] provided an efficient buffer management scheme for a video player in mobile video services. It used a logistic regression algorithm to fill the video player buffer in a bad wireless network. Kim et al. [Kim, Jeong and Kim (2017)] proposed a data filtering system in a network. In that research paper, the system used naïve Bayesian classifier algorithm to find and block malfunctioning and intrusion data. Kim et al. [Kim and Kim (2018)] studied data forwarding in a delay tolerant system with a mobile sink, where a naïve Bayesian classifier algorithm was used to determine the connectivity status with a mobile sink terminal. Kim et al. [Kim and Kim (2019)] determined wireless channel migration using naïve Bayesian classifier algorithm. It maintained a better communication environment through the channel migration. As mentioned earlier, UPF integrated with MEC in 5G mobile networks can gathers traffic information of radio access technologies. If we use the gathered network information for machine learning, efficient decision-making for traffic steering is available. This leads to efficient mobile services in 5G networks through an intelligent traffic steering method.

The rest of this paper is organized as follows. Section 2 discusses related work to establish the conceptual background for this study. The network architecture for intelligent traffic steering is then discussed in Section 3. Section 4 proposes an intelligent traffic steering method using machine learning, and Section 5 presents the performance evaluation thereof. Finally, Section 6 concludes the paper.

2 Related work

5G mobile systems allow a variety of radio access networks and can use multiple connections. This situation leads to traffic steering for better mobile services. 3GPP defines traffic steering as three functions: steering, switching, and splitting [3GPP TS 23.793 (2018)]. Traffic steering is to select an appropriate radio access before data transfer, while traffic switching is to change a radio access to another one during data transfer. Traffic splitting, then, is to use multiple radio accesses by splitting a data flow. These functions are represented by access traffic steering, switching, and splitting (ATSSS). This ATSSS is a key technology in 5G mobile networks to provide various ranges of services.

There are several studies for traffic steering in mobile networks. Condoluci et al. [Condoluci, Johnson, Ayadurai et al. (2019)] provides a hybrid access gateway (HAG) using traffic splitting. The HAG is used as a traffic aggregation point in a core network so that fixed and mobile broadband services can be used simultaneously. A data flow, which the HAG receives from a content server for broadband services, is separated to a mobile network and a wired infra network and mobile terminal receives traffic from the separated data flows. Therefore, it is possible to satisfy QoS for broadband services from sufficient data traffic. Barmpounakis et al. [Barmpounakis, Magdalinos, Alonistioti et al. (2018)] proposes an analytics framework to support radio access technology (RAT) selection. It characterizes a mobile network with high data volumes and a large number of users. Traffic load is determined by the behavioral profiles of users, which are constructed by users' collected network information. According to the traffic load, an appropriate RAT is selected to avoid traffic load of a radio access network. Prasad et al. [Prasad, Moya, Ericson et al. (2016)] deals with traffic steering in the tight integration of LTE and 5G communication. An integration layer is placed in bearer (i.e., between a core network and radio access networks). It receives various data flows from a core network and selects a proper RAT for each data flow using radio link feedback. Through the integration layer entity, dynamic QoS management for data flows is provided. Nguyen et al. [Nguyen and Pham (2018)] provides a traffic steering solution to minimize routing costs in a network. In a multipath routing protocols, a routing path is determined through network function virtualization (NFV). The routing costs are calculated by a heuristic algorithm.

When multiple paths are available (i.e., there are multiple RATs or transmission paths), the traffic steering technology can provide the best connection for a given service. Therefore, 3GPP defines traffic steering modes in 5G mobile networks. There are five traffic steering modes: active-standby, priority-based, best-access, redundant, and load-balance mode [3GPP TS 23.793 (2018)]. According to the traffic steering mode, the usage of RATs in 5G mobile networks can be different. Fig. 1 represents traffic steering modes in 5G mobile networks.

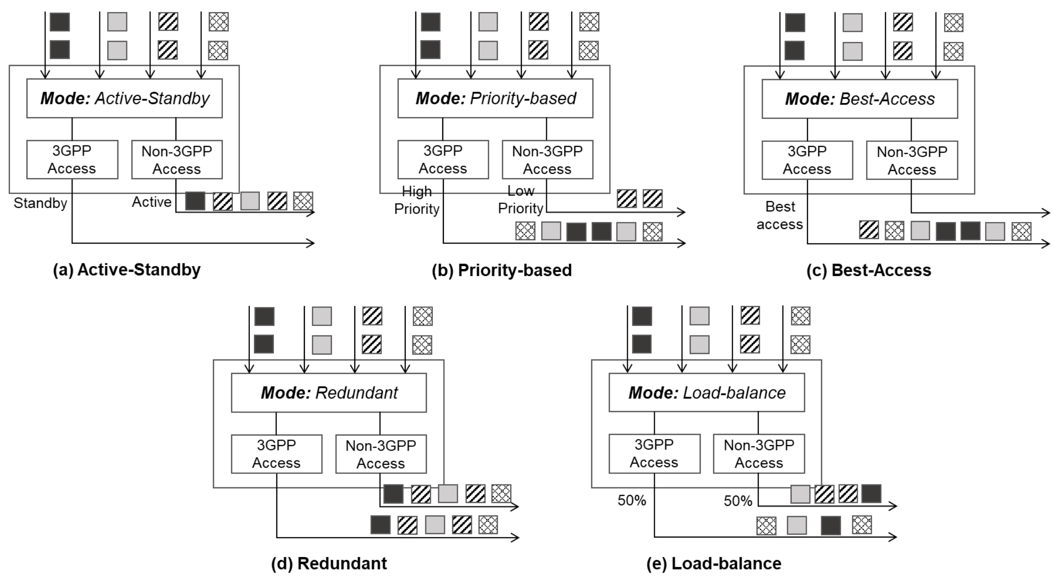


Figure 1: Steering modes for 5G mobile networks [3GPP TS 23.793 (2018)]

In the active-standby steering, all data traffic is only delivered to one RAT in the active state. The RAT in the standby state waits until the RAT in the active state becomes unavailable. If the RAT re-enters the active state, data traffic can be transferred to the RAT. This mode provides a continuity of connection. If one RAT has a problem during a mobile service, another RAT in the standby state can be used in the active state. That is, the radio access is switched. In the priority-based steering, priorities are assigned to the RATs. All data traffic is delivered to the high-priority access and if congestion occurs in the high priority access, the low priority access is used for additional data flows. This mode also provides improved connection continuity and RAT coexistence. In the best-access steering, the RAT with the smallest RTT is selected. The radio access for data transmission can be dynamically changed because this mode depends on measured RTT in RATs. This mode is known to have better performance than the priority-based steering mode. The redundant steering uses both radio accesses to increase reliability. The same data traffic is transmitted to both radio accesses. Then, even though traffic loss occurs in one radio access, lost traffic can be delivered through another radio access. The Load-balance steering exploits the percentage of data flows to indicate traffic load. The traffic load of each radio access is assigned and the data traffic is split according to the given traffic load. This mode can provide bandwidth aggregation with balanced load.

3 5G mobile network architecture

The 5G mobile network is composed of a 5G core, 3GPP radio access, and Non-3GPP radio access networks. In the 5G core network, the access and mobility management function (AMF) interacts with a radio network through signaling to manage user mobility. The unified data management function (UDM) deals with authentication data for subscribers. The session management function (SMF) manages user session (i.e., creation and release) and IP allocation for UE. The policy control function (PCF) manages policy rules for the network and services for users. The user plane function (UPF) has an important role in processing and transmitting user data. In the 3GPP radio access network, the UE connects to a next generation NodeB (gNB). As a base station, it receives a radio signal and relays to the core network. In case of the non-3GPP radio access network, an access point (AP) receives the data signal and delivers it to the UPF via the non-3GPP interworking function (N3IWF). The N3IWF provides tightly coupled integration with 3GPP and non-3GPP mobile networks. For 5G core elements, several signaling interfaces are used to transmit information for mobile networks. Data flows in 5G mobile networks are created via the UPF. Fig. 2 shows the 5G mobile network architecture. As shown in the figure, data sessions of 3GPP and non-3GPP are constructed via the UPF. The UPF manages data flows for users and provides a gate to connect to Internet. In 5G standard, multipath TCP (MPTCP) [Bonaventure and Seo (2016)] is allowed to support multiple data sessions for a single data transmission. It can use multiple network interfaces by generating subflows. This is useful when performing traffic steering.

In 5G mobile networks, the UPF gathers all data flows and integrates with radio access networks. This allows for radio access control through the UPF. That is, the UPF can select or share network interfaces for users. Thus, 3GPP 5G standard includes traffic steering techniques (i.e., ATSSS) in 5G specification as mentioned in Section 2. However, the schemes are quite simple. Mobile networks have various service conditions. For user satisfaction, it is necessary to provide intelligent traffic steering per service contents. This is why network-aided intelligent traffic steering is proposed.

Table 1: 5G mobile network notations

Elements	Descriptions
UE	User equipment
gNB	Next generation NodeB
UPF	User plane function
AMF	Access and mobility management function
SMF	Session management function
PCF	Policy control function
UDM	Unified data management function
N3IWF	Non-3GPP interworking function

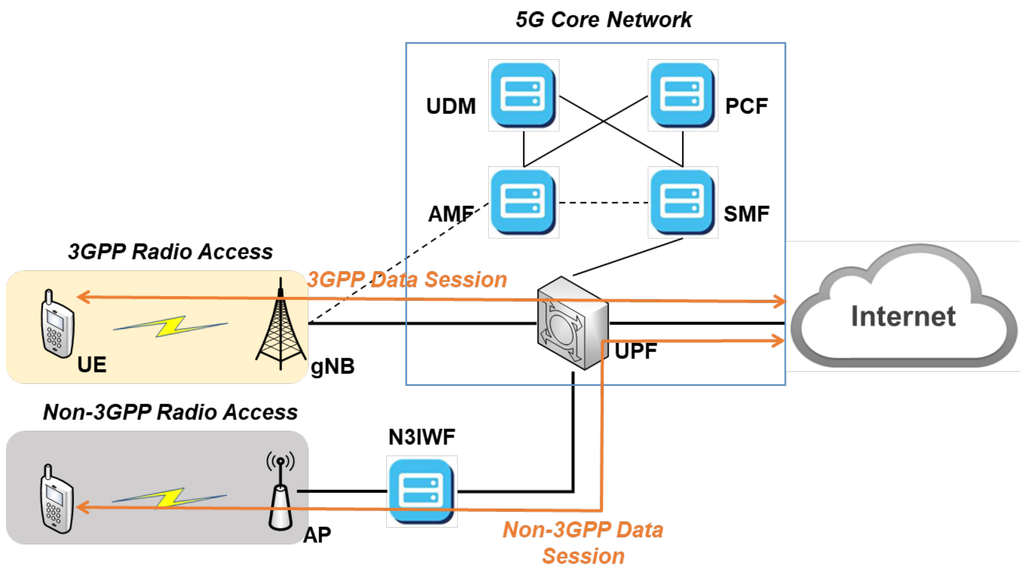


Figure 2: 5G mobile network architecture

4 Intelligent traffic steering

4.1 Network architecture for network-aided intelligent traffic steering

As mentioned earlier, mobile traffic connects to the Internet through UPF. Thus, among integration scenarios of mobile edge cloud (MEC) in 5G mobile networks, the UPF and MEC integration scenario may be appropriate. MEC provides computing resources to radio access networks so that radio access networks can offload their computing loads. It also provides local storage for radio access networks. Thus, service data for UEs can be cached in MEC for QoS. In addition, MEC monitors radio access networks and supports virtualization interfaces [Hu, Patel, Sabella et al. (2015); Mach and Becvar (2017)]. Therefore, the network-aided traffic steering method can be implemented as a function of MEC. Fig. 3 represents the network architecture for network-aided traffic steering.

Integrated UPF (i.e., UPF with MEC) can performs deep neural network learning on radio network status through network data such as signal strength, transmission delay, RTT, transmission rate, etc. It monitors RATs during data transmission and gathers network data of RATs. Learning to predict RAT status is periodically performed using updated network data. The learning results are used for traffic steering of RATs. In wireless networks, network control by learning can provide better performance because it performs decision-making using intelligence instead of a simple threshold. The integrated UPF has a role of MPTCP [Bonaventure and Seo (2016); Nguyen, Kibria, Ishizu et al. (2019); Lee and Chung (2019)] proxy and supports MPTCP flows. UEs with MPTCP functionality can use multiple subflows of MPTCP in an access network by connecting to the MPTCP proxy of the integrated UPF.

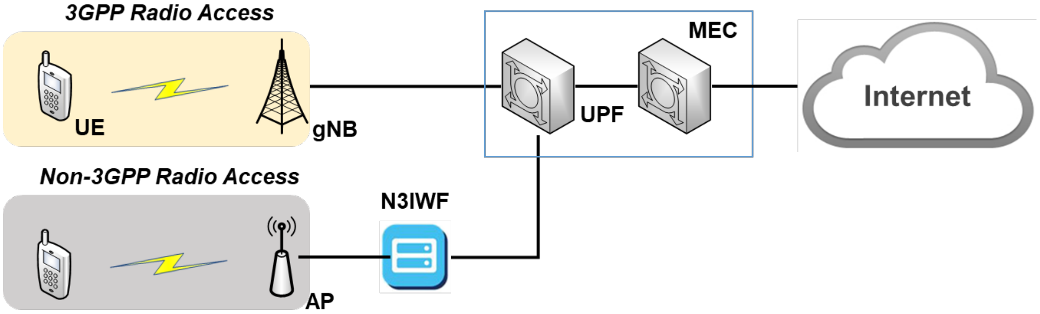


Figure 3: UPF integrated with mobile edge cloud (MEC) in 5G mobile network architecture

4.2 Network-aided traffic steering

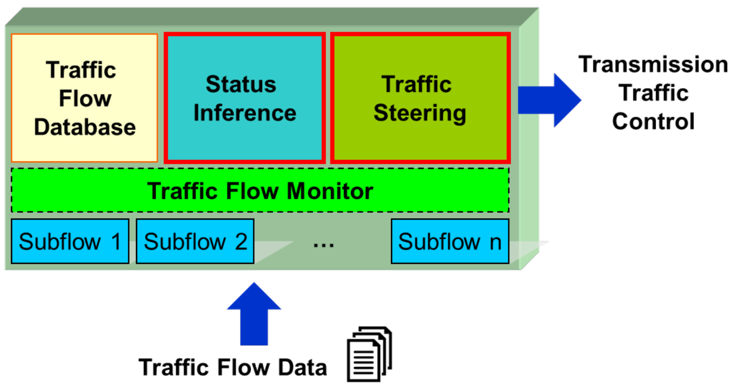


Figure 4: System architecture of the proposed method

The system for the network aided traffic steering is composed of four modules: traffic flow monitor, traffic flow database, status inference module, and traffic steering module. Fig. 4 represents a system architecture of the proposed method. In the integrated UPF, the traffic flow monitor gathers network data of traffic subflows. Gathered network data is managed by the traffic flow database. The status inference module performs learning to predict subflow status using data in the traffic flow database. This module employs a learning model using a deep neural network algorithm. The learning for network data (e.g., signal strength, delay, RTT, jitter, etc.,) is not required complex computation such as convolutional neural network for image processing. The learning results are applied to transmission traffic control by the traffic steering module. The proposed traffic steering uses intelligence in the integrated UPF. Through the intelligence by the collected network data, suitable traffic control for mobile services can be enabled.

Traffic-Steering-Algorithm ():
line 1: $f_1 \leftarrow D, f_2 \leftarrow D$
line 2: Loop:
line 3: $N_1 \leftarrow$ {RSS, delay, throughput, etc.}$_1$
line 4: $N_2 \leftarrow$ {RSS, delay, throughput, etc.}$_2$
line 5: $S_1 \leftarrow$ *Learning-Status-Inference* (N_1)
line 6: $S_2 \leftarrow$ *Learning-Status-Inference* (N_2)
line 7: If ($S_1 \neq S_2$):
line 8: $n \leftarrow floor\ (TP_g/TP_b)$
line 9: If $n > THRD$:
line 10: $f_g \leftarrow M \times D, f_b \leftarrow 0$
line 11: Else
line 12: $f_g \leftarrow n \times D, f_b \leftarrow (M-n) \times D$
line 13: End if
line 14: Else
line 15: $f_1 \leftarrow M/2 \times D, f_2 \leftarrow M/2 \times D$
line 16: End if
line 17: End loop

Figure 5: The proposed traffic steering algorithm

Table 2: Notations for the proposed traffic steering

Elements	Descriptions
f_i	Subflow i
D	Unit data size for requests
N_j	Network data of subflow j
S_j	Network state of subflow j
TP_g	Throughput of the subflow with good state
TP_b	Throughput of the subflow with bad state
$THRD$	Threshold to select a network interface
n	Throughput ratio of two subflows
M	Maximum number of request data
f_g	Subflow with good state
f_b	Subflow with bad state

Fig. 5 represents the proposed traffic steering algorithm to adjust the amount of traffic to network interfaces. Tab. 2 shows notations for the proposed traffic steering algorithm. The proposed algorithm starts with the collected network data. A status inference of network interfaces by learning is carried out in *lines 3~6*. If the states of two network interfaces are different (i.e., they are in a good and bad state, respectively), the throughput ratio of the network interfaces is calculated in *line 8*. If the radio is greater than THRD, only one network interface with a good state is selected to transmit the data traffic (*lines 9~10*). Otherwise, the amount of traffic is assigned to each network interface according to the throughput ratio (*lines 11~12*). When the network interfaces are in the same state, the same amount of traffic is assigned to each network interface (*lines 14~15*). The proposed algorithm is periodically performed with updated network data in the integrated UPF. According to the result of the proposed algorithm, traffic steering for network interfaces is carried out. The proposed algorithm employs deep neural network–based learning to improve the recognition of network status. It allocates data traffic to the network interfaces through accurate state awareness of networks. Thus, it can use network resources efficiently in 5G mobile networks. This can greatly the improve user's QoE.

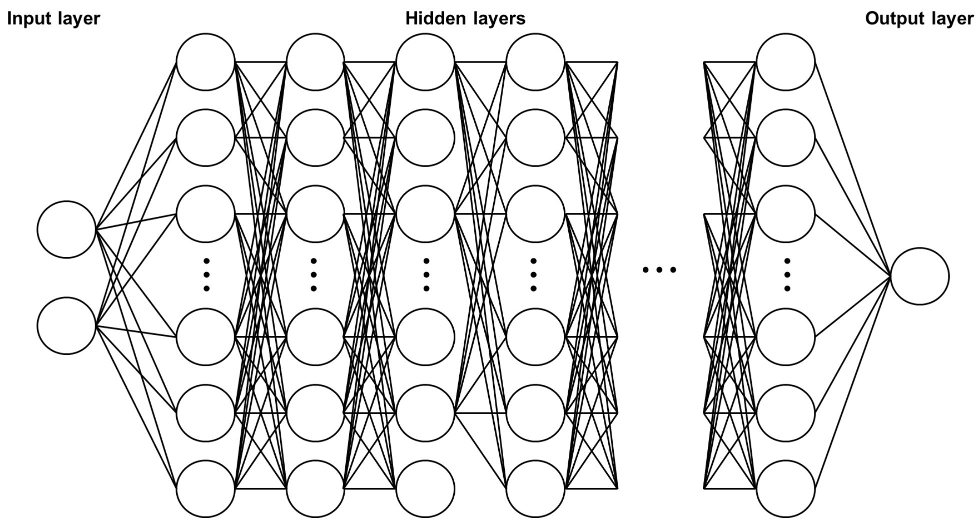


Figure 6: The deep neural network model for the network status prediction

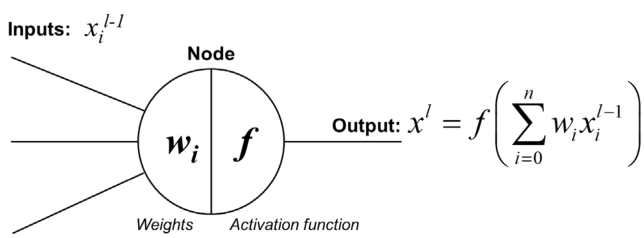


Figure 7: A node of the deep learning model: the number of inputs becomes the number of activated nodes (n) in the previous layer

The *Learning-Status-Inference* function in Fig. 5 is placed in the status inference module of the system architecture, called by the traffic steering module. This function is based on a deep neural network for learning of network status. The deep neural network model consists of 1 input layer, 8 hidden layers and 1 output layer. The number of nodes in a hidden layer is 30. Fig. 6 shows the deep learning model to predict the status of network interfaces. The learning outputs of the previous layer are used as the input of the next layer. In each node, the attributes of the input are multiplied by the weight and then used for an activation function. A result of the activation function is an output of each node. In this learning model, we use *ReLu* [Goodfellow, Bengio and Vourville (2016)] as the activation function. In addition, the dropout technique (50%) is applied to every three layers in hidden layers to reduce overfitting. Fig. 7 represents a node of the learning model. The output layer consists of a single node using *sigmoid* [Goodfellow, Bengio and Vourville (2016)] as an activation function and its output becomes the final result. An error occurs when there is a difference between this final result and the real output. To minimize this error, the deep neural network performs optimization using gradient. We employ the *Adam* [Goodfellow, Bengio and Vourville (2016)] method to optimize the weights.

4.3 Cooperation with neighbor MECs

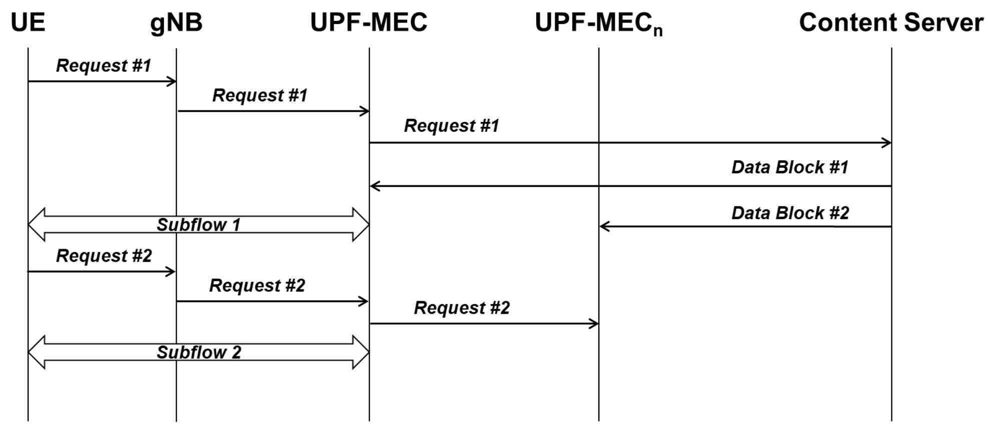


Figure 8: Collaboration with an adjacent integrated UPF

In the proposed method, the UPF is integrated with the MEC. This integrated UPF can temporarily store requested data by users. Data exchange between the integrated UPFs is also possible. Thus, when a UE attempts to request data using MPTCP, the integrated UPF can piggyback neighboring UPF information. The piggybacked request is delivered to a content server, which then divides the requested data and transmits the data blocks to the requested UPF and its neighbor UPF. When the UE creates the second subflow, the UPF requests data from its neighbor UPF. Data in the neighboring UPF is delivered to the requested UPF, which then transmits the data to the second subflow for UE. Fig. 8 shows the collaboration with an adjacent integrated UPF. Because the second subflow's request is not transmitted to a content server, delays for service time may be reduced.

5 Performance evaluation

Performance evaluations are carried out with computer simulations. The proposed method is compared with the existing traffic steering modes (i.e., best-access, priority-based and load-balance) of 5G mobile networks, using event-driven simulations. The simulator is implemented by C language with SMPL [MacDougall (1987)] library, which provides APIs for event scheduling. A mobile node (i.e., UE) has two network interfaces: 3GPP and Non-3GPP. It performs data download for performance evaluation. Both 3GPP and Non-3GPP access networks have two states of Markov chain model: a good and a bad state. The network channel state changes according to the model. In the simulation, traffic download events occur with a uniform distribution of 1 min on average. The UE requests 500 MB for a content server to download. To download the entire data, several requests are needed, each having up to 40 MB. The total simulation time is set to 24 hours. Fig. 9 represents a state diagram for the simulation. In the TFGEN state, the download events occurs and the ratio of priority download is 30%. In the NET state, the channel states of access networks are changed. If a download event occurs, the UE downloads data traffic from the C-SERV via the I-UPF and the gNB. Transmission delays are assumed to $T_{internet}$, T_{core}, and T_{access}. T_{core}, and $T_{internet}$ are set to 10 ms and 200 ms, respectively. T_{access} depends on data rate of access networks. In the AL state, traffic steering algorithms are run. In the proposed method, access network status prediction by the proposed deep neural network model is used. According to the traffic steering in the AL state, the UE adjusts the amount of requested traffic.

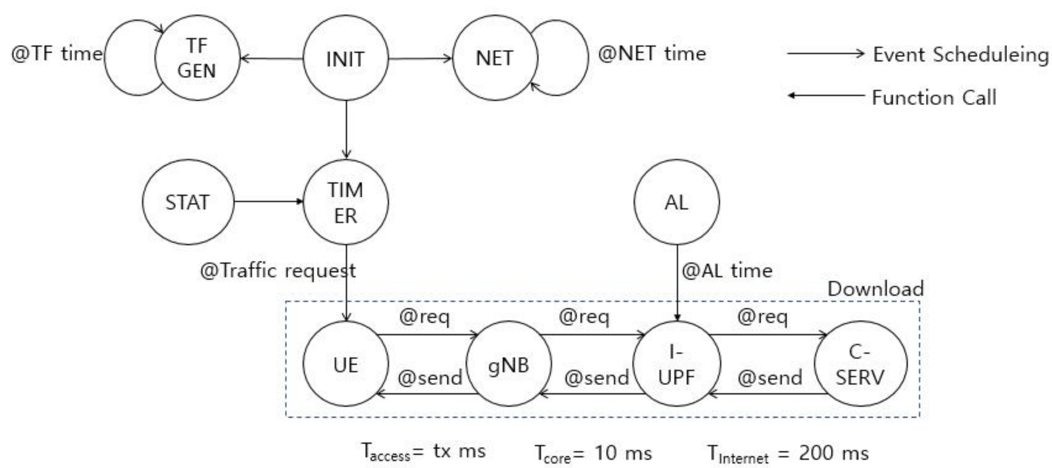


Figure 9: State diagram for the simulation

5.1 Network channel model

As shown in Fig. 10, a state transition occurs with a given probability. The probability p is used to change the state from good to bad. The good state is maintained with the probability 1-p, and the bad state is changed to the good state with the probability q, which is maintained with the probability 1-q. During the simulation, the state transition is checked every NET time. For the 3GPP access network, p and q are assumed to be 0.2 and 0.9, respectively. For the non-3GPP (i.e., high speed WiFi) network, they are assumed to be 0.4 and 0.7, respectively. The NET time is set to 5 sec. In the access network model, the non-3GPP access network has a more frequent network state transition than 3GPP. In the good state of 3GPP, the data rate and received signal strength are randomly determined in the range of 90 to 130 MB/s, and in the range of -70 to -55 dBm. In the bad state of 3GPP, they are randomly determined in the range of 10 to 50 MB/s, and in the range of -85 to -70 dBm. In the good state of non-3GPP, the data rate and received signal strength are randomly determined in the range of 20 to 100 MB/s, and in the range of -65 to -45 dBm. In the bad state of non-3GPP, they are randomly determined in the range of 62 to 500 KB/s, and in the range of -85 to -65 dBm. Tab. 3 shows the simulation parameters.

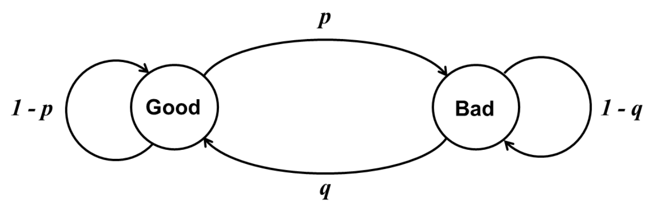


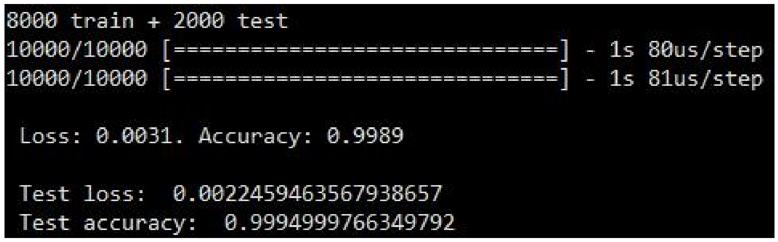
Figure 10: Network channel model for the simulation

Table 3: Simulation parameters

Description	Parameter values
Simulation Time	24 hours
NET time	5 sec
Traffic event	Uniform distribution (60 sec)
Priority traffic ratio	30%
Total download	500 MB
Max. request (M)	40
Unit size (D)	1 MB
Channel prob. (3GPP)	p=0.2 and q=0.9
Channel prob. (non-3GPP)	p=0.4 and q=0.7
THRD (in proposed al.)	20
T_{core}	10 ms
$T_{Internet}$	200 ms
Network status prediction prob.	99.9%

5.2 Network status prediction

The learning model for network status prediction in Section 4.2 is implemented with Keras [Keras (2019)], which is an open source library for deep learning. Training data consists of throughput and received signal strength as a tuple. 8,000 data tuples from a training data set with 10,000 data tuples are used for training and 2,000 data tuples are used for validation. The batch size for training samples is set to 1,000, and training is repeated for a given number of epochs-3,000 in the case of our study. As mentioned in Section 4.2, the learning model employs *ReLu* as an activation function in the hidden layers and *sigmoid* in the output layer. A loss function is binary cross-entropy and the loss is optimized by the optimizer *Adam*. Fig. 11 represents the learning result of the prediction model in Section 4.2. The model accuracy is 99.89%—a high result.



```
8000 train + 2000 test
10000/10000 [==============================] - 1s 80us/step
10000/10000 [==============================] - 1s 81us/step

 Loss: 0.0031. Accuracy: 0.9989

 Test loss:  0.0022459463567938657
 Test accuracy:  0.9994999766349792
```

Figure 11: Learning result of network status prediction

5.3 Simulation results

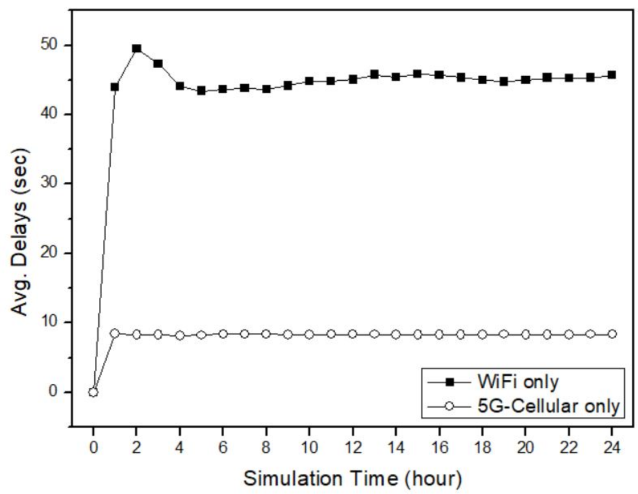


Figure 12: Average download delays in active-standby mode

Fig. 12 shows the result of active-standby mode, namely the average delays for simulation time when WiFi or 5G-cellular is only activated. The average delays are 45.105 sec in non-3GPP (WiFi only) and 8.324 sec in 3GPP (5G-cellular) only. Because the wireless conditions of the access network are frequently changed in non-3GPP, delays in the data downloads are longer than in 3GPP. In 3GPP, the wireless conditions remain relatively more stable than in non-3GPP. The data download in 3GPP causes less transmission delays but incurs more transmission costs. The transmission cost of non-3GPP using an unlicensed radio band is relatively low.

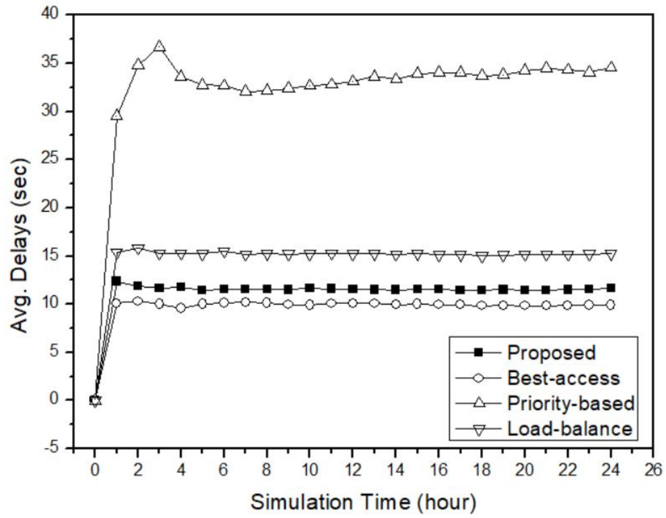


Figure 13: Average download delays

Fig. 13 shows the average delays of traffic steering modes and the proposed method. The average delay is 11.588 sec in the proposed method and 9.976 sec in the best-access mode. In the priority-based mode, the average delay is 33.485 sec, while it is 15.25 sec in the load-balance mode. As mentioned earlier, wireless conditions in non-3GPP networks are frequently changed. Thus, wireless access in bad conditions causes long transmission delays. In the case of the best-access mode, because it selects the best network for data download, it has a high 3GPP network usage rate. Thus, it shows the lowest download delays. In the load-balance mode, downloaded traffic is split into 3GPP and non-3GPP networks. Because the non-3GPP usage rate is high, it shows long download delays. In the priority-based mode, the high-priority data downloads use the 3GPP network. Thus, although the ratio of high-priority download rate is 30%, 3GPP usage can reduce the download delays. The proposed method steers access networks according to status prediction and the algorithm in Fig. 5. Therefore, the proposed method can reduce the download delays as the best-access mode. However, it does not use 3GPP as much as the best-access mode. The proposed method can be affected by the *THRD* parameter. In the simulation, the value showing the best performance in delay was set as a parameter. When the average delay was measured by changing to 3, 5, 10 and 20, the result of *THRD* at 3 and 5 were 11.84 sec and 11.65 sec. The result was the same when the *THRD* was 10 or higher.

Fig. 14 represents the average delays in high priority downloads. The priority-based mode allows high priority downloads to assign better network connection such as 3GPP. Thus, the average delays of high priority downloads in the priority-based and best-access modes are very similar. The proposed method has more delay, but the difference between the proposed method and other methods (i.e., priority-based and best-access) is only 1.2 sec. In the case of the load-balance mode, it uses 3GPP and non-3GPP networks at the same rate. Thus, due to non-3GPP usage, the average delay of high-priority downloads is longer than other methods. As shown in Figs. 13 and 14, in terms of average download delays, the best-access mode and the proposed mode provide better performance. However, the best-access mode uses more expensive 3GPP network resources.

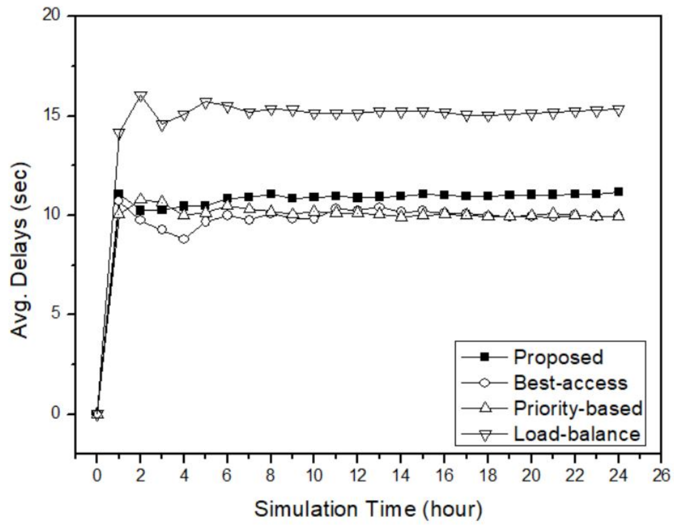


Figure 14: Average download delays of high priority downloads

Fig. 15 represents the amount of downloaded data traffic in each wireless network. The amount of traffic received in the proposed method is 181,689,915 KB and 555,610,112 KB, respectively, in non-3GPP and 3GPP. In the case of the best-access mode, the amount of traffic is 54,060,370 KB and 683,253,760 KB respectively. In the proposed method, the amount of non-3GPP traffic is more than three times higher than that of the best-access mode. Therefore, although the proposed method shows more delay, it has an offloading effect on the traffic load by using more non-3GPP. In addition, the difference in delay between the proposed method and best-access mode is not significant. In the priority-based mode, only the high-priority downloads use 3GPP. Because the high priority traffic rate is 30%, the amount of traffic downloaded from non-3GPP is higher (496,047,755 KB in non-3GPP and 241,254,400 KB in 3GPP). In the load-balance mode, when a download request occurs, both 3GPP and non-3GPP are used. In general, non-3GPP access networks are not stable and change frequently. Thus, the amount of traffic downloaded from non-3GPP (213,788,795 KB) is less than 3GPP (523,509,760 KB). Because the priority-based and the load-balance modes use more non-3GPP than the best-access mode and the proposed method, the amount of non-3GPP traffic is usually more.

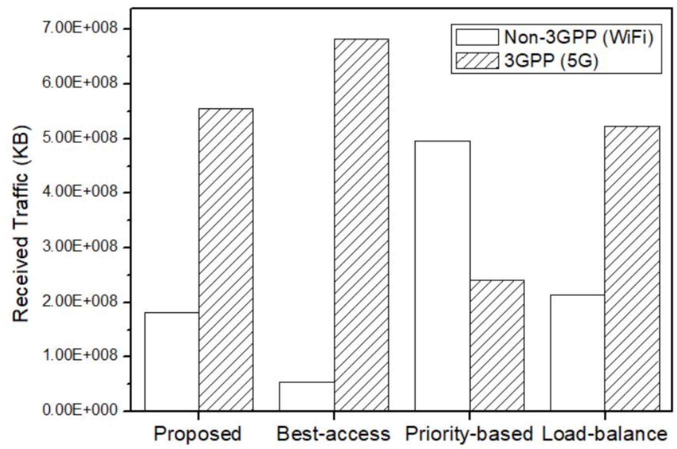


Figure 15: Amount of downloaded data traffic

In Section 4.3, the proposed method uses neighbor UPF-MEC as temporary data storage. If there is data traffic to be downloaded in the neighboring UPF-MEC, the delay to request/transmit traffic to/from a content server on the Internet can be reduced because it is possible to request a download to a neighboring UPF-MEC instead of a content server. Fig. 16 shows the download delay with and without the neighboring UPF-MEC in the proposed method. It can be seen that the download delay is longer when the neighboring UPF-MEC is not used. The difference in delay is about 200 ms and this can be viewed as Internet access time.

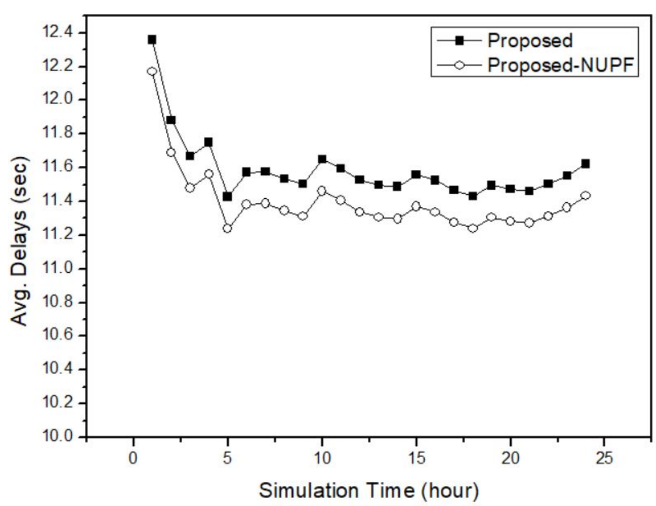


Figure 16: Download delay with and without neighbor UPF-MEC in the proposed method

6 Conclusion

Traffic steering technology in 5G mobile network systems is important for the efficient use of radio resources. The 5G mobile network standard provides five traffic steering modes for 3GPP and non-3GPP radio access networks: the active-standby, priority-based, best-access, redundant and load-balance mode. According to each steering mode, the usage of radio resources can be different. Although the five steering modes in the standard can be supported for mobile services, they do not lead to efficient usage of radio resources because they do not consider radio conditions dynamically. Therefore, a traffic steering method that considers dynamic radio conditions is proposed. The proposed method employs a deep neural network to predict radio conditions. According to the learning result, it steers the amount of traffic of radio networks. Among the five steering modes in the standard, the best-access mode shows the best performance in aspect of transmission delay. However, it uses more expensive 3GPP radio access. The proposed method has a slightly longer delay time (1.2 sec), but performs three times the data offloading than that of the best-access mode during data downloads. Therefore, the proposed method enables mobile services to use radio resources more efficiently.

Funding Statement: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2015-0-00403) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation), and this work was supported by the Soonchunhyang University Research Fund.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

3GPP TS 23.501 (2019): System architecture for the 5G system (Release 15). *3GPP Technical Specification*, TS 23.501, ver. 15.5.0.

3GPP TS 23.793 (2018): Study on access traffic steering, switch and splitting support in the 5G system architecture (Release 16). *3GPP Technical Specification*, TS 23.793, ver. 16.0.0.

5G PPP (2019): View on 5G architecture. *5G PPP Architecture Working Group White Paper*, ver. 3.0.

Agyapong, P. K.; Iwamura, M.; Staehle, D.; Kiess, W.; Benjebbour, A. (2014): Design considerations for a 5G network architecture. *IEEE Communications Magazine*, vol. 52, no. 11, pp. 65-75.

Alsheikh, M. A.; Lin, S.; Niyato, D.; Tan, H. P. (2014): Machine learning in wireless sensor networks: algorithms, strategies, and applications. *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1996-2018.

Barmpounakis, S.; Magdalinos, P.; Alonistioti, N.; Kaloxylos, A.; Spapis, P. et al. (2018): Data analytics for 5G networks: a complete framework for network access selection and traffic steering. *International Journal on Advances in Telecommunications*, vol. 11, no. 3 & 4, pp. 101-114.

Bonaventure, O.; Seo, S. (2016): Multipath TCP deployments. *IETF Journal*, vol. 12, no. 2, pp. 24-27.

Condoluci, M.; Johnson, S. H.; Ayadurai, V.; Lema, M. A.; Cuevas, M. A. et al. (2019): Fixed-mobile convergence in the 5G era: from hybrid access to converged core. *IEEE Network*, vol. 33, no. 2, pp. 138-145.

Eriksson, A. C.; Forsman, M.; Ronkainen H.; Willars, P.; Östberg, C. (2019): 5G new radio RAN & transport choices that minimize TCO. *Ericsson Technology Review*, https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/5g-nr-ran-and-transport-choices-that-minimize-tco.

Frank, H.; Fuhrmann, W.; Ghita, B. (2016): Mobile edge computing: requirements for powerful mobile near real-time applications. *Proceedings of the 11th International Network Conference*, pp. 63-66.

Goodfellow, I.; Bengio, Y.; Courville, A. (2016): *Deep Learning*. The MIT Press, Cambridge, London, England.

https://doi.org/10.1186/s13638-019-1396-2.

Hu, Y. C.; Patel, M.; Sabella, D.; Sprecher, N.; Young, V. (2015): Mobile edge computing-a key technology towards 5G. *ETSI White Paper*, no. 11.

Kekki, S.; Featherstone, W.; Fang, Y.; Kuure, P.; Li, A. et al. (2018): MEC in 5G networks. *ETSI White Paper*, no. 28.

Keras (2019): *The Python Deep Learning Library*. https://keras.io.

Kim, D. Y.; Jeong, Y. S.; Kim, S. (2017): Data-filtering system to avoid total data distortion in IoT networking. *Symmetry Journal*, vol. 9, no. 1, ID 16.

Kim, D. Y.; Kim, S. (2019): Gateway channel hopping to improve transmission efficiency in long-range IoT networks. *KSII Transactions on Internet and Information Systems*, vol. 13, no. 3, pp. 1599-1610.

Kim, S.; Kim, D. Y. (2018): Efficient data-forwarding method in delay-tolerant P2P networking for IoT services. *Peer-to-Peer Networking and Applications*, vol. 11, no. 6, pp. 1176-1185.

Kim, S.; Kim, D. Y.; Park, J. H. (2018): Traffic management in the mobile edge cloud to improve the quality of experience of mobile video. *Computer Communications*, vol. 118, pp. 40-49.

Lee, S.; Chung, K. (2019): Reducing the flow completion time for multipath TCP. *KSII Transactions on Internet and Information Systems*, vol. 13, no. 8, pp. 3900-3916.

MacDougall, M. H. (1987): *Simulating Computer Systems, Techniques and Tool*. The MIT Press, Cambridge, USA.

Mach, P.; Becvar, Z. (2017): Mobiled edge computing: a survey on architecture and computation offloading. *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1628-1656.

Nguyen, K.; Kibria, M. G.; Ishizu, K.; Kojima, F.; Sekiya, H. (2019): An approach to reinforce multipath TCP with path-aware information. *Sensors*, vol. 19, no. 3, ID. 476.

Nguyen, T.; Pham, T. (2018): Optimization model and algorithm for dynamic service-aware traffic steering in network functions virtualization. *Proceedings of IEEE International Conference on Communications and Electronics*, pp. 107-112.

Prasad, A.; Moya, F. S.; Ericson, M.; Fantini, R.; Bulakci, Ö. (2016): Enabling RAN moderation and dynamic traffic steering in 5G. *Proceedings of IEEE Vehicular Technology Conference*, pp. 1-6.

Rommer, S.; Hedman, P.; Olsson, M.; Frid, L.; Sultana, S. et al. (2019): *5G Core Networks: Powering Digitalization*. Academic Press, London, UK.

Rost, P.; Banchs, A.; Berberana, I.; Breitbach, M.; Doll, M. et al. (2016): Mobile network architecture evolution toward 5G. *IEEE Communications Magazine*, vol. 54, no. 5, pp. 84-91.

Rostami, S. M. H.; Sangaiah, A. K.; Wang, J.; Liu, X. Z. (2019): Obstacle avoidance of mobile robots using modified artificial potential field algorithm. *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, ID 70,

Zhang, C.; Patras, P.; Haddadi, H. (2019): Deep learning in mobile and wireless networking: a survey. *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3. pp. 2224-2287.