공학석사 학위논문

셀룰러 네트워크에서 YouTube 서비스를 위한 자원 할당 방법

Resource Allocation Scheme for YouTube Service in Cellular Networks

2015년 2월

서울대학교 대학원

전기 · 컴퓨터 공학부

허 원 선

셀룰러 네트워크에서 YouTube 서비스를 위한 자원 할당 방법

Resource Allocation Scheme for YouTube Service in Cellular Networks

지도교수 전 화 숙

이 논문을 공학석사 학위논문으로 제출함

2015 년 1 월

서울대학교 대학원

전기 · 컴퓨터 공학부

허 원 선

허원선의 공학석사 학위논문을 인준함

2015 년 1 월

위 원 장 : ______________________ (인)

부위원장 : ______________________ (인)

위 원 : ______________________ (인)

Abstract

Due to the growth of network capacity and computing power, many multimedia services such as video on demand (VOD) and video streaming have appeared. To accommodate those services in cellular networks, an efficient radio resource management scheme that considers traffic characteristics is needed.

Especially, among multimedia services, video streaming traffic has distinct characteristics. That is, video streaming does not require user to download all data composing video clip before playback. To identify the properties of video streaming traffic more precisely, many researchers have investigated YouTube, which is one of the most famous social networking services that carries a lot of video streaming traffic. Although there have been many studies on YouTube traffic itself in the literature, few studies have tried to apply those findings into radio resource allocation.

In this paper, we propose a resource allocation scheme for the YouTube service in cellular networks that maximizes throughput while maintaining client buffer size in a predefined range. For this purpose, we use minimum rate requirement adaptation as an indirect method to adjust the buffer sizes of clients. Proposed scheme determines the minimum rate requirement of each user at each time slot by the information reported from each user such as buffer amount of the user and received data. Simulation results show that our resource allocation algorithm outperforms existing resource allocation schemes in terms of the QoS satisfaction of users by using average user satisfaction ratio and the number of pauses during playback.

keywords : Cellular network, radio resource management, quality of service

student number : 2013-20906

Contents

Abstract **i**

Contents **ii**

List of Figures **iii**

List of Tables **iv**

Chapter 1 Introduction **1**

Chapter 2 System model **4**

Chapter 3 Proposed Resource Allocation Scheme **6**

Chapter 4 Simulation Results **12**

Chapter 5 Conclusion **20**

Bibliography **21**

Abstract in Korean **23**

List of Figures

Figure 1. Proposed minimum rate requirement adaptation flow. 7

Figure 2. Average user satisfaction ratio according to frames. 15

Figure 3. User snapshot of instantaneous rate and minimum rate requirement, according to frames. 16

Figure 4. User snapshot of client buffer amount and minimum rate requirement, according to frames. 17

List of Tables

Table 1. Simulation parameters. 13

Table 2. Minimum rate requirement adaptation parameters. 13

Table 3. Total user throughput and average standard deviation of user rates. 18

Table 4. Number of pauses and total pause duration. 19

Chapter 1

Introduction

Due to the growth of network capacity and computing power, many multimedia services such as video on demand (VOD) and video streaming have appeared. To accommodate those services in cellular networks, an efficient radio resource management scheme that considers traffic characteristics is needed.

There have been radio resource allocation schemes to support different traffic types in cellular networks. Specifically in [1], a general resource allocation scheme for multimedia services in orthogonal frequency division multiple access (OFDMA) networks was proposed. Most research including [1] have treated multimedia services as a data traffic which has requirements of rate and delay. However, among multimedia services, video streaming traffic has distinct characteristics. That is, video streaming does not require user to download all data composing video clip before playback. To identify the properties of video streaming traffic more precisely, many researchers have investigated YouTube, which is one of the most famous social networking services that carries a lot of video streaming traffic. Several publications [2] have revealed that the YouTube server performs transmission rate adaptation, which is composed of two phases: initial burst and throttling. In the initial burst phase, the server transmits video data at the end-to-end maximum rate between the server and client to insert enough data into client's buffer. After that, a reduced data rate is applied. Also, authors in [3] investigated YouTube streaming traffic in a 3G environment. They identified that in a 3G environment, because of the channel variation, servers transmit data at a higher rate than in a wired environment. Furthermore, they found that YouTube clients exploit several buffer management policies, which aim to maintain a client's buffer in a considerable amount.

Although there have been many studies on YouTube traffic itself in the literature, few studies have tried to apply those findings into radio resource allocation. J. Navarro-Ortiz et.al. [4] presented a quality of experience (QoE) based scheduler for OFDMA systems for the purpose of minimizing pauses during the video playback. However, they did not consider the transmission rate adaptation of YouTube servers.

In this paper, we propose a resource allocation scheme for the YouTube service in cellular networks that maximizes throughput while maintaining client buffer size in a predefined range. For this purpose, we use minimum rate requirement adaptation as an indirect method to adjust the buffer sizes of clients. Proposed scheme determines the minimum rate requirement of each user at each time slot by the information reported from each user such as buffer amount of the user and received data. Also, this scheme can accommodate as much users as possible when channel condition is bad by dividing minimum rate requirements into several levels. Simulation results show that our resource allocation algorithm outperforms existing resource allocation schemes in terms of the QoS satisfaction of users.

This thesis is organized as follows: Chapter 2 introduces the system model we considered, Chapter 3 explains the main idea of proposed resource allocation scheme, Chapter 4 presents simulation setup and numerical results, and Chapter 5 concludes this thesis.

Chapter 2

System Model

We consider the downlink of an OFDMA system which is composed of one macro base station (macro BS) and six neighboring BSs. Also, we assume each BS serves N users. Time is slotted with a duration T_s, which is equivalent to the coherent time. The system bandwidth W is divided by K subchannels. The interference from neighboring BSs is denoted by I. Then the achievable data rate to user n at the subchannel k in time slot t is given by

$$r_{n,k}(t) = \frac{W}{K}\log\left(1 + \frac{\eta\left|h_{n,k}\right|^2 p_{n,k}}{\sigma^2 + I}\right)$$

where η is the signal-to-noise (SNR) gap between Shannon capacity and practical capacity when adaptive modulation and coding (AMC) scheme is used[5], $h_{n,k}$ is the channel gain which can be calculated as $\phi_{\mathrm{x}} D^{-a_{\mathrm{x}}} G_{\mathrm{x}}$, where the fixed channel gain is denoted by ϕ_{x}, the distance from a user to cell is denoted by D, path loss exponent is denoted by a_{x}, and G_{x} is used to model the lognormal shadowing and multipath fading with composite random variable [6], $p_{n,k}$ is the power allocated to user n at the subchannel k, and σ is the noise power over a subchannel.

We divide the traffic of users into two types: YouTube-H and YouTube-L. YouTube-H denotes high quality video clips with encoding rates of 256kbps, and YouTube-L indicates low quality video clips with encoding rates of 64kbps. Although each user can have multiple traffic demands, each user is assumed to have one type of QoS requirement for simplicity. Each user reports the following information to its BS: bitrate of the video, amount of received data, and amount of buffer. Bitrate of the video is defined as the amount of data dedicated to a second of video clip in bits.

Chapter 3

Proposed Resource Allocation Scheme

3.1 Minimum Rate Requirement Adaptation

To support user's QoS in cellular networks, several approaches can be applied. One of them is to use delay as a constraint. This approach may be useful because QoS is usually defined using delay requirements. However, the delay of a user could not be calculated accurately in general cellular networks, although the calculation has significant computing overhead. Instead, we use minimum rate requirement adaptation as an indirect approach to support QoS in cellular networks. By adjusting the buffer sizes of clients, users does not experience any pauses in the video playback. That is, QoS of user can be improved. Additionally, minimum rate requirement can be easily added into resource allocation problem.

Figure 1 shows the proposed minimum rate requirement adaptation for the YouTube service. The value inside the circle represents the minimum rate requirement. By adapting minimum rate requirement, number of subchannels allocated to each user is changed. The minimum rate is determined by information reported by the clients such as client buffer amount and client received data. The values of *RX* can be set arbitrary rates which have descending order and the number of *RX*s can also be adjusted according to channel conditions in order to maintain the buffer size between 40 s and 110 s, as measured in a 3G environment [3]. By utilizing several levels of adaptation, this scheme can accommodate as much users as possible in many different situations.

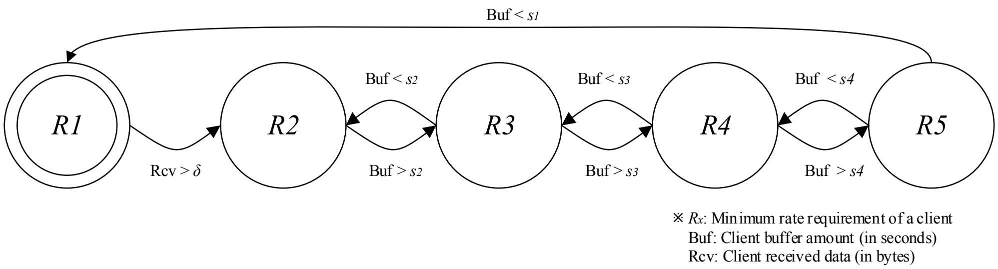

Figure 1. Proposed minimum rate requirement adaptation flow.

3.2 Resource Allocation with Minimum Rate Requirement

The subcarrier and power allocation with minimum rate requirement is formulated as:

$$\max_{S} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{n,k}(t) s_{n,k}(t)$$

Subject to

$$\sum_{n=1}^{N} s_{n,k}(t) = 1, \forall k$$

$$s_{n,k}(t) \in \{0,1\}$$

$$\sum_{k=1}^{K} r_{n,k}(t) \geq R_n, \forall n$$

$$\sum_{n=1}^{N} \sum_{k=1}^{K} p_{n,k}(t) s_{n,k}(t) \leq P_m$$

where R_n is the minimum rate requirement determined for each user in every frame and P_m is the maximum power in a cell. By solving this problem, we can obtain the subcarrier allocation matrix S and power allocation matrix P. Given optimization problem is solved by Lagrangian Dual method. First, given primal problem is decomposed using lagrangian multipliers $\boldsymbol{\lambda}$ and μ as:

$$\mathcal{L}(\boldsymbol{p}, \boldsymbol{s}, \boldsymbol{\lambda}, \mu) = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{n,k}(t) s_{n,k}(t) + \sum_{n=1}^{N} \lambda_n \left(\sum_{k=1}^{K} r_{n,k}(t) - R_n \right)$$

$$+ \mu \left(P_m - \sum_{n=1}^{N} \sum_{k=1}^{K} p_{n,k}(t) s_{n,k}(t) \right)$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} r_{n,k}(t)\, s_{n,k}(t) + \sum_{n=1}^{N} \lambda_n \left(\sum_{k=1}^{K} r_{n,k}(t) - R_n \right)$$

$$- \mu \left(P_m - \sum_{n=1}^{N} \sum_{k=1}^{K} p_{n,k}(t) s_{n,k}(t) \right)$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} r_{n,k}(t)\, s_{n,k}(t) + \sum_{n=1}^{N} \lambda_n \sum_{k=1}^{K} r_{n,k}(t) - \mu \sum_{n=1}^{N} \sum_{k=1}^{K} p_{n,k}(t) s_{n,k}(t)$$

$$- \sum_{n=1}^{N} \lambda_n\, R_n + \mu P_m$$

$$\mathcal{L}(\boldsymbol{p}, \boldsymbol{s}, \boldsymbol{\lambda}, \mu) = \sum_{n=1}^{N} \sum_{k=1}^{K} \frac{W}{K} \log_2(1 + \eta p_{n,k}\, g_{n,k})\, s_{n,k}(t)$$

$$+ \sum_{n=1}^{N} \lambda_n \sum_{k=1}^{K} \frac{W}{K} \log_2(1 + \eta p_{n,k}\, g_{n,k}) s_{n,k}(t)$$

$$- \mu \sum_{n=1}^{N} \sum_{k=1}^{K} p_{n,k}(t) s_{n,k}(t) - \sum_{n=1}^{N} \lambda_n\, R_n + \mu P_m$$

From the partial derivative of g with respect to p being zero, optimal power allocation can be derived as:

$$p_{n,k} = \left[\frac{\frac{W}{K}(1+\lambda_n)}{\mu \ln 2} - \frac{1}{\eta g_{n,k}}\right]^+$$

where $[\mathrm{z}]^+ = \max\{\mathrm{z}, 0\}$.

By relaxing the value of $s_{n,k}(t)$ from 0 to 1, lagrangian dual function of given lagrangian function is given as:

$$g(\boldsymbol{\lambda}, \mu) = \max_{\boldsymbol{p},\boldsymbol{s}} \mathcal{L}(\boldsymbol{p}, \boldsymbol{s}, \boldsymbol{\lambda}, \mu)$$

Subject to

$$0 \le \sum_{n=1}^{N} s_{n,k}(t) \le 1$$

From this, dual problem of primal problem is derived as:

$$\min_{\boldsymbol{\lambda},\mu} g(\boldsymbol{\lambda}, \mu)$$

Subject to

$$\boldsymbol{\lambda} \succcurlyeq 0$$

$$\mu \ge 0$$

By continuously updating lagrangian multipliers with certain step size, the value of $\boldsymbol{\lambda}$ and μ converges and optimal solution $(\boldsymbol{\lambda}^*, \mu^*)$ can be found. Langrangian multiplier update for $\boldsymbol{\lambda}$ and μ is calculated as follows:

$$\lambda_i(n+1) = \left[\lambda_i(n) - \delta_{\lambda_i}(n)\left(\sum_{k=1}^{K} r_{i,k}(t) - R_i\right)\right]^+$$

$$\mu(n+1) = \left[\mu(n) - \delta_{\mu}(n)\left(P_m - \sum_{n=1}^{N}\sum_{k=1}^{K} p_{n,k}(t)s_{n,k}(t)\right)\right]^+$$

Where $\delta(n)$ is positive step size at iteration n.

Chapter 4

Performance Evaluation

We evaluated proposed resource allocation using custom java based routine. In the first experiment, number of adaptation levels is investigated. Followed by this, a user snapshot of instantaneous rate and buffer amount is presented. And total user throughput and average standard deviation of user rate is showed in terms of total available power and number of users. Finally, number of pauses during video playback is compared with typical throughput maximization scheme.

4.1 Simulation Setup

For the YouTube clips, video clips are ranged from 150 s to 1000 s, since the minimum rate adaptation process is observed more clearly in relatively long video clips. At each timeslot, buffer amount of clients are calculated. If a certain video clips finishes downloading, another video clip is requested. To obtain results in earlier frames, encoding rate of video clips are divided into 20,000. For node movement, the model in [7] is used. Other simulation parameters are listed in Table 1 and Table 2.

Table 1. Simulation parameters

Total Bandwidth	10 MHz
Number of Subcarriers	48
Cell Radius	500 m
User Speed	50 km/h
BER Required	10^{-3}
Path Loss	$31 + 40 \log d$
Shadowing	Log normal shadowing, mean = 0dB, SD = 8dB
Correlation distance	50 m
Multipath fading	Jakes spectrum model [8] with 8 taps

Table 2. Minimum rate requirement adaptation parameters

	$\boldsymbol{R1}$	$\boldsymbol{R2}$	$\boldsymbol{R3}$	$\boldsymbol{R4}$	$\boldsymbol{R5}$	$\boldsymbol{\delta}$	$\boldsymbol{s_1}$	$\boldsymbol{s_2}$	$\boldsymbol{s_3}$	$\boldsymbol{s_4}$
Low quality	500kbps	400kbps	200kbps	100kbps	50kbps	0.8MB	40s	50s	80s	110s
High quality	500kbps	500kbps	256kbps	128kbps	50kbps	1.28MB	40s	50s	80s	110s

4.2 Numerical Results

Figure 2 deploys the comparison result when the number of levels used in the minimum rate requirement adaptation. Type 1 uses 5 levels of minimum rate requirements as shown in Table 2. However, Type 2 uses 4 levels ($\boldsymbol{R1}, \boldsymbol{R2}, \boldsymbol{R4},$ and $\boldsymbol{R5}$) and Type 3 uses 3 levels ($\boldsymbol{R1}, \boldsymbol{R2},$ and $\boldsymbol{R5}$), respectively. In terms of total user throughput, three types of adaptation shows almost same result. However, when average user satisfaction ratio, defined as the user throughput divided by the minimum rate requirement [9], is used as a QoS metric, Type 1 shows the highest average user satisfaction ratio among three types. These findings indicate that as number of adaptation level increases, users get higher average user satisfaction ratio. Also, these results suggest that the perceived QoS of user can be improved by using higher levels of minimum rate requirements.

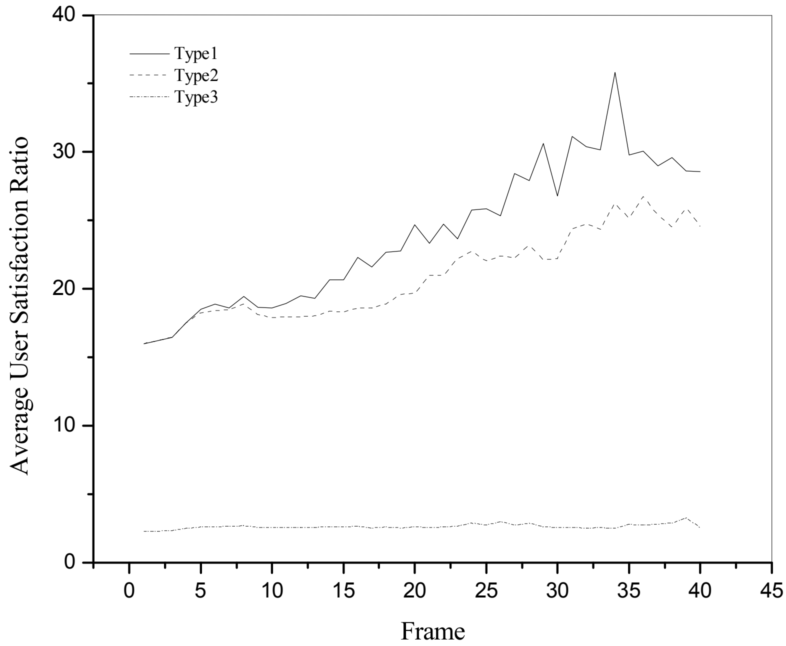

Figure 2. Average user satisfaction ratio according to frames.

In figure 3, a user snapshot of instantaneous rate and minimum rate requirements is depicted. As minimum rate requirement decreases, instantaneous rate shows the tendency to decrease. Also, instantaneous rate shows peaks at certain points. This may be explained by the variation of the channel conditions. Likewise, Figure 4 shows a user snapshot of client buffer amount and minimum rate requirement. Client buffer amount grows slowly when minimum rate requirement is low. Client buffer amount decreases when the received data is smaller than consumed data in slots.

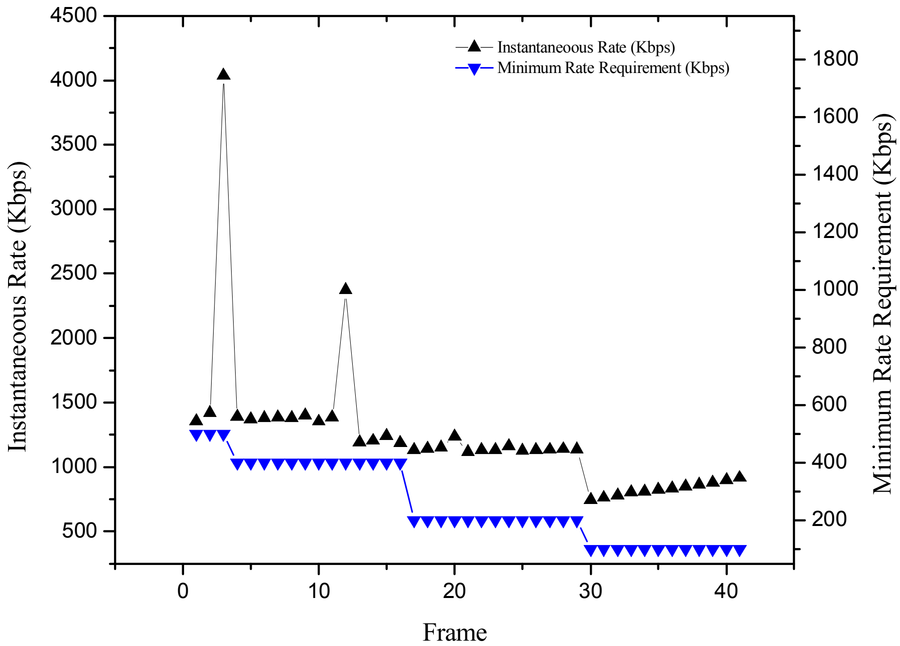

Figure 3. User snapshot of instantaneous rate and minimum rate requirement, according to frames.

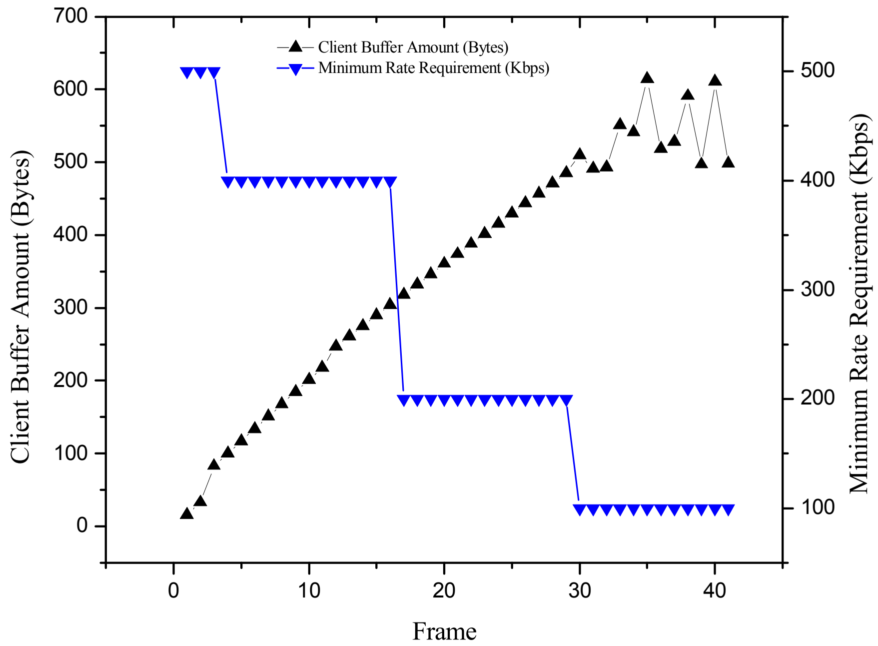

Figure 4. User snapshot of client buffer amount and minimum rate requirement, according to frames.

Table 3. Total user throughput and average standard deviation of user rates

P_m (dBm)	users in a cell	total user throughput (Mbps)	standard deviation of rates (Mbps)
38	40	274	0.98
	45	267	0.52
39	40	276	0.93
	45	269	0.53

Also, total user throughput and average standard deviation of user rates in terms of total available power in a cell (P_m) and number of users in a cell are presented in Table 3. When more power is provided, throughput increases accordingly. However, as proposed scheme is designed to guarantee minimum rate requirement and to use remaining radio resources to maximize throughput, average standard deviation of user rates also increases. Likewise, when less user is served in a cell, throughput increases and standard deviation of users decreases.

Table 4. Number of pauses and total pause duration

	No minimum rate req.	proposed
number of pauses	3	0
total pause duration (frames)	31	0

In Table 4, we compare our proposed scheme with a typical throughput maximization scheme. Typical throughput maximization scheme aims to maximize throughput, without minimum rate requirement. Number of pauses incremented when client buffer size is zero after the download started. Total pause duration is the sum of paused user frames. As shown in Table 4, in the experiment conducted for 40 frames, three pauses are observed and total pause duration is 31 frames. However, proposed scheme makes number of pauses into zero. From this result, it appears that proposed scheme is more suitable for QoS of YouTube users, since the number of pauses during playback can be used an indicator of perceived QoS.

Chapter 5

Conclusion

In this paper, we propose a resource allocation scheme for YouTube, which considers the rate adaptation policy of YouTube servers and the buffer management policy of YouTube clients. In the proposed scheme, the minimum rate requirement is determined in each timeslot, and the optimal resource allocation for maximizing throughput is presented using those constraints. Since many streaming services such as Netflix and Pandora radio apply similar strategies on data transmission as YouTube does, this approach can be utilized to improve the QoS of users.

Bibliography

[1] K. W. Choi, W. S. Jeon, and D. G. Jeong, “Resource allocation in OFDMA wireless communications systems supporting multimedia services,” *IEEE/ACM Trans. Netw.*, vol. 17, no. 3, pp. 926–935, Jun. 2009.

[2] S. Alcock and R. Nelson, “Application flow control in YouTube video streams,” *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 2, p. 24, Apr. 2011.

[3] J. Ramos-munoz, J. Prados-Garzon, P. Ameigeiras, J. Navarro-Ortiz, and J. Lopez-soler, “Characteristics of mobile youtube traffic,” *IEEE Wirel. Commun.*, vol. 21, no. 1, pp. 18–25, Feb. 2014.

[4] J. Navarro-Ortiz, P. Ameigeiras, J. M. Lopez-Soler, J. Lorca-Hernando, Q. Perez-Tarrero, and R. Garcia-Perez, “A QoE-aware scheduler for HTTP progressive video in OFDMA systems,” *IEEE Commun. Lett.*, vol. 17, no. 4, pp. 677–680, Apr. 2013.

[5] X. Qiu, K. Chawla, "On the performance of adaptive modulation in cellular systems," *Communications, IEEE Transactions on* , vol.47, no.6, pp.884,895, Jun 1999.

[6] W. S. Jeon, J. Kim, and D. G. Jeong, "Downlink radio resource partitioning with fractional frequency reuse in femtocell networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 1, pp. 308-321, Jan. 2014.

[7] W. S. Jeon, J. A. Han, and D. G. Jeong, "Distributed resource allocation for multi-cell relay-aided OFDMA systems," *IEEE Transactions on Mobile Computing*, vol. 13, no. 9, pp. 2003-2015, Sept. 2014.

[8] W. C. Jakes, Microwave Mobile Communications, *IEEE press*, 1994.

[9] H. M. Chaskar, U. Madhow, "Fair scheduling with tunable latency: a round robin approach," *Global Telecommunications Conference, 1999. GLOBECOM '99* , vol.2, no., pp.1328,1333 vol.2, 1999

요약

최근 네트워크 용량과 컴퓨팅 능력의 향상으로 인해 VOD나 비디오 스트리밍 서비스와 같은 많은 멀티미디어 서비스가 등장했고, 이러한 다양한 트래픽 요구사항에 맞는 무선 자원 할당 방법이 주목을 받고 있다. 특히, 멀티미디어 서비스 중에서도 비디오 스트리밍 서비스는 구분되는 특성을 가지고 있다. 즉, 비디오 스트리밍 서비스에서는 재생 전에 사용자가 동영상을 구성하는 모든 데이터를 다운로드할 필요가 없다는 것이다. 이러한 스트리밍 서비스의 특성을 확인하기 위해 많은 연구자들이 모바일 트래픽의 많은 부분을 차지하는 YouTube 서비스에 대해 연구하였으나, 이러한 특성을 실제 무선 자원 관리에 적용하는 기법에 대한 연구는 거의 이루어지지 않았다. 따라서 본 논문에서는 셀룰러 네트워크에서 YouTube 트래픽의 특성을 고려하여, 주파수 효율을 최대화하면서 각 사용자의 버퍼 크기를 일정 수준으로 유지하는 무선 자원 관리 기법에 대해 제안한다. 제안하는 무선 자원 관리 기법은 매 타임슬롯마다 각 사용자로부터 받은 버퍼 크기, 전송된 데이터 양 등의 정보를 바탕으로 각 사용자의 최소 전송 속도 요구치를 결정한다. 제안하는 무선 자원 기법은 사용자의 QoS 측면에서 기존의 기법들보다 우수한 성능을 보임을 평균 사용자 만족도와 재생 중 일시 정지 횟수를 통하여 확인하였다.

주요어 : 셀룰러 네트워크, 무선 자원 관리, QoS
학 번 : 2013-20906