

Research Article

Passive Method for Estimating Available Throughput for Autonomous Off-Peak Data Transfer

Kozo Satoda ^{1,2} **Eiji Takahashi**¹ **Takeo Onishi**¹ **Takayuki Suzuki**¹ **Daisuke Ohta**¹ **Kosei Kobayashi**¹ and **Tutomu Murase**³

¹System Platform Research Laboratories, NEC Corporation, Kawasaki 211-8666, Japan

²Graduate School of Informatics, Nagoya University, Nagoya 464-8601, Japan

³Information Technology Center, Nagoya University, Nagoya 464-8601, Japan

Correspondence should be addressed to Kozo Satoda; k-satoda@cb.jp.nec.com

Received 22 November 2019; Accepted 22 January 2020; Published 17 February 2020

Academic Editor: Telmo Reis Cunha

Copyright © 2020 Kozo Satoda et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Large demands for mobile traffic subject base stations to frequent short-term and sharp peak loads. Recent analysis of data traffic on commercial mobile networks reported that the traffic peaks can be reduced by an average of 40% without compromising the quality of experience provided to the end user, if a peak load can be shifted for at most 20 s. To reduce peak traffic, we previously proposed a method for off-peak data transfer, with which user equipment (UE) autonomously delays receiving data, and a peak load on a base station can be shifted. In terms of off-peak transfer of data, a significant problem is determining how each UE estimates available throughput. In this paper we propose a method of passively estimating available throughput of each UE. We evaluated the effectiveness of the proposed method through experiments on experimental and commercial LTE networks. The results indicate that our method obtains more than a 0.7 correlation between actual available throughput and estimated throughput.

1. Introduction

Thanks to the recent spread in mobile broadband networks and evolving smart phones, worldwide mobile traffic increased to 14 Exabytes/month at the end of 2017 and is expected to keep increasing to 110 Exabytes/month by 2023 [1].

However, large mobile traffic demands subject base stations to frequent short-term and sharp peak loads. AT&T reported that peak loads on base stations occur frequently throughout the day by analyzing the data traffic on commercial mobile networks [2]. This report also pointed out that it is possible to reduce the traffic peaks by an average of 40% without additional network equipment and deterioration in the quality of experience (QoE) provided to the end user when a peak load can be shifted for at most 20 s.

Cisco's report said that more than 80% of traffic through base stations come from the Internet, and this percentage is increasing [3]. Internet traffic should be classified as either

delay-sensitive data (e.g., voice over IP, web-browsing data, and video streaming) or non-delay-sensitive data (e.g., content-download and software-update data). Since base stations cannot distinguish between these two types of data, they treat them equally. Therefore, the quality of delay-sensitive traffic deteriorates when a peak load occurs. For example, when download packets fill a base station and the base station is subject to peak load, voice packets wait for the base station to allocate wireless resources to the voice packets. This delay might deteriorate user perceived quality.

To solve this problem, we previously proposed a method for off-peak data transfer [4]. With this method, user equipment (UE) autonomously delays non-delay-sensitive data until a load on a base station decreases. This method reduces traffic peaks on base stations and improves quality of delay-sensitive traffic during a peak load on a base station.

In terms of off-peak transfer of data, a significant problem is how each UE monitors available throughput, which is the expected communication rate between the UE

and a base station. There has been extensive research on available throughput on the Internet and mobile networks. Sato et al. and Kitahara et al. proposed available throughput estimation methods for mobile networks using trial downloads and packet-pair probes [5, 6]. These methods are called active measurement methods since they require additional data traffic for estimation. However, this causes additional load on a base station resulting in heavier traffic load. Therefore, such methods are not suitable for off-peak data transfer systems, and a passive method that does not require additional traffic is needed.

We also previously proposed methods of passively and accurately estimating the resource block (RB) usage rate of a serving cell [4, 7]. In this paper we propose a passive estimation method of UE's available throughput. By estimating the available throughput, a UE can create a more efficient download schedule for off-peak data transfer rather than only using the RB usage rate of a base station.

The contributions of this paper are as follows.

- (i) A method for passively estimating a UE's available throughput in addition to a base-station load is proposed.
- (ii) The effectiveness of the proposed method was validated through experiments on both experimental and commercial LTE networks. The results indicate that our method obtains more than a 0.7 correlation between actual available throughput and estimated throughput.
- (iii) Our proposed method can be used for off-peak transfer of data, which can mitigate the peak load of a base station and improve communication quality without compromising the QoE provided to the end user.

The rest of this paper is organized as follows. We first discuss related work. We then define off-peak data transfer system architecture. Next we discuss our proposed method for passively estimating available throughput of each UE. We then present our evaluation of estimation accuracy involving both experimental and commercial LTE networks. Finally we summarize the key points of this study.

2. Related Work

For both network applications and network management, available throughput estimation on mobile networks is becoming important because mobile networks are widely used, and the throughput of mobile access networks between a base station and UE fluctuates quite a bit due to radio quality fluctuation such as radio interference and UE motion.

Available throughput is typically estimated using active measurement methods which involve sending packets for estimation. Several research papers have been published on active measurement methods and several tools are available [5, 8–13]. However, probe packets, which active measurement methods send for estimating available throughput, increase a base-station load. In addition, even a single-packet probe causes the radio link to remain in a high-energy state on UEs, which consumes much energy on UEs. Therefore,

for mobile networks, a passive measurement method that does not generate additional traffic load is needed.

Aceto et al. reported that the achievable TCP throughput has a relationship with available bandwidth, which can be measured using measurement method such as Yaz and Pathload [14]. They also suggested that available throughput can be estimated through radio quality between a base station and UE, and radio quality can be acquired only by a UE.

A method that allows a UE to passively measure cellular load was proposed [15]. This method computes the ratio of pilot power to total power as a power ratio. However, it only roughly predicts broad trends of cell-load fluctuation without taking interference due to neighboring cells into account.

A simple model that regresses LTE download throughput on independent metrics, i.e., the reference signal receive power (RSRP), reference signal receive quality (RSRQ), and signal-to-interference-plus-noise ratio (SINR), was developed [16, 17]. However, while the correlation with each parameter can be checked, it does not show how to accurately predict the throughput.

Iordache and Marghescu investigated the relationship between the SINR and LTE downlink throughput (bit/s/Hz) [18]. This method can calculate throughput between a base station and UE once the RB usage rate of a serving cell is known [4, 7]. However, the 3GPP specifies that total data bit per RB that an RB is able to convey is dependent on the RB usage rate [19]. Therefore, it requires a calculation process based on the RB usage rate for more accurate throughput estimation.

Recent research estimated available throughput using machine learning technology [20–24].

Both Yue et al. and Raca et al. proposed available throughput estimation methods based on machine learning using the time series of wireless information such as the RSRP, RSRQ, and channel quality indicator (CQI) [20–22]. They also use the time series of additional probe packets for estimation, such as application throughput, packet data convergence protocol (PDCP) throughput, and PDCP delay. These methods can be categorized as active estimation methods.

Falkenberg et al. propose a machine learning-based passive estimation method that only uses wireless information [23]. However, this method uses an additional sniffer for measuring wireless information besides a UE.

Nikolov et al. passively estimates an available uplink throughput on universal mobile telecommunications system (UMTS) and LTE networks using a neural network [24]. Machine learning is a powerful tool for estimating available throughput. However, validity of learning data and the possibility of applying other environments should be considered. Therefore, machine-learning-based method should be combined with analytical estimation methods.

3. Architecture for Off-Peak Data Transfer System

This section describes the system architecture of an off-peak data transfer system.

Since the traffic load of a base station fluctuates, the base station is subjected to short- to long-term peak loads [25, 26]. Human behavior is the major cause of long-term

peak loads. For example, the difference in human behavior between day and night induces daily peak loads. A large event encourages people to watch live video streaming, which also causes a huge amount of data traffic. The intervals of long-term peak loads are minutes, hours, weeks, or even months.

As mentioned in the Introduction, AT&T reported that short-term peak loads occur in seconds [2]. Unlike long-term peak loads, short-term peak loads are caused by variation in data communication such as downloaded data size, session duration, and traffic pattern of each session. Short-term peak loads occur every several seconds. We focus on controlling only data-communication traffic, not human behavior. Therefore, the remainder of this paper discusses the off-peak data transfer of short-term peak loads.

We assume the simple system architecture of an off-peak data transfer system shown in Figure 1. The system consists of servers, a base station, and UEs. UEs communicate with servers through the base station. When the amount of traffic through the base station becomes large (peak-load status), a UE and server, whose data communication is not delay-sensitive, delay the data communication until the base-station load decreases. By putting off transmission of non-delay-sensitive data communication, the base-station load decreases and a UE can receive delay-sensitive traffic more quickly.

To realize an off-peak data transfer system, we first define a system architecture. An off-peak data transfer system can be classified in terms of where base-station load or available throughput is monitored and the traffic is controlled. In this section, we analyze the characteristics of candidate architectures.

3.1. Base-Station Load Monitoring. Either a server, base station, or UE can monitor a base-station load in theory. We summarize the characteristics for each place where base-station load is monitored.

Since a server cannot directly determine base-station load, it requires a new protocol to do this. Active estimation methods are one solution to estimate base-station load or available throughput. Although various active estimation methods have been studied, the packets that an active estimation sends increase base-station load.

Another method involves deploying a new protocol that reports metrics about base-station load. However, this requires significant time to develop and distribute base stations that deploy a new protocol.

A base station can essentially detect its load.

Similar to server-side monitoring, a UE cannot directly determine base-station load. However, the UE has metrics on the radio quality between the base station and UE. Some studies reported that these metrics are related to base-station load and available throughput [7, 20, 23]. Therefore, client-side monitoring can be adopted by analyzing base-station load using these metrics.

For monitoring base-station load, base-station monitoring and client-side monitoring are possible candidate architecture of off-peak data transfer system.

3.2. Traffic Control. The same as a base-station monitoring, either a server, a base station, or a UE is able to control traffic in theory. We summarize the characteristics for each place where non-delay-sensitive traffic is controlled.

Since a server sends its content to a UE, it is easy for the server to delay its traffic if the content the server sends is not delay-sensitive. However, a transport protocol should stay connected between a server and a UE while non-delay-sensitive data are delayed. This could consume server resources due to a lot of transport connections such as denial of service (DoS) attacks.

It is difficult to control traffic at a base station because the base station cannot determine the content or type of content of the traffic such as source/destination of packet and whether it is delay-sensitive.

Similar to server-side controlling, a client can easily control the data-downloading time of its downloaded data if the content the client receives is not delay-sensitive. In addition, a client can disconnect transport connection while non-delay-sensitive data are delayed and reconnect transport connection and continue downloading. Client-side controlling could be better than server-side controlling in terms of server resources used.

3.3. Candidate Architecture. Table 1 summarizes the characteristics of each candidate architecture. It is better for an off-peak data transfer system to monitor base-station load and control traffic at the same place because it is not necessary to send monitoring results to the controller. Therefore, client-side monitoring and controlling constitute the best architecture for off-peak data transfer systems.

The main problem remaining is how clients estimate base-station load or available throughput. Estimating them usually requires active measurement such as trial downloads or packet-pair probes. However, passive measurement involves estimating a load or throughput without additional traffic. Passive measurement is better in terms of traffic amount, computation load, and battery consumption. A client can obtain radio quality metrics between a base station and client, and base-station load or available throughput can be analyzed using these metrics through passive measurement.

We focus on short-term off-peak data transfer by passively monitoring base-station traffic on a client. In the next section, we explain our proposed method for estimating base-station load and available throughput from metrics that the client can acquire.

4. Proposed Method for Estimating Available Throughput

Our proposed method estimates a UE's available throughput on the UE side. We define the RB usage rate as base-station load and the available throughput as the data-communication rate between a base station and UE. We infer a UE's available throughput from metrics that the UE can acquire.

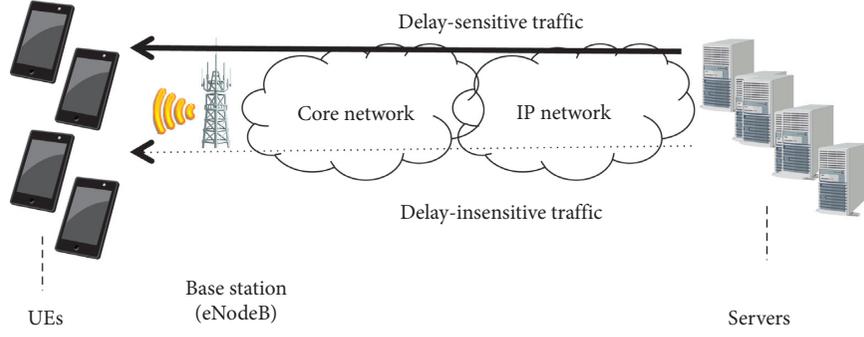


FIGURE 1: System architecture of off-peak data transfer system.

TABLE 1: Classification of off-peak data transfer system architecture.

	Monitoring	Controlling
Server-side	Requires new protocol or active measurement method	Easy for controlling sending content; many transport connections occur
Base station	Essentially possible	Difficult to understand content
Client-side	Passive measurement method should be used	Easy for controlling receiving content; transport connection can be disconnected

4.1. RB Usage Rate Estimation. We previously proposed methods for passively estimating base-station load [4, 7]. In this subsection, we first briefly explain these methods because estimation of a UE's available throughput uses the result of base-station load estimation.

The structure of the RBs of an LTE downlink is shown in Figure 2. This figure includes the RB structure for both one and two transmission antennas. An RB is the smallest allocation unit of a radio resource in a cell. The smallest modulation structure in an LTE downlink is the resource element (RE). An RE is defined as one 15 kHz subcarrier by one orthogonal frequency division multiplexing (OFDM) symbol. An RB has dimensions of twelve consecutive subcarriers in the frequency domain and six or seven OFDM symbols in the time domain form each RB.

For this structure, the RB usage rate of the serving cell is estimated from the radio conditions of the measurement-target OFDM symbol. The RB usage rate is given by dividing the number of used RBs by the number of all RBs at each time slot.

The main metrics considered when estimating the RB usage rate are the SINR and RSRQ [19, 27]. The SINR is the ratio of the power of reference signals (RSs) to that of the interference and thermal noise. The RSRQ is an index that indicates the quality of the received RS. These metrics can be acquired by a UE or calculated using other metrics such as the RSRP and RSSI acquired by a UE.

From these metrics, a base-station load u_k for one transmission antenna can be derived as

$$u_k = \frac{1}{5} \left(\frac{1}{2q_k} - \frac{6}{s_k} - 1 \right), \quad (1)$$

where q_k and s_k denote the RSRQ and SINR, respectively.

Since an RB structure for two transmission antennas is different from that for one transmission antenna, as shown in Figure 2, estimation of u_k for two transmission antennas is not identical to equation (1) and given as

$$u_k = \frac{1}{8} \left(\frac{1}{2q_k} - \frac{6}{s_k} - 2 \right). \quad (2)$$

4.2. Available Throughput Estimation. In this subsection, we explain our proposed method with which a UE estimates the available throughput based on metrics it can acquire. By using the RB usage rate, a UE can only detect congestion on a base station. However, the UE can determine how much data it can receive from a base station by estimating the available throughput, and the UE can create more efficient download schedule for the off-peak data transfer.

4.2.1. Estimation Method. Available throughput means how much throughput a UE can use for off-peak data transfer. It can be basically estimated by the number of unused RBs and using a modulation method. The number of unused RBs is the number of RBs that the base station can allocate to the UE during one time slot (transmission time interval, TTI). The modulation method shows how much data the base station can contain in each RB, which is determined by the signal quality between the UE and base station.

We can already estimate the RB usage rate of a base station using the RB-usage-rate-estimation methods. The number of unused RBs is simply acquired by multiplying the total number of RBs the base station served by the RB usage rate. Each RB occupies 180 kHz in the frequency domain and the total number of RBs in 1 TTI can be calculated according to the LTE system bandwidth [19]. Table 2 summarizes the

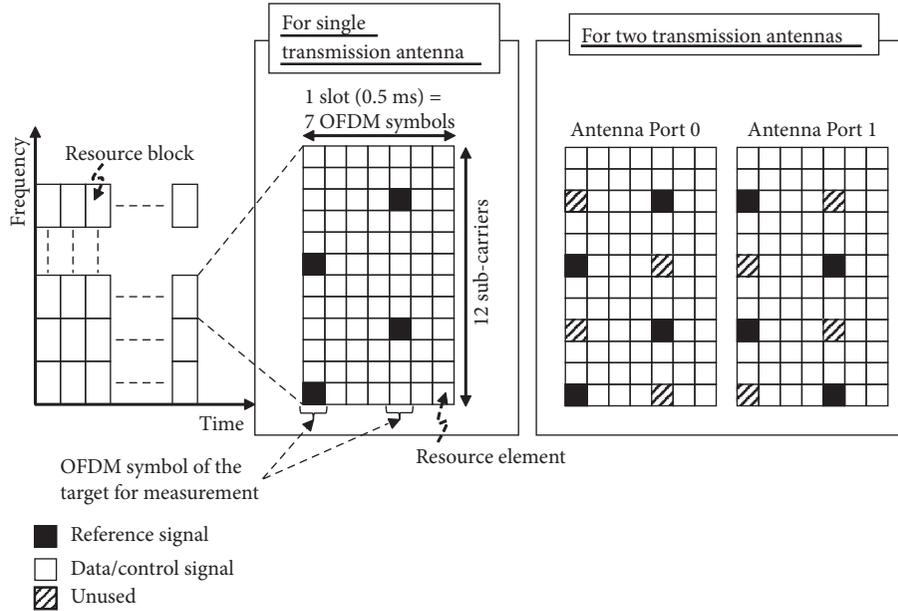


FIGURE 2: RB structure of LTE downlink for one and two transmission antennas.

relation between LTE system bandwidth and number of RBs. If a UE cannot measure the LTE system bandwidth passively, the total number of RBs is calculated from the RSRQ, RSRP, and RSSI. RSRQ is given

$$q = \frac{p}{r/b}, \quad (3)$$

where q , p , r , and b denote the RSRQ, RSRP, RSSI, and total number of RBs, respectively [7].

From equation (3), b can be calculated by the following equation.

$$b = \frac{rq}{p}. \quad (4)$$

Then, the number of unused RBs N_{rb} is calculated using b and estimated base-station load u_k from equation (1).

$$N_{rb} = u_k \cdot b. \quad (5)$$

Next we should determine a modulation method for the LTE downlink from the base station to UE.

The amount of data each RB can carry is dependent on the modulation method the base station used, and such a method is selected based on the signal quality between the base station and the UE.

In the case of good signal quality, the base station selects a more efficient modulation method such as 64QAM (quadrature amplitude modulation) because errors between the base station and UE occurs with low possibility. On the other hand, the base station uses not efficient but a higher error-resilient modulation method such as QPSK (quadrature phase shift keying) because errors are likely to occur.

Figure 3 shows how a base station (eNodeB) selects a modulation method. The eNodeB sends RS to a UE. From the RS status, the UE calculates the CQI, which shows the signal channel quality between the UE and eNodeB. The CQI

TABLE 2: LTE system bandwidth and number of RBs supported by LTE.

Total bandwidth (MHz)	Number of RBs b
1.4	6
3	15
5	25
10	50
15	75
20	100

is represented as a 4-bit discrete value and CQI = 1 and 15 means worst and best signal quality, respectively. The calculation of the CQI is vendor specific, and it is known that the CQI has a close relationship with the SINR [28, 29]. The UE sends the feedback of channel state information (CSI), which includes the CQI to the eNodeB [30].

When eNodeB receives the CQI, it selects the modulation and coding scheme (MCS) for downlink to the UE. The LTE system defines guidelines of selecting a modulation method according to the CQI. However, how a base station determines MCS for the received CQI is also vendor specific. We create and use a mapping table, Table 3, as one mapping example that follows the LTE system guideline [19].

Then the eNodeB modulates the data on each RB in the modulation mode using the MCS. The amount of data the eNodeB transmits to the UE (total block size; TBS) depends on both MCS and number of RBs the eNodeB allocates to the UE. Table 4 shows the mapping between the MCS and TBS index, which is defined in 3GPP standards [19]. Therefore, we can derive the mapping of both the CQI to MCS and MCS to TBS and the merged table of both mappings, Table 3.

Table 5 shows the TBS obtained through the TBS index and the number of RBs. This table is extracted from a table in 3GPP standards [19]. The TBS table in 3GPP is huge because 3GPP

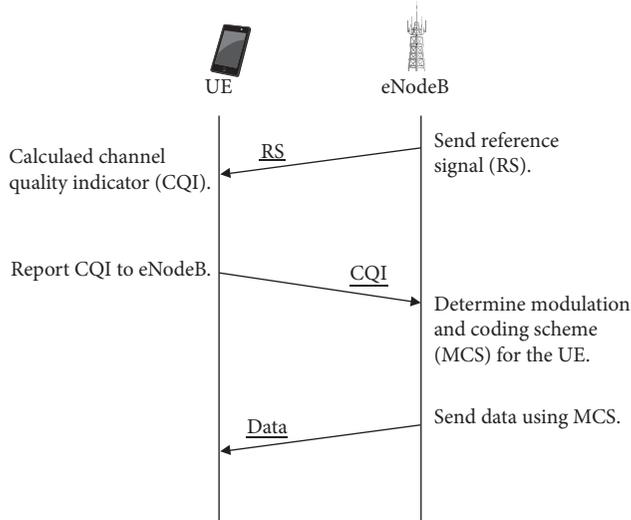


FIGURE 3: Sequence of determining MCS.

standards defined TBS of the number of RBs from 1 to 110. We just extracted the number of RBs up to 10, as shown in Table 5.

Since the CQI and number of RBs are calculated in each UE, a UE can estimate the TBS and available throughput as well.

This paragraph explains the proposed method using examples. We assume the total number of RBs is 50 and estimation value of the RB usage rate is 80% for these examples. In this case, ten RBs are available for the UE. We estimate available throughputs for three CQIs (CQI = 3, 10, 15) by using Table 5.

CQI = 3. When the CQI is 3, which shows quite low radio quality between a base station and UE, the MCS and TBS indices are both 2, as shown in Table 3. Table 5 shows that the TBS is 424 bits when the TBS index is two and at most 10 RBs can be allocated to the UE. In the base station, RBs are allocated at every TTI and the TTI of LTE is 1 ms. Therefore, the available throughput is estimated as 424 kbps.

CQI = 10. When the CQI is 10, the MCS and TBS indices are 18 and 16, respectively, as shown in Table 3. Table 5 shows that the TBS is 3240 bits when the TBS index is 18 and at most 10 RBs can be allocated to the UE. The available throughput is then estimated as 3.24 Mbps.

CQI = 15. When the CQI is 15, which shows the highest radio quality, the MCS and TBS indices are 28 and 26, respectively, as shown in Table 3. Table 5 shows that the TBS is 7480 bits when the TBS index is 28 and at most 10 RBs can be allocated to the UE. Then the available throughput is estimated as 7.48 Mbps.

Reprinted from “3GPP TS 36.213 Rel.15, Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures,” Sept. 2018.

5. Evaluation of Estimated RB Usage Rate and Available Throughput

We conducted experiments for evaluating our proposal method. One experiment was for evaluating the accuracy of estimating the RB usage rate in an experimental LTE network. Another was for evaluating the accuracy of estimating the available throughput in a commercial LTE network.

TABLE 3: CQI, MCS, and TBS indices mapping table.

CQI index	MCS index	TBS index
1	0	0
2	0	0
3	2	2
4	4	4
5	6	6
6	8	8
7	11	10
8	13	12
9	15	14
10	18	16
11	20	18
12	22	20
13	24	22
14	26	24
15	28	26

TABLE 4: MCS and TBS indices table.

MCS index	TBS index
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	9
11	10
12	11
13	12
14	13
15	14
16	15
17	15
18	16
19	17
20	18
21	19
22	20
23	21
24	22
25	23
26	24
27	25
28	26/26A
29	Reserved
30	Reserved
31	Reserved

5.1. Evaluation Accuracy of Estimating RB Usage Rate in Experimental LTE Network

5.1.1. Experimental Setting. We first evaluated the accuracy of estimating the RB usage rate using an experimental LTE network. Figure 4 shows the system architecture for this experiment. Two UEs, UE A and UE B, were attached to the

TABLE 5: TBS table (nRB = 1, ..., 10).

TBS index	Number of RBs (nRB)									
	1	2	3	4	5	6	7	8	9	10
0	16	32	56	88	120	152	176	208	224	256
1	24	56	88	144	176	208	224	256	328	344
2	32	72	144	176	208	256	296	328	376	424
3	40	104	176	208	256	328	392	440	504	568
4	56	120	208	256	328	408	488	552	632	696
5	72	144	224	328	424	504	600	680	776	872
6	328	176	256	392	504	600	712	808	936	1032
7	104	224	328	472	584	712	840	968	1096	1224
8	120	256	392	536	680	808	968	1096	1256	1384
9	136	296	456	616	776	936	1096	1256	1416	1544
10	144	328	504	680	872	1032	1224	1384	1544	1736
11	176	376	584	776	1000	1192	1384	1608	1800	2024
12	208	440	680	904	1128	1352	1608	1800	2024	2280
13	224	488	744	1000	1256	1544	1800	2024	2280	2536
14	256	552	840	1128	1416	1736	1992	2280	2600	2856
15	280	600	904	1224	1544	1800	2152	2472	2728	3112
16	328	632	968	1288	1608	1928	2280	2600	2984	3240
17	336	696	1064	1416	1800	2152	2536	2856	3240	3624
18	376	776	1160	1544	1992	2344	2792	3112	3624	4008
19	408	840	1288	1736	2152	2600	2984	3496	3880	4264
20	440	904	1384	1864	2344	2792	3240	3752	4136	4584
21	488	1000	1480	1992	2472	2984	3496	4008	4584	4968
22	520	1064	1608	2152	2664	3240	3752	4264	4776	5352
23	552	1128	1736	2280	2856	3496	4008	4584	5160	5736
24	584	1192	1800	2408	2984	3624	4264	4968	5544	5992
25	616	1256	1864	2536	3112	3752	4392	5160	5736	6200
26	712	1480	2216	2984	3752	4392	5160	5992	6712	7480

Note. Reprinted from "3GPP TS 36.213 Rel.15, Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures," Sept. 2018.

same base station. UE A placed a load on the LTE network by communicating with a server. While changing the LTE load, we estimated the RB usage rate of the base station on UE B and available throughput between UE B and the base station. The UE model used for UE A or UE B was Sony Xperia with Android 4.2. The server was connected directly to the experimental LTE network.

Both UEs used in this experiment were set up for directly obtaining low-level radio condition metrics from wireless equipment every 10 ms by using a dedicated application interface (API) provided by Qualcomm. This dedicated API is compatible with Qualcomm eXtensible Diagnostic Monitor (QxDM) [31]. A server ran an iperf server for sending user datagram protocol (UDP) traffic on Ubuntu 14.04.

We developed an experimental LTE network consisting of two UEs (UE A and UE B), an eNodeB, and evolved packet core. The specifications of the experimental LTE network are listed in Table 6. When we used the LTE testbed, we covered the eNodeB and UEs with an electromagnetic shield to prevent the radio waves of the experimental LTE network from interfering with other radio waves. Therefore, only our experimental UEs were attached to the eNodeB.

We placed two load patterns by the server sending UDP packets to UE A. One was gradual traffic-load increase and the other was random load variation.

For the gradual traffic-load increase, the server sent UDP packets for 10 s after the server stopped sending packets for 10 s. Then the server restarted sending UDP packets for 10 s. The packet-sending rate increased 7 Mbit/s every time the server restarted sending and the packet-sending rate was from 7 Mbit/s to 35 Mbit/s. This means that the base-station load increased from 20 to 100%. Then, UE B estimated the RB usage rate according to the measured RSRQ and SINR.

When using the random traffic variation, the server changed the UDP sending rate from 8.75 to 26.25 Mbit/s every second. This means that the base-station load randomly changed between 25 and 75% every second.

We compared the actual and estimated RB usage rate using three metrics: the correlation coefficient r , mean absolute error (MAE), and absolute value of residuals (AVR).

The r is a metric that shows the strength of the relationship between two variables. The r between x_i and y_i ($i = 1, 2, \dots, N$) is defined as

$$\begin{aligned}
 r &= \frac{s_{xy}}{s_x s_y}, \\
 &= \frac{1/n \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{1/n \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{1/n \sum_{i=1}^n (y_i - \bar{y})^2}}
 \end{aligned} \tag{6}$$

where s_{xy} , s_x , and s_y are the covariance of x and y , standard deviation of x , and standard deviation of y , respectively.

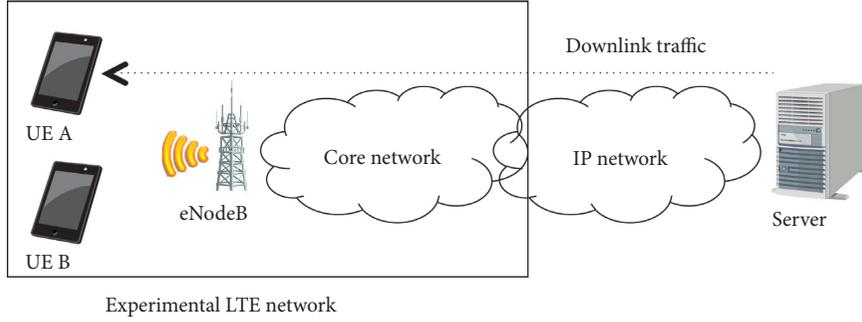


FIGURE 4: Experimental environment.

TABLE 6: Specifications of experimental LTE network.

EPC	Conforming to 3GPP Rel. 11
eNodeB	Conforming to 3GPP Rel. 11
Operation band	2.66 GHz (band 7)
Channel bandwidth	10 MHz
Transmission mode	TM 4 (closed-loop spatial multiplexing)
Maximum throughput	35 Mbit/s

The MAE is a metric that shows the average difference between the actual and estimated values and is defined as

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i|, \quad (7)$$

where x_i and \hat{x}_i are the actual and the estimated values, respectively.

The AVR is the ratio of absolute residual error to the actual value. The AVR_i is defined as

$$\text{AVR}_i = \frac{|x_i - \hat{x}_i|}{x_i}, \quad (8)$$

where x_i and \hat{x}_i are the actual and the estimated values, respectively. The mean absolute percentage error (MAPE) is the average AVR. Since the MAE is dependent on the actual value and is not scaled, MAPE is the relative error to the actual value. To avoid dividing by zero, the minimum x_i and \hat{x}_i were limited to 0.01 when we calculated the AVR.

5.1.2. Estimating RB Usage Rate for Gradual Traffic-Load Increase. The time variation of the RB usage rate for gradual traffic-load increase is shown in Figure 5. Note that “actual” and “estimated” show the actual RB usage rate measured at the eNodeB and estimated RB usage rate at UE B, respectively, and the RB usage rate was averaged by 1 s. This figure shows that estimation of a base-station load follows the UDP rate. The r was 0.898 and MAE was 0.058. The cumulative distribution function (CDF) of the AVR for estimating the RB usage rate for gradual traffic-load increase is shown in Figure 6. The proposed method estimated the RB usage rate with an AVR less than 0.1 for almost 80% estimation. The MAPE for this part of the experiment was 0.779.

5.1.3. RB-Usage-Rate Estimation for Random Load Variation. The time variation of the RB usage rate for

random load variation is shown in Figure 7. This figure also shows that estimating base-station load follows the UDP rate. The r was 0.861 and MAE was 0.085. Figure 8 shows the CDF of the AVR for random load variation. The proposed method estimated the RB usage rate with an AVR less than 0.3 for more than 80% estimation. The MAPE for this part of the experiment was 0.181.

5.1.4. Results for RB-Usage-Rate Estimation. These verification experiments showed that a UE can accurately estimate the RB usage rate of the connected cell from the measured RSRQ and SINR. However, the MAPE for gradual traffic-load increase was relatively high, although residuals for almost 80% estimation were less than 0.1. This is because the error ratio was high when the actual RB usage rate was close to 0.

5.2. Evaluation of Available Throughput Estimation in Commercial LTE Network

5.2.1. Experimental Setting. We experimentally verified the accuracy of the estimated available throughput in the commercial LTE system shown in Figure 9.

UE A placed a load on the LTE network by communicating with a server. While changing the LTE load, we estimated the RB usage rate of the eNodeB on UE B and available throughput between UE B and the eNodeB.

The bandwidth and number of RBs we used for this LTE network were 10 MHz and 50, respectively. Both UEs were connected to the same serving cell with two transmission antennas. The model used for both UEs was Samsung Galaxy S4 with Android 4.2.2. These UEs were also set up for directly obtaining low-level radio condition metrics from wireless equipment every 10 ms by using a dedicated API provided by Qualcomm. The server was located in the Internet and there were network operator core networks, which were operated by Internet service providers between the server and the eNodeB. This server is identical to that in the previous experiment which ran an iperf server for sending UDP traffic on Ubuntu 14.04.

Accurate available throughput was not available since we used a commercial LTE network. Therefore, we measured the downloaded data rate in the following manner and used it as the correct available throughput.

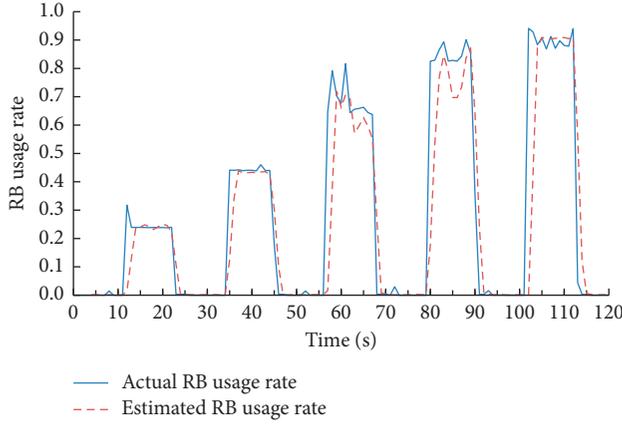


FIGURE 5: RB usage rate for gradual load increase.

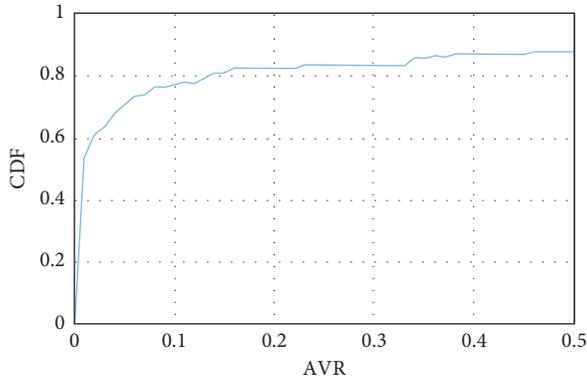


FIGURE 6: CDF of AVR for estimating RB usage rate for gradual load increase.

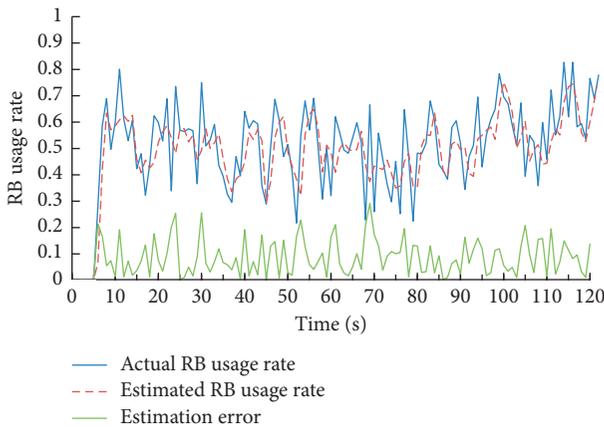


FIGURE 7: RB usage rate for random load variation.

- (i) The UE B estimated the available throughput.
- (ii) The server sent 50 Mbit/s of UDP data for 1 s just after the UE B finished estimation.
- (iii) The data size the UE B received in 1 s was considered the “correct available throughput.”

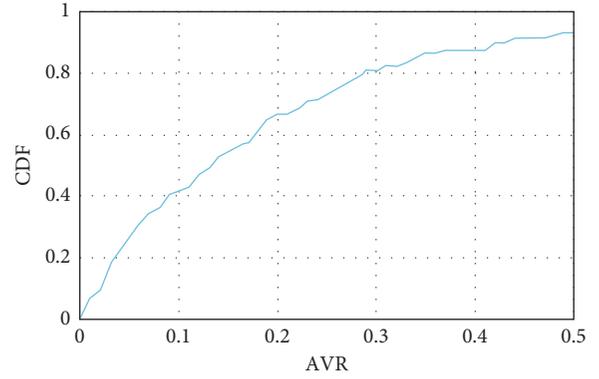


FIGURE 8: CDF of AVR for estimating RB usage rate for random load variation.

Note that we assumed that the available throughput the UE estimated was similar to that just after estimation. The difference between them should be considered for future work.

The UE B estimated the available throughput from the radio metrics every 3 s. The accuracy of the estimated available throughput was analyzed by comparing the estimated available throughput and measured downlink throughput.

5.2.2. Experimental Results. The experimental results are shown in Figure 10. “Downlink throughput” denotes the correct available throughput measured at the UE B, and the “estimated available throughput” denotes the estimated available throughput by using the proposed method. The r was 0.719 and MAE was 6.97. Figure 10 shows that the estimated available throughput was overestimated. When we used $0.82 \times$ the estimated available throughput as the estimated value, the MAE improved to 3.70. We assume that this is because the actual throughput decreases in accordance with the overhead of the header and data retransmission.

Figure 11 shows the CDF of the AVR for estimating available throughput. The solid and dashed lines are for the estimated available throughput and for $0.82 \times$ the estimated available throughput, respectively. The MAPEs for the estimated available throughput and $0.82 \times$ the estimated available throughput were 0.273 and 0.143, respectively. This shows that the proposed method estimated the available throughput with an AVR less than 0.4 for more than 80% estimation. It is also possible for the proposed method to improve estimation accuracy of available throughput with an AVR less than 0.2 for more than 80% estimation.

We also evaluated estimation accuracy for high and low throughput. We categorized the available throughput into high throughput and low throughput according to whether the actual throughput was higher than 20 Mbps. The MAPEs for the $0.82 \times$ the estimated high throughput and low throughput were 0.109 and 0.417, respectively. This means the proposed method can estimate available throughput for high throughput more accurately than for low throughput. For future work, we plan to improve estimation accuracy in low

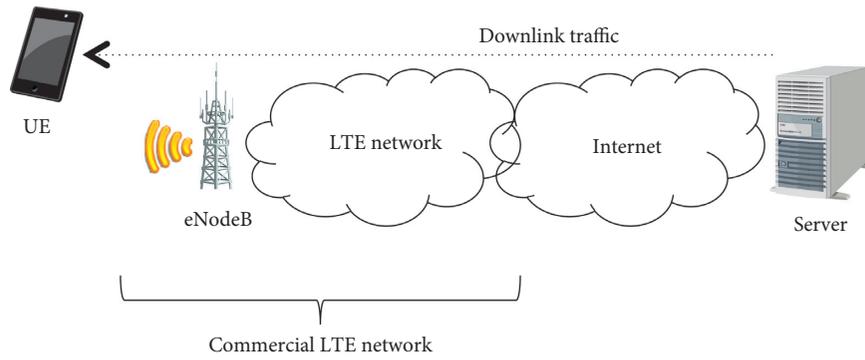


FIGURE 9: Evaluation of available throughput estimation.

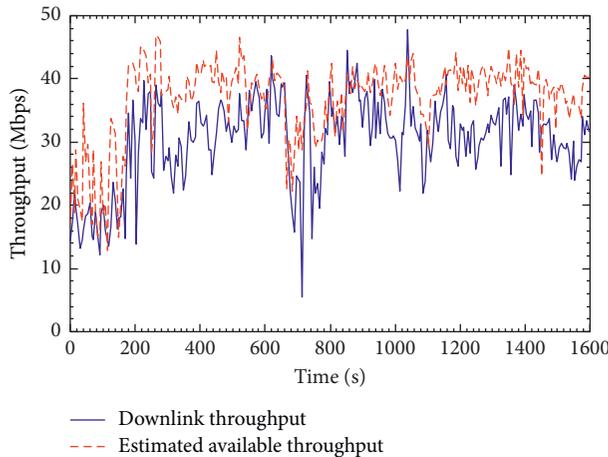


FIGURE 10: Accuracy of estimated available throughput.

throughput, because applications should be adaptively controlled to improve their quality especially in low throughput.

6. Off-Peak Data Transfer as Application of Passive Estimation of RB Usage Rate and Available Throughput

As mentioned earlier, off-peak data transfer system can be realized with available throughput estimation. When non-delay-sensitive traffic during a peak-load period can be shifted to a non-peak-load period, users using delay-sensitive traffic experience higher quality, even though this traffic shift does not deteriorate non-delay-sensitive-traffic user experience.

In our previous study, we showed the possibility of realizing off-peak transfer system by measuring the RB usage rate in commercial LTE networks at four areas in Tokyo [4]. We reported that off-peak came every several 10s and continued for several tens of seconds even in congested area. Our results indicated that there is room where the off-peak data transfer can be put into effect even in congested area.

In the previous study, we evaluated an off-peak transfer system using only the RB usage rate. In addition, available throughput estimation of our current study is expected to improve the efficiency of radio wave utilization.

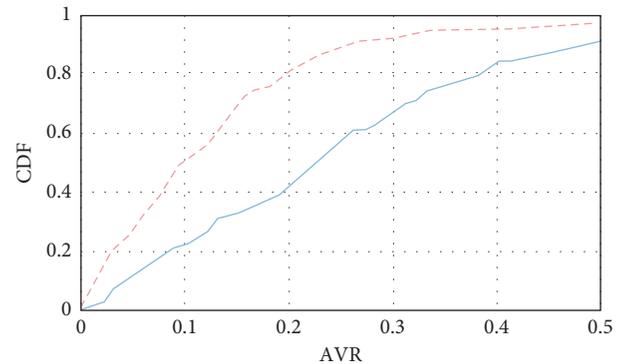


FIGURE 11: CDF of AVR for available throughput estimation.

UE Implementation In this subsection, we discuss implementation of an off-peak data transfer system on a UE. An off-peak data transfer system runs either on a normal OS, such as Android or iOS, or on a custom OS, either as an application or daemon software. For easier implementation and deployment, we used an off-peak data transfer system running on a normal OS as an application that delays receiving non-delay-sensitive content. However, it is difficult for the application on normal OS to estimate base-station load and available throughput. We used UEs that can obtain low-level radio quality every 10 ms. Low-level radio quality is not available for applications on current normal OSes.

As Yue et al. reported the difference between the dedicated API and Android API [20], we also observed a similar difference. Rather than a dedicated API such as QxDM, a standard API provided with a normal OS for obtaining low-level radio quality returns radio quality metrics less frequently only when the radio quality changes. The radio quality from a standard API is less accurate than that from a dedicated API since a standard API returns an integer-type value, not a floating-point type value, which a dedicated API returns.

Therefore, the estimation accuracy must be lower using only a standard API. We can show the possibility that base-station load and available throughput can be estimated using only the metrics a UE acquires. When a standard API can return more accurate radio quality more frequently, off-peak data transfer can be easily implemented on UEs.

7. Conclusion

We proposed and evaluated a method for passively estimating the available bandwidth between a base station and UE. Experiments on experimental and commercial LTE networks confirm that available throughput can be estimated passively and accurately by using the RSRQ, SINR, and CQI measured at a UE; correlation with actual throughput was more than 0.7. We argued that this method may reduce base-station load in an off-peak data transfer system, by which a UE autonomously shifts a peak load on a base station. An off-peak data transfer system can benefit both consumers and network operators in which UEs are smart in shifting their communication data based on information acquired indirectly by UEs. Now, 5G is expected to accommodate large mobile data demand. Since 5G is based on a similar idea regarding RB structure, the basic idea of the proposed method can be applied to estimate 5G base-station load. In addition, LTE/4G and older networks will still occupy around 90% of total mobile subscriptions at the end of 2023 [1]. The proposed method can mitigate the peak load of a base station and improve communication quality without compromising the QoE provided to the end user.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] Ericsson, "Ericsson mobility report," Ericsson, Stockholm, Sweden, 2017.
- [2] C. Shi, K. Joshi, R. K. Panta, M. H. Ammar, and E. W. Zegura, "CoAST: collaborative application-aware scheduling of last-mile cellular traffic," in *Proceedings of the 12th Annual International Conference on Mobile systems, Applications, and Services—MobiSys '14*, pp. 245–258, Bretton Woods, NH, USA, June 2014.
- [3] Cisco, "Cisco visual networking index: global mobile data traffic forecast update 2014–2019," 2015.
- [4] E. Takahashi, T. Suzuki, T. Onishi, and K. Satoda, "Autonomous off-peak data transfer by passively estimating overall LTE cell load," in *Proceedings of the 14th Annual IEEE Consumer Communications & Networking Conference (CCNC 2017)*, Las Vegas, NV, USA, January 2017.
- [5] N. Sato, T. Oshiba, K. Nogami, A. Sawabe, and K. Satoda, "Experimental comparison of machine learning-based available bandwidth estimation methods over operational LTE networks," in *Proceedings of the 2017 IEEE Symposium on Computers and Communications (ISCC)*, pp. 339–346, Heraklion, Greece, July 2017.
- [6] T. Kitahara, Y. Hiehata, H. Koto, N. Fukumoto, H. Nakamura, and S. Ano, "Autonomous data transmission using off-peak hours for cellular-based M2M communications," in *Proceedings of the 2013 IEEE 37th Annual Computer Software and Applications Conference*, pp. 167–168, Kyoto, Japan, July 2013.
- [7] E. Takahashi, T. Onishi, T. Shiroshima, and H. Sugahara, "A passive method for estimating overall LTE cell traffic load by using radio quality information of a user terminal," in *Proceedings of the 2015 21st Asia-Pacific Conference on Communications (APCC)*, pp. 17–20, Kyoto, Japan, October 2015.
- [8] C. Dovrolis, P. Ramanathan, and D. Moore, "Packet-dispersion techniques and a capacity-estimation methodology," *IEEE/ACM Transactions on Networking*, vol. 12, no. 6, 2004.
- [9] N. Hu and P. Steenkiste, "Evaluation and characterization of available bandwidth probing techniques," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 6, pp. 879–894, 2003.
- [10] M. Jain and C. Dovrolis, "End-to-end available bandwidth: measurement methodology, dynamics, and relation with TCP throughput," in *Proceedings of the 2002 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications—SIGCOMM '02*, Pittsburgh, PA, USA, August 2002.
- [11] J. Strauss, D. Katabi, and F. Kaashoek, "A measurement study of available bandwidth estimation tools," in *Proceedings of the 2003 ACM SIGCOMM conference on Internet measurement—IMC '03*, Miami Beach FL USA, October 2003.
- [12] K. Lakshminarayanan, V. N. Padmanabhan, and J. Padhye, "Bandwidth estimation in broadband access networks," in *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement—IMC '04*, Taormina, Sicily, Italy, October 2004.
- [13] M. Li, M. Claypool, and R. Kinicki, "Wbest: a bandwidth estimation tool for IEEE 802.11 wireless networks," in *Proceedings of the 2008 33rd IEEE Conference on Local Computer Networks (LCN)*, Montreal, Canada, October 2008.
- [14] G. Aceto, F. Palumbo, V. Persico, and A. Pescape, "Available bandwidth vs. achievable throughput measurements in 4G mobile networks," in *Proceedings of the 2018 14th International Conference on Network and Service Management (CNSM)*, Rome, Italy, November 2018.
- [15] A. Chakraborty, V. Navda, V. N. Padmanabhan, and R. Ramjee, "Coordinating cellular background transfers using loadsense," in *Proceedings of the 19th Annual International Conference on Mobile computing & Networking—MobiCom'13*, pp. 63–74, Miami, FL, USA, September 2013.
- [16] J. Caine, B. Gill, S. Johnston, J. Robinson, and S. Westwood, "Modelling download throughput of LTE networks," in *Proceedings of the 39th Annual IEEE Conference on Local Computer Networks Workshops IEEE P2MNET*, Edmonton, AB, Canada, September 2014.
- [17] F. Pervez, M. Jaber, J. Qadir, S. Younis, and M. A. Imran, "Fuzzy Q-learning-based user-centric backhaul-aware user cell association scheme," in *Proceedings of the 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pp. 1840–1845, Valencia, Spain, June 2017.
- [18] A. V. Iordache and I. Marghescu, "LTE downlink performance," in *Proceedings of the 2012 10th International Symposium on Electronics and Telecommunications*, pp. 179–182, Timisoara, Romania, November 2012.
- [19] 3GPP TS 36.213 Rel.15, Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures, 2018.
- [20] C. Yue, R. Jin, K. Suh, Y. Qin, B. Wang, and W. Wei, "Linkforecast: cellular link bandwidth prediction in LTE networks," *IEEE Transactions on Mobile Computing*, vol. 17, no. 7, pp. 1582–1594, 2018.
- [21] D. Raca, A. H. Zahran, C. J. Sreenan et al., "Back to the future: throughput prediction for cellular networks using radio KPIs," in *Proceedings of the 4th ACM Workshop on Hot Topics in Wireless (HotWireless)*, Snowbird, UT, USA, October 2017.

- [22] D. Raca, A. H. Zahran, C. J. Sreenan et al., "Empowering video players in cellular: throughput prediction from radio network measurements," in *Proceedings of the ACM Multimedia Systems Conference (MMSys)*, Amherst, MA, USA, June 2019.
- [23] R. Falkenberg, K. Heimann, and C. Wietfeld, "Discover your competition in LTE: client-based passive data rate prediction by machine learning," in *Proceedings of the GLOBECOM 2017-2017 IEEE Global Communications Conference*, Singapore, December 2017.
- [24] G. Nikolov, M. Kuhn, and B.-L. Wenning, "UE-based estimation of available uplink data rates in cellular networks," in *Proceedings of the 2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Limassol, Cyprus, October 2018.
- [25] J. Ding, Y. Li, and D. Jin, "Characterizing the phenomenon of traffic tide for large-scale mobile cellular data networks," in *Proceedings of the 2015 IEEE Conference on Computer Communications Workshops*, pp. 45-46, Hong Kong, China, April 2015.
- [26] RIPE Network Coordination Center, *Internet Traffic During the 2012 Olympics*, RIPE Network Coordination Center, Amsterdam, Netherlands <https://labs.ripe.net/Members/fergalc/internet-traffic-during-olympics-2012>.
- [27] 3GPP TS. 36.214 Rel.15, evolved universal terrestrial radio access (E-UTRA); physical layer; Measurements, 2018.
- [28] M. T. Kawser, N. I. B. Hamid, M. N. Hasan, M. S. Alam, and M. M. Rahman, "Downlink SNR to CQI mapping for different multiple antenna techniques in LTE," *International Journal of Information and Electronics Engineering*, vol. 2, no. 5, pp. 757-760, 2012.
- [29] S. N. Donthi and N. B. Mehta, "An accurate model for EESM and its application to analysis of CQI feedback schemes and scheduling in LTE," *IEEE Transactions on Wireless Communications*, vol. 10, no. 10, pp. 3436-3448, 2011.
- [30] I. Siaud and A.-M. Ulmer-Moll, "Green-oriented multi-techno link adaptation metrics for 5G heterogeneous networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, 2016.
- [31] Qualcomm, "Qualcomm extensible diagnostic monitor (QXDM professional)," in *4G Mobile Communications*, Wiley, Hoboken, NJ, USA, 2013, <https://www.qualcomm.com/documents/qxdm-professional-qualcomm-extensible-diagnostic-monitor>.