We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Scheduling and Capacity of VoIP Services in Wireless OFDMA Systems

Jaewoo So
Sogang University
Republic of Korea

1. Introduction

The voice over Internet protocol (VoIP) service is widely supported in wireless orthogonal frequency division multiple access (OFDMA) systems such as a mobile worldwide interoperability for microwave access (WiMAX) system and a long term evolution (LTE) system. In wireless OFDMA systems, a base station (BS) broadcasts information to users about new resource assignments for every frame, where each resource is represented by time symbols and subchannels (IEEE, 2009; Ghosh et al., 2005). The representations of the allocated resources are usually broadcast at the level of a low modulation and coding scheme (MCS) because the BS must ensure that all users can receive the signaling information. The allocation process generates a substantial signaling overhead that influences the system resource utilization. In particular, the performance of VoIP services is seriously affected by the signaling overhead because of following reasons: First, the amount of signaling information is too large compared with the small-sized VoIP packets. Second, the symmetry between the downlink and uplink causes immense downlink overheads. Third, a BS may periodically allocate resources to VoIP users because the voice traffic are periodically generated and the voice traffic is delay sensitive.

In OFDMA-based systems such as IEEE 802.16e/m or 3GPP LTE, a BS allocates resources to users on a frame-by-frame basis and does not remember allocation information from one frame to next. This type of scheduling is referred to as *dynamic scheduling*. Dynamic scheduling allows the BS to schedule each frame independently. However, the signaling overhead increases with the increase of users that are served in the frame. As a means of reducing the signaling overhead, *persistent scheduling* has been proposed for VoIP services which has a periodic traffic pattern and a relatively fixed payload size. The persistent scheduling allows a BS to allocate resources persistently for multiple frames and therefore the BS can reduce the signaling overhead by obviating the need to send signaling information in every frame. The IEEE 802Rev2, the IEEE 802.16m and the 3GPP LTE standards support the persistent scheduling for efficient VoIP services.

Many researchers have evaluated the performance of VoIP services in wireless OFDMA systems. In (Kwon et al., 2005), the capacity of VoIP services was evaluated through a simulation framework in the IEEE 802.16e OFDMA system but without the development of an analytical model. The performance of wireless OFDMA systems was studied in (Niyato & Hossain, 2005a;b). None of these studies, however, considered the signaling overhead. Although other studies have evaluated how the signaling overhead affects the system performance in the wireless OFDMA system, they failed to consider the algorithm

for reducing the signaling overhead (Gross et al., 2006; So, 2008). In (Ben-Shimol et al., 2006), persistent scheduling was introduced for constant bit rate voice sessions; however, no analytical model was used and no consideration was given to the adaptive modulation and coding (AMC) scheme for data transmissions. In (Wan et al., 2007), a cross-layer packet scheduling and subchannel allocation scheme was proposed for IEEE 802.16e OFDMA systems. Each packet is prioritized in relation to its channel quality but no consideration is given to the signaling overhead. Furthermore, scheduling based on channel quality is problematic when applied to delay-sensitive VoIP services. In (Jiang et al., 2007) and (Shrivastava & Vannithamby, 2009b), the performance of persistent scheduling in wireless OFDMA systems was evaluated in terms of the VoIP capacity but no analytical model was developed. In (Shrivastava & Vannithamby, 2009a) and (McBeath et al., 2007), group scheduling was proposed as a solution to the problem of persistent scheduling. Users are clustered into multiple groups, and the resource allocation for individual users has some persistence within each group's resources. However, none of these studies developed an analytical model. In (So, 2009), the performance of persistent scheduling was mathematically analyzed but the downlink resources for data transmissions and the signaling message transmissions were assumed to be separated. In a practical system, the downlink resources are shared by the data transmissions and the signaling message transmissions.

This chapter introduces the concepts of two scheduling schemes for VoIP services, dynamic scheduling and persistent scheduling, in terms of resource allocations. Moreover, we develop an analytical model to evaluate the capacity of VoIP services according to the scheduling schemes by considering the AMC scheme in data transmission. The remainder of the chapter is organized as follows: Section 2 gives a description of the system model; Section 3 introduces the dynamic scheduling and the persistent scheduling for VoIP services; Section 4 analyzes the capacity of VoIP services in view of the throughput and the signaling overhead; Section 5 shows the numerical and simulation results; and finally, Section 6 presents conclusions.

2. System model

2.1 System description

We considers a downlink (DL) VoIP transmission from a BS to users in a time division duplex (TDD)-based mobile WiMAX system of the IEEE 802.16Rev2 standard. In an OFDMA-based WiMAX system, each resource is represented in slot units; a slot is a two-dimensional entity with a time symbol space and a subchannel space. One slot carries 48 data subcarriers (IEEE, 2009). The TDD-based mobile WiMAX system is operated on a frame basis, where each frame consists of a DL subframe and an uplink (UL) subframe (IEEE, 2009). The DL subframe consists of a preamble, a frame control header (FCH), a DL-MAP message, a UL-MAP message, and data bursts. By broadcasting a MAP message, the BS indicates the location, size, and encoding of data bursts. The duration of a frame is denoted by T_f.

2.2 Channel model

The probability density function of the instantaneous received signal-to-noise ratio (SNR), γ, at the user is denoted by $f_\gamma(\gamma)$. If N denotes the total number of MCS levels available in the downlink, there are N regions defined by the thresholds $\gamma_1 < \gamma_2 < \cdots < \gamma_{N+1}$. When the instantaneous received SNR, γ, falls in region n, that is, when $\gamma_n \leq \gamma < \gamma_{n+1}$, the MCS level n is used, where $n \in \mathcal{N} = \{1, 2, \cdots, N\}$. When $\gamma < \gamma_1$, no data is assumed to be sent. The

probability that the SNR γ falls in the nth region is given by (Alouini & Goldsmith, 2000)

$$
\begin{aligned}
P_\gamma(n) &= \int_{\gamma_n}^{\gamma_{n+1}} f_\gamma(\gamma)d\gamma \\
&= \frac{\Gamma(m, m\gamma_n/\overline{\gamma}) - \Gamma(m, m\gamma_{n+1}/\overline{\gamma})}{\Gamma(m)},
\end{aligned} \tag{1}
$$

where $\Gamma(m)$ is the gamma function which equals $\Gamma(m) = \int_0^\infty t^{m-1}\exp(-t)dt$, $\Gamma(m, x)$ is the complementary incomplete gamma function which equals $\Gamma(m, x) = \int_x^\infty t^{m-1}\exp(-t)dt$, m is the Nakagami fading parameter, and $\overline{\gamma}$ is the average SNR.

Wireless channel is described by a finite state Markov chain taking the discrete adaptive modulation and coding into consideration, as shown in Fig. 1. Assuming slow fading conditions, the state transition probability of the MCS level during the frame duration T_f is given by (Liu et al., 2005; Razavilar et al., 2002)

$$
P_t(i,j) = \begin{cases}
(N_{i+1}T_f)/P_\gamma(i), & \text{if } j = i+1, j \in \mathcal{N} \\
(N_iT_f)/P_\gamma(i), & \text{if } j = i-1, j \in \mathcal{N} \\
1 - P_t(i, i+1) - P_t(i, i-1), & \text{if } j = i, j \in \mathcal{N} \\
0, & \text{otherwise,}
\end{cases} \tag{2}
$$

where i is the MCS level in the current frame and j is the MCS level in the next frame. The level crossing rate, N_i, is expressed as follows (Liu et al., 2005):

$$
N_i = \sqrt{2\pi \frac{m\gamma_i}{\overline{\gamma}}} \frac{f_d}{\Gamma(m)} \left(\frac{m\gamma_i}{\overline{\gamma}}\right)^{m-1} \exp\left(-\frac{m\gamma_i}{\overline{\gamma}}\right), \tag{3}
$$

where f_d is the maximum Doppler shift given in hertz.

2.3 VoIP traffic model

The G.729 codec generates a 20 byte encoded voice frame every $T_v = 20$ milliseconds (Bi et al., 2006). Hence, the average size of voice data per a medium access control (MAC) packet can be expressed as follows:

$$
L_v = \frac{T_s}{20 \text{ milliseconds}} \times 20 \text{ bytes}, \tag{4}
$$

where T_s is the scheduling period. For example, if the BS schedules voice frames every $T_s = 40$ milliseconds, the value of L_v becomes 40 bytes. The constant overhead at the MAC layer is 13 bytes including a 6 byte generic MAC header, a 4 byte cyclic redundancy check (CRC), and a 3 byte IP header, because the IP header can fit into 3 bytes as a result of robust header compression. The packet structures are depicted in Fig. 2. The VoIP packets are assumed to be transmitted in accordance with a simplified first-in-first-out scheduling model. Moreover, the VoIP packet uses an AMC scheme at the physical layer.
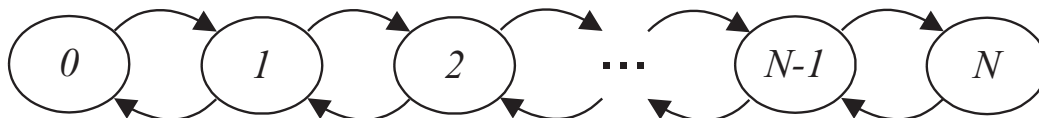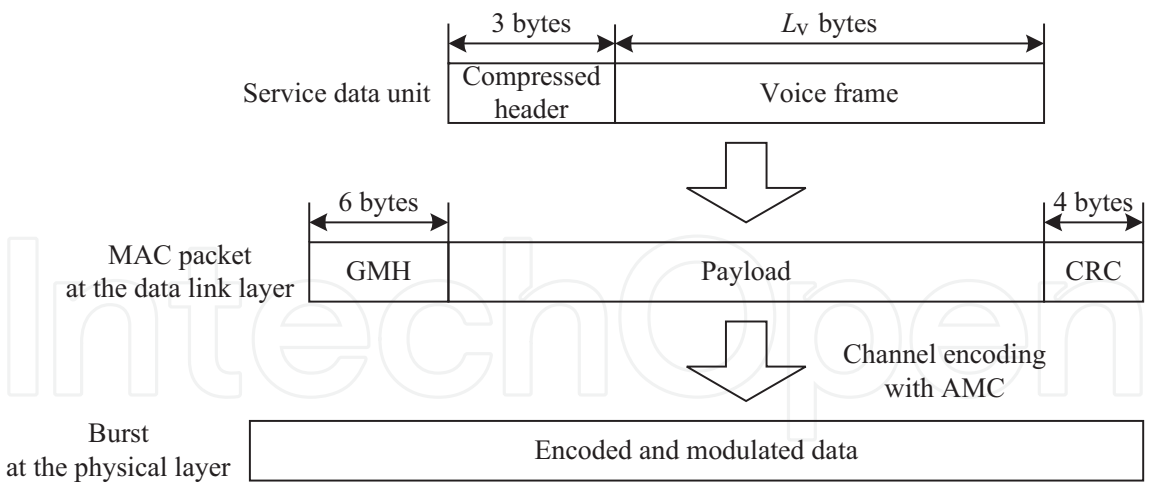

Fig. 1. Finite states Markov channel model

Fig. 2. Packet structure

The VoIP traffic has been modeled as an exponentially distributed on-off model with a mean on-time of $\alpha^{-1} = 1$ second and a mean off-time of $\beta^{-1} = 1.5$ second (Ozer et al., 2000). We use the two-state Markov-modulated Poisson process (MMPP) to model the aggregate VoIP traffic requested from N_v users (Heffes & Lucantoni, 1986; Shah-Heydari & Le-Ngoc, 1998). The two-state MMPP is represented by the transition rate matrix, $\mathbf{R}$, and the Poisson arrival rate matrix, $\mathbf{\Lambda}$, as follows:

$$\mathbf{R} = \left[\begin{array}{cc} -r_1 & r_1 \\ r_2 & -r_2 \end{array} \right], \qquad \mathbf{\Lambda} = \left[\begin{array}{cc} \lambda_1 & 0 \\ 0 & \lambda_2 \end{array} \right]. \tag{5}$$

We determine the four parameters, λ_1, λ_2, r_1, and r_2, by using the index of dispersion for counts (IDC) matching technique as follows (Shah-Heydari & Le-Ngoc, 1998; Baiocchi et al., 1991; Huang et al., 1996):

$$\lambda_1 = A \frac{\sum_{j=0}^{M_v} j\, \pi_j}{\sum_{i=0}^{M_v} \pi_i}, \qquad \lambda_2 = A \frac{\sum_{j=M_v+1}^{N_v} j\, \pi_j}{\sum_{i=M_v+1}^{N_v} \pi_i}, \tag{6}$$

where $\pi_j = \binom{N_v}{j} p^j (1-p)^{N_v-j}$, $p = \beta/(\alpha+\beta)$, $M_v = \lfloor N_v \cdot p \rfloor$, and A, which is the emission rate in the on state, equals $1/T_v$. The transition rates are as follows:

$$r_1 = \frac{2(\lambda_2 - \lambda_{avg})(\lambda_{avg} - \lambda_1)^2}{(\lambda_2 - \lambda_1)\lambda_{avg}(\mathrm{IDC}(\infty) - 1)} \tag{7}$$

$$r_2 = \frac{2(\lambda_2 - \lambda_{avg})^2(\lambda_{avg} - \lambda_1)}{(\lambda_2 - \lambda_1)\lambda_{avg}(\mathrm{IDC}(\infty) - 1)}, \tag{8}$$

where $\lambda_{avg} = N_v \cdot A \cdot p$ and $\mathrm{IDC}(\infty)$ is taken from (Heffes & Lucantoni, 1986).

3. Scheduling schemes

3.1 Dynamic scheduling

In the conventional mobile WiMAX system, the BS broadcasts a DL-MAP message for every frame to inform the allocations of radio resources in the downlink. A DL-MAP message

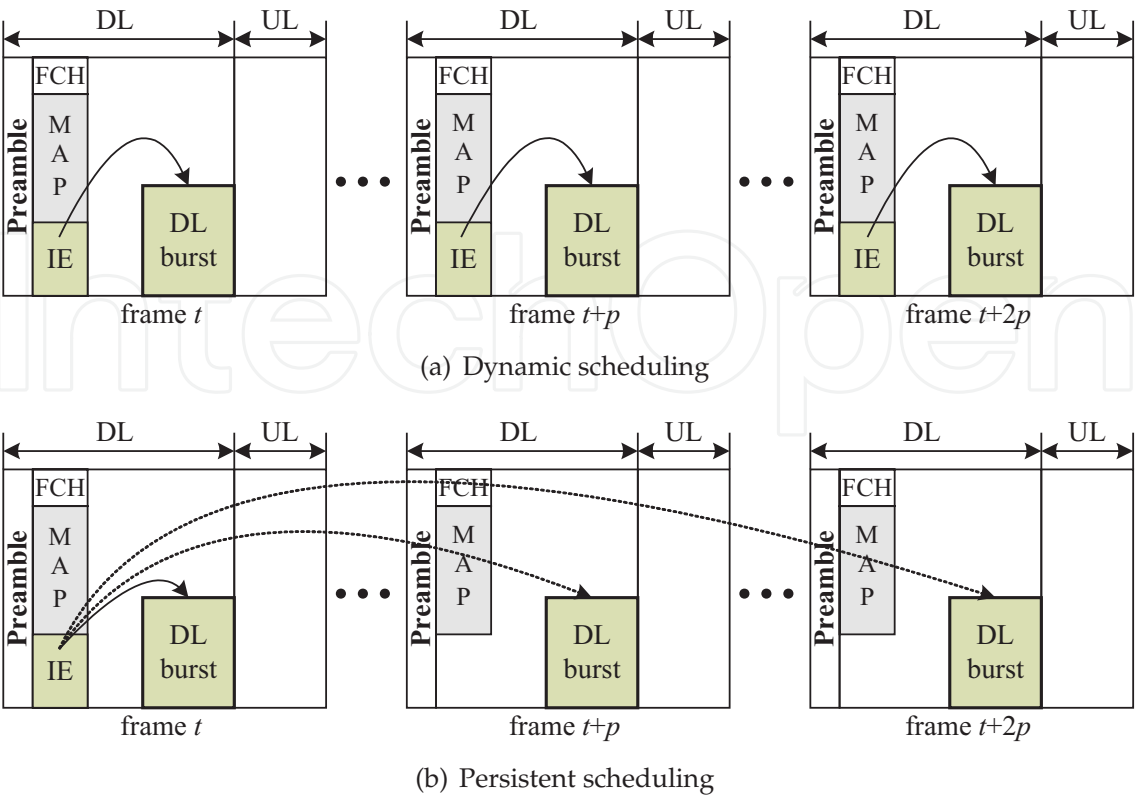(a) Dynamic scheduling

(b) Persistent scheduling

Fig. 3. Dynamic scheduling and persistent scheduling

contains DL-MAP information elements (IEs) that indicate the location, size, and encoding of data bursts directed to the users. The flow between the BS and a user is identified by a connection identifier (CID). Packets directed to different users are integrated into a single burst if the MCS levels of the packets are identical. Let all VoIP packets scheduled for the downlink frame t be denoted by $\mathbf{X}^{(t)} = (x_1^{(t)}, x_2^{(t)}, \cdots, x_N^{(t)})$, where $x_n^{(t)}$ is the number of packets modulated with the nth MCS level and N is the total number of MCS levels available in the downlink. The superscript (t) can be omitted for the steady state analysis. In dynamic scheduling, a DL-MAP IE uses a constant 44 bits to indicate the location, size, and encoding of a data burst; it also uses a 16 bit CID field. Accordingly, in dynamic scheduling, the size of the DL-MAP IEs can be expressed as follows (IEEE, 2009):

$$h_{\text{IE}}^{(ds)}(\mathbf{X}) = \sum_{n=1}^{N} (44 + 16x_n) \cdot J(x_n) \text{ [bits]}, \tag{9}$$

where $J(x_n)$ is an index function expressed as follows: if $x_n > 0$, $J(x_n) = 1$; otherwise $J(x_n) = 0$.

3.2 Persistent scheduling

For VoIP services, the packet arrival rate is somewhat predictable. Hence, the BS can reduce the signaling overhead by transmitting an initial assignment message, which is valid in a periodic sequence of future frames. This type of scheduling is referred to as *persistent scheduling* (IEEE, 2009; 2010).

Figure 3 illustrates a high-level concept of dynamic scheduling and persistent scheduling for when a BS transmits a burst for every p frame in a downlink. In dynamic scheduling, as shown

in Fig. 3(a), the BS broadcasts a DL-MAP IE in the MAP message for frame t, frame $t + p$, frame $t + 2p$, and so on, where p is the period of the allocation. The DL-MAP IEs indicate the location, size, and encoding of the DL burst in each frame. Because the BS allocates resources by using the DL-MAP IEs for every frame, the BS can change the modulation and coding schemes from frame to frame. However, in persistent scheduling, the BS allocates a persistent resource to a user when it first schedules the user in frame t; and the allocated resource is valid in frame $t + p$, frame $t + 2p$, and so on. Hence, as shown in Fig. 3(b), the BS broadcasts a DL-MAP IE in the MAP message only for frame t and does not broadcast the DL-MAP IEs for frame $t + p$, frame $t + 2p$, and so on. Accordingly, the signaling overhead decreases and the effective downlink resource increases. However, persistent scheduling may result in some inefficiency because the BS cannot change both the MCS level and the locations of persistently allocated resources on a frame-by-frame basis.

The main problems of persistent scheduling are the resource hole and the MCS mismatch. The term *resource hole* is used to describe sets of successive slots that are not allocated between persistently allocated resources. A resource hole is generated whenever an already allocated burst is deallocated because the resource hole can be completely filled by the new user with the exact same resource requirements. The term *MCS mismatch* is used to describe the difference between the optimized MCS level at the current frame and the latest MCS level indicated by the BS through the persistent scheduling. The MCS mismatch is caused by variation of the radio channel during the session. The MCS mismatch causes a link adaptation error or an additional overhead due to signaling the change to the user (Shrivastava & Vannithamby, 2009a). The resource hole and the MCS mismatch both degrade the efficiency of the resource utilization.

We propose a new format of a DL-MAP IE for persistent scheduling. The format is shown in Table 1. The proposed persistent DL-MAP IE follows the format of the standard DL-MAP extended-2 IE (IEEE, 2009). The format of the proposed DL-MAP IE has two parts. The first part indicates the location, size, and encoding of a burst that the BS transmits to a user every p frames. The allocation of the bandwidth starts from the *slot offset* of the last zone, and the allocated bandwidth is represented by the *allocation size*. The encoding is implicitly determined by the mapping relation between the MCS level and the size of the burst, as shown in Table 2. The second part is the adjustment part. The BS performs an adjustment procedure to eliminate the problems of persistent scheduling by configuring the two fields shown in Table 1: the *adjustment offset* and the *adjustment size*. The user, which uses a persistent allocation, updates its location and size in relation to these two fields. If the value of the adjustment offset is not equal to its slot offset, the user increases or decreases its slot offset by the value of the adjustment offset; otherwise the user does not update its slot offset. If the value of the adjustment offset is equal to the slot offset of an user, the user increases or decreases its bandwidth by the value of the adjustment size and changes its MCS level in accordance with the mapping relation between the MCS level and the burst size. Hence, through these adjustments, the proposed DL-MAP IE prevents the resource hole and the MCS mismatch from degrading the performance.

Although the IEEE 802.16Rev2 and the IEEE 802.16m standard include a format for a persistent DL-MAP IE (IEEE, 2009; 2010), the proposed persistent DL-MAP IE has the advantage of being able to reduce the size of the standard persistent DL-MAP IE. The size reduction is as follows: first, the proposed DL-MAP IE eliminates the CID field whenever the BS adjusts the persistently allocated resources because the CID information can be implicitly determined by the location of the allocated resources. Second, as shown in Table 2, the

Syntax	Bits	Notes
DIUC	4	if (DIUC==14) Extended-2
Extended-2 DIUC	4	
Length	8	Length in bytes of the following data
Allocation Flag	1	Indicate a resource allocation
if (Allocation Flag == 1) {		
N_Alloc	4	Number of allocations
for (i=0; i<N_Alloc; i++) {		
CID	16	Connection indentifier
Slot Offset	8	Offset from the last of zone
Allocation Size	8	Bandwidth in units of slots
Allocation Period	4	Allocation period, p
}		
}		
Adjustment Flag	1	Indicate an adjustment
if (Adjustment Flag == 1) {		
N_Adj	6	Number of adjustments
for (i=0; i<N_Adj; i++) {		
Adjustment Offset	8	Offset from the last of zone
Adjustment Size	8	Increase/decrease of bandwidth in units of slots (signed value)
}		
}		

Table 1. Format of the proposed persistent DL-MAP IE

proposed DL-MAP IE eliminates the encoding fields because the MCS level can be implicitly determined by the mapping relation between the MCS level and the allocated size.

The size of the proposed persistent DL-MAP IE depends on the number, u, of new allocations and the number, v, of existing allocations that changed in size during the p frames. The signaling overhead due to new allocations can be neglected because the talk spurt time is relatively long compared to the frame time, usually in hundreds of milliseconds in contrast to several milliseconds. The proposed persistent DL-MAP IE uses constant 18 bits to indicate the extended-2 IE and flags; it also uses 6 bits to indicate the number of adjustment bursts. In addition, two adjustment fields use 16 bits to adjust the location, size, and encoding of a persistently allocated burst. Accordingly, in persistent scheduling, the size of the DL-MAP IEs can be approximated as follows:

$$h_{\text{IE}}^{(ps)}(v) \approx \left\{ 18 + (6 + 16v) \right\} \cdot J(v) \text{ [bits]}, \qquad (10)$$

MCS level, n	Modulation and Coding	bits/ symbol	Burst size (slots), l_n		Threshold, dB
			when $T_s = 20$ ms	when $T_s = 40$ ms	
1	QPSK 1/12	0.17	36	54	-5.6
2	QPSK 1/8	0.25	24	36	-3.8
3	QPSK 1/4	0.5	12	18	-1.4
4	QPSK 1/2	1.0	6	9	2.1
5	QPSK 3/4	1.5	4	6	6.6
6	16-QAM 1/2	2.0	3	5	7.2
7	16-QAM 3/4	3.0	2	3	12.5

Table 2. Modulation and coding schemes for VoIP traffic

where $J(v)$ is an index function expressed as follows: if $v > 0$, $J(v) = 1$; otherwise $J(v) = 0$.

4. Performance analysis

4.1 MCS variation in persistent scheduling

In the persistent scheduling, the last allocation is used to transmit a VoIP packet without any notification of a DL-MAP IE if the MCS level is unchanged. However, the MCS level may vary in every frame in accordance with the time-varying channel conditions. The probability of staying at the same MCS level, n, during p frames is

$$
\begin{aligned}
\Omega_p(n) &= \sum_{\forall m_i} \Big\{ P_t(n, m_2) P_t(m_2, m_3) \cdots P_t(m_p, n) \Big\} \\
&= \sum_{m_i \in \mathcal{Z}} \prod_{i=1}^{p} P_t(m_i, m_{i+1}),
\end{aligned}
\tag{11}
$$

where $\mathcal{Z} = \{\forall (m_i, m_{i+1}) \mid m_i \leq m_{i+1} \leq m_i + 1, m_1 = m_{p+1} = n, m_i \in \mathcal{N}, m_{i+1} \in \mathcal{N}\}$ and the state transition probability of the MCS level during the frame duration, $P_t(m_i, m_{i+1})$, is obtained from (2). Hence, the average probability of staying at the same MCS level during p frames is

$$
\xi = \sum_{n=1}^{N} \Omega_p(n) P_\gamma(n),
\tag{12}
$$

where $P_\gamma(n)$ is obtained from (1). When the MCS levels of all the users are distributed with $\mathbf{X} = (x_1, x_2, \cdots, x_N)$, the probability of the MCS levels of v users being changed during the p frames is given by

$$
P_c(v|\mathbf{X}) = \sum_{\forall \mathcal{Y}} \prod_{n=1}^{N} \binom{x_n}{y_n} (1 - \Omega_p(n))^{y_n} (\Omega_p(n))^{x_n - y_n},
\tag{13}
$$

where $\mathcal{Y} = \{(y_1, y_2, \cdots, y_n) \mid \sum_{n=1}^{N} y_n = v, y_n \leq x_n\}$.

4.2 Scheduling feasibility condition

For simplicity, the UL-MAP message and the UL bursts are not considered. In the MAP message, a BS may transmit a 12 bit CID-switch IE to toggle the inclusion of the CID parameter. With the subsequent inclusion of a 88 bit constant overhead and a 32 bit CRC, the size of the compressed MAP message in units of bits can be expressed as follows (IEEE, 2009):

$$
h_{\text{MAP}}(\cdot) = \left\lceil \frac{88 + 12 + h_{\text{IE}}(\cdot) + 32}{8} \right\rceil \cdot 8,
\tag{14}
$$

where $h_{\text{IE}}(\cdot)$, which is the size of the DL-MAP IEs, is obtained from (9) or (10) according to the scheduling scheme. The MAP message is generally modulated with a QPSK rate of $1/2$ and broadcast after six repetitions; and one slot carries 48 data subcarriers (IEEE, 2009; So, 2008). Accordingly, when the MCS levels of all the users are distributed in the manner of $\mathbf{X} = (x_1, x_2, \cdots, x_N)$ in dynamic scheduling, the size of the MAP message in units of slots is given by

$$
H_{\text{MAP}}^{(ds)}(\mathbf{X}) = \lceil h_{\text{MAP}}(\mathbf{X})/48 \rceil \cdot 6.
\tag{15}
$$

Similarly, in persistent scheduling, the average size of the MAP message in units of slots is given by

$$H_{\text{MAP}}^{(ps)}(\mathbf{X}) = \sum_{v=0}^{\sum_{n=1}^{N} x_n} \left(\lceil h_{\text{MAP}}(v)/48 \rceil \cdot 6 \right) P_c(v|\mathbf{X}). \tag{16}$$

The DL scheduling is feasible if the resources occupied by the FCH, the MAP message, and the data bursts are less than or equal to the total available resources in units of slots, N_{tot}. Then, when the MCS levels of all the scheduled users are distributed in manner of $\mathbf{X} = (x_1, x_2, \cdots, x_N)$, the feasibility condition is

$$
\begin{aligned}
\Gamma(\mathbf{X}) &= H_{\text{FCH}} + H_{\text{MAP}}(\mathbf{X}) + \sum_{n=1}^{N} (x_n \cdot l_n) \\
&\leq N_{\text{tot}},
\end{aligned}
\tag{17}
$$

where H_{FCH}, which denotes the number of slots used to transmit the FCH, is 4 (IEEE, 2009); x_n denotes the number of packets modulated by the nth MCS level; and l_n denotes the size of the data burst, which is modulated with the nth MCS level, after the encoding and repetition in units of slots. The value of l_n is shown in Table 2.

4.3 Queuing analysis

The performance of VoIP services is analyzed with a discrete time Markov chain model. A discrete-time MMPP can be equivalent to an MMPP in continuous time (Niyato & Hossain, 2005a). Arrival and service process of the queue is depicted in Fig. 4. The queueing analysis is based on our earlier work (So, 2008).

4.3.1 Arrival process

We define the diagonal probability matrix, $\mathbf{D}_k$. Each diagonal element of $\mathbf{D}_k$ is the probability of k packets transmitting from users during the frame duration, T_f, and this probability is given by $(\lambda_i T_f)^k e^{-\lambda_i T_f}/k!$ for $i = 1, 2$ where λ_i is obtained from (6). Furthermore, the average packet arrival rate at the queue during the frame duration is

$$\rho = \mathbf{s} \left(\sum_{k=0}^{N_v \cdot A_{\max}} k \, \mathbf{D}_k \right) \mathbf{1}, \tag{18}$$

where $A_{\max}$ is the maximum number of packets that can be transmitted during T_f per user; $\mathbf{1}$ is a column matrix of ones; and $\mathbf{s} = [s_1, s_2]$ is obtained by solving $\mathbf{s}\mathbf{U} = \mathbf{s}$ and $s_1 + s_2 = 1$, where the matrix $\mathbf{U}$ is given by (Heffes & Lucantoni, 1986)

$$\mathbf{U} = (\mathbf{\Lambda} - \mathbf{R})^{-1} \mathbf{\Lambda}, \tag{19}$$
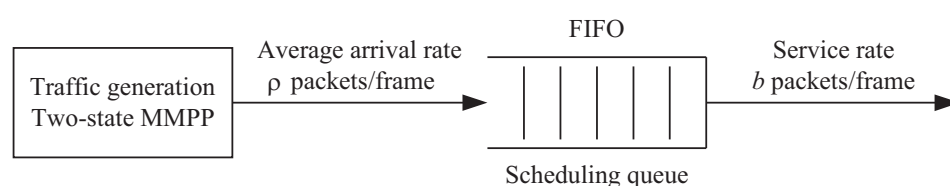

Fig. 4. Arrival and service process of the queue

where $\mathbf{\Lambda}$ and $\mathbf{R}$ are obtained from (5). The transition probability matrix $\mathbf{U}$ keeps track of the phase during an idle period. Each element U_{ij} of the matrix $\mathbf{U}$ is the transition probability that the first arrival to a busy period arrives with the MMPP in phase j, given that the last departure from the previous busy period departs with the MMPP in phase i (Heffes & Lucantoni, 1986).

4.3.2 Service process

The BS schedules VoIP packets from the queue in accordance with the FIFO policy. The number of the scheduled VoIP packets depends on the channel condition of each VoIP packet. Let b denote the number of VoIP packets scheduled at frame time t, i.e., $b = x_1 + x_2 + \cdots + x_N$, where x_n is the number of VoIP packets modulated with the nth MCS level. At frame time t, if the (17) is satisfied when the BS services the b packets and the (17) is not satisfied when the BS services the $(b+1)$ packets, then the BS will schedule b packets in the frame. Let the parameters $\mathbf{X}_b$ and $\mathbf{X}'_{b+1}$ be denoted as follows: $\mathbf{X}_b = \{\forall(x_1, x_2, \cdots, x_N) \mid \sum_{n=1}^{N} x_n = b, x_n \geq 0\}$; and $\mathbf{X}'_{b+1} = \{\forall(x'_1, x'_2, \cdots, x'_N) \mid \sum_{n=1}^{N} x'_n = b+1, x_n \leq x'_n \leq x_n + 1\}$. The cases where the BS schedules b packets are then represented by

$$\psi_b = \{\forall \mathbf{X}_b \mid \Gamma(\mathbf{X}_b) \leq N_{\text{tot}} \text{ and } \Gamma(\mathbf{X}'_{b+1}) > N_{\text{tot}}\}. \tag{20}$$

Let the index function be defined as follows:

$$I_n(\mathbf{X}_b) = \begin{cases} 1, & \text{if } \mathbf{X}_b \notin \psi_b \text{ when } x_n \text{ increases} \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

Two conditions should be satisfied for the BS to schedule b VoIP packets from the queue: the first condition is that the MCS-level distribution of b packets satisfies (17) and the second condition is that the MCS-level distribution of $(b+1)$ packets does not satisfy (17) when the BS schedules the $(b+1)$th packet. Thus, the probability of the BS scheduling b VoIP packets from the queue is (So, 2008)

$$\begin{aligned} P_s(b) &= \Pr\{\text{the number of scheduled packets} = b\} \\ &= \Pr\{\mathbf{X}_b \in \psi_b \text{ and } \mathbf{X}_{b+1} \notin \psi_{b+1}\} \\ &= \sum_{\forall \mathbf{X}_b \in \psi_b} \left[\left(b! \prod_{n=1}^{N} \frac{P_\gamma(n)^{x_n}}{x_n!} \right) \left(1 - \sum_{n=1}^{N} \left(P_\gamma(n) I_n(\mathbf{X}_b) \right) \right) \right], \end{aligned} \tag{22}$$

where $P_\gamma(n)$ is obtained from (1). The probability $P_s(b)$ is the sum of the products of two equations. The left side of the equation is the probability that b packets are distributed with a specific MCS-level distribution, $\mathbf{X}_b$. The right side of the equation is the probability that the $(b+1)$th packet is not a specific MCS level.

4.3.3 State transition probability

The state is defined as the number of packets in the queue and is expressed as follows: $\boldsymbol{\pi} = [\pi_0 \, \pi_1 \cdots \pi_{2K+1}]$. Then, the state transition matrix $\mathbf{P}$ of the queue can be expressed as follows:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{0,0} & \mathbf{P}_{0,1} & \cdots & \mathbf{P}_{0,K} \\ \mathbf{P}_{1,0} & \mathbf{P}_{1,1} & \cdots & \mathbf{P}_{1,K} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{P}_{K,0} & \mathbf{P}_{K,1} & \cdots & \mathbf{P}_{K,K} \end{bmatrix} \tag{23}$$

where K is the maximum size of the queue. The element $\mathbf{p}_{i,j}$ represents the transition probability that the number of packets in the queue will be j at the next frame when the number of packets is i at the current frame. If the number of packets in the queue of the current frame is i and the BS schedules b packets during the frame duration, a new batch of $\{j - \max(i - b, 0)\}$ packets should arrive so that the number of packets in the queue of the next frame is j. Hence, each element of the matrix $\mathbf{P}$ is obtained as follows:

$$\mathbf{p}_{i,j} = \sum_{b=b_{\min}}^{b_{\max}} \mathbf{U}\mathbf{D}_{j-\max(i-b,0)}P_s(b). \tag{24}$$

The matrix $\boldsymbol{\pi}$ is obtained from the equations $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}\mathbf{1} = 1$. The probability of k packets being in the queue is $\pi(k) = \pi_{2k} + \pi_{2k+1}$.

4.4 Throughput analysis

The average number of VoIP packets transmitted to users during the frame duration is

$$\overline{b} = \sum_{b=b_{\min}}^{b_{\max}} \sum_{k=0}^{K} \min(k,b)\pi(k)P_s(b), \tag{25}$$

where K is the maximum queue size, $b_{\min}$ is the minimum number of scheduled packets, and $b_{\max}$ is the maximum number of scheduled packets in the downlink. Accordingly, the average throughput, which is defined as the average amount of voice data successfully transmitted per second, is

$$S = \overline{b} \cdot L_v / T_f, \tag{26}$$

where L_v, which is the size of voice data in a VoIP packet, is obtained from (4).

4.5 Signaling overhead

Let the signaling overhead be defined as the size of the DL-MAP IEs. In dynamic scheduling, the average signaling overhead can then be expressed as follows:

$$\begin{aligned}
H_{\text{sig}}^{(ds)} &= \sum_{b=b_{\min}}^{b_{\max}} \sum_{k=0}^{b-1} \sum_{\forall \mathbf{X}_k} \left[\left(k! \prod_{n=1}^{N} \frac{P_\gamma(n)^{x_n}}{x_n!} \right) P_s(b)\pi(k)h_{\text{IE}}^{(ds)}(\mathbf{X}_k) \right] \\
&+ \sum_{b=b_{\min}}^{b_{\max}} \sum_{k=b}^{K} \sum_{\forall \mathbf{X}_b \in \psi_b} \left[\left(b! \prod_{n=1}^{N} \frac{P_\gamma(n)^{x_n}}{x_n!} \right) \right. \\
&\qquad \left. \times \left(1 - \sum_{n=1}^{N} P_\gamma(n)I_n(\mathbf{X}_b) \right) \pi(k)h_{\text{IE}}^{(ds)}(\mathbf{X}_b) \right].
\end{aligned} \tag{27}$$

Similarly, when persistent scheduling is applied, the average signaling overhead is given by

$$\begin{aligned}
H_{\text{sig}}^{(ps)} &= \sum_{b=b_{\min}}^{b_{\max}} \sum_{k=0}^{b-1} \sum_{v=0}^{k} \sum_{\forall \mathbf{X}_k} \left[\left(k! \prod_{n=1}^{N} \frac{P_\gamma(n)^{x_n}}{x_n!} \right) P_s(b)\pi(k)h_{\text{IE}}^{(ps)}(v)P_c(v|\mathbf{X}_k) \right] \\
&+ \sum_{b=b_{\min}}^{b_{\max}} \sum_{k=b}^{K} \sum_{v=0}^{b} \sum_{\forall \mathbf{X}_b \in \psi_b} \left[\left(b! \prod_{n=1}^{N} \frac{P_\gamma(n)^{x_n}}{x_n!} \right) \right. \\
&\qquad \left. \times \left(1 - \sum_{n=1}^{N} P_\gamma(n)I_n(\mathbf{X}_b) \right) \pi(k)h_{\text{IE}}^{(ps)}(v)P_c(v|\mathbf{X}_b) \right],
\end{aligned} \tag{28}$$

where $P_c(v|\mathbf{X})$ is obtained from (13).

5. Numerical and simulation results

The downlink performance of VoIP services is evaluated in a mobile WiMAX system with a Rayleigh channel environment of $f_\gamma(\gamma) = 1/\overline{\gamma}\exp(-\gamma/\overline{\gamma})$, where $\overline{\gamma}$ is the average received SNR. On the assumption of a partial usage of subchannels (PUSC), a diversity subcarrier permutation is used to build a subchannel. For a downlink PUSC, one slot consists of one subchannel and two OFDMA symbols and one slot carries 48 data subcarriers (IEEE, 2009). The total number of MCS levels available in the downlink is assumed to be $N = 7$ with the thresholds as shown in Table 2. The thresholds were obtained by computer simulation under a practical environment with the channel ITU-R recommendation M.1225 (Leiba et al., 2006). For a mobile WiMAX system with a bandwidth of 8.75 MHz, the simulation uses a frame structure of $T_f = 5$ milliseconds and $N_{\text{tot}} = 390$ slots (IEEE, 2009; So, 2008). Figure 5 and Figure 6 assume that the BS schedules the voice frames every 20 millisecond, i.e., $T_s = 4$ frames. Accordingly, in persistent scheduling, the persistent allocation period is $p = 4$ frames. Figure 5 shows the average throughput as the number of active voice users increases. The average throughput linearly increases when the number of active voice users is less than a certain number of active voice users. However, the throughput approaches an asymptotic limit after the offered load overwhelms the system capacity. The asymptotic limit of the average throughput is higher in the persistent scheduling than in the dynamic scheduling because persistent scheduling increases the effective downlink resources by reducing the signaling overhead. For example, for $\overline{\gamma} = 9$ dB, the asymptotic limit of the average throughput is about 1.41 Mbps in persistent scheduling and 1.14 Mbps in dynamic scheduling.
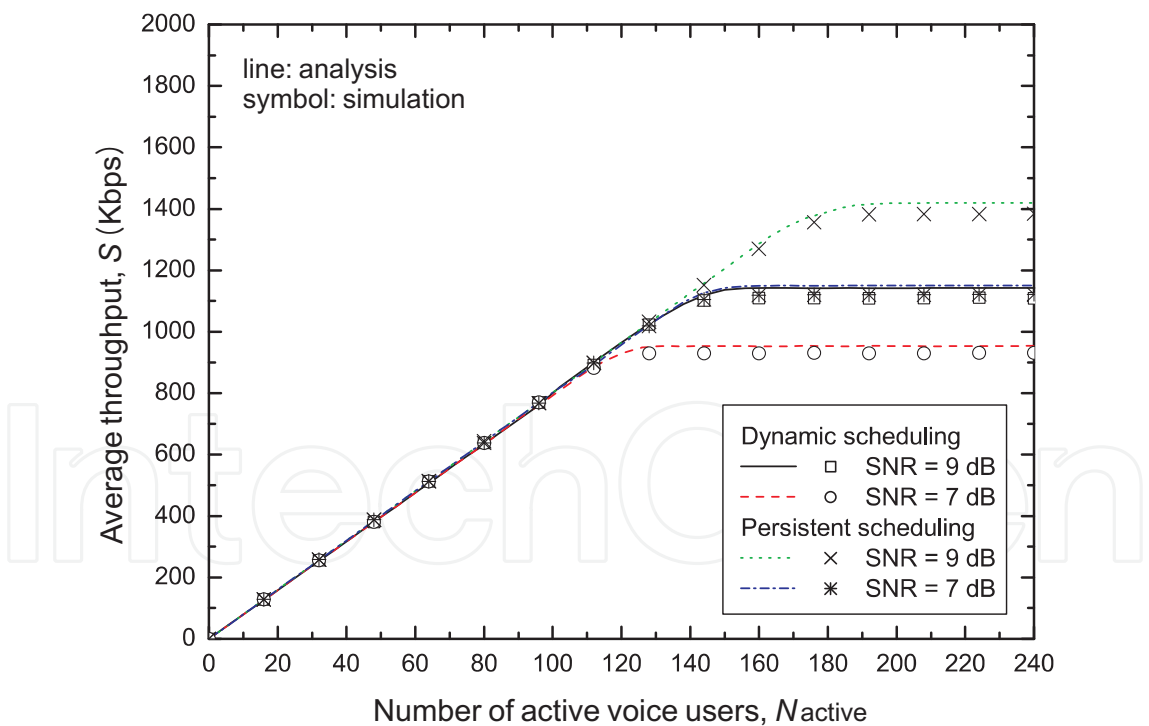

Fig. 5. Average throughput versus the number of active voice users when $T_s = 40$ milliseconds

Figure 6 shows the average signaling overhead for both dynamic scheduling and persistent scheduling. In dynamic scheduling, the signaling overhead linearly increases as the number of scheduled VoIP packets increases. Under high loading conditions, the signaling overhead of dynamic scheduling is about 772 bits when $\overline{\gamma} = 9$ dB and about 685 bits when $\overline{\gamma} = 7$ dB.

However, in persistent scheduling, the signaling overhead is not dependent on the number of scheduled packets but on the number of packets whose MCS levels change during the allocation period. In the simulation environments, the average probability of staying at the same MCS level when $p = 4$ frames is about $\xi = 0.64$, regardless of the value of $\overline{\gamma}$. The value of ξ directly decreases the signaling overhead. Under high loading conditions, the signaling overhead of persistent scheduling is approximately 235 bits, regardless of the value of $\overline{\gamma}$.

Fig. 6. Average signaling overhead versus the number of active voice users when $T_s = 40$ milliseconds

Figure 7 and Figure 8 assume that the BS schedules the voice frames every 20 milliseconds or 40 milliseconds; that is, $T_s = 4$ or 8 frames. Accordingly, in persistent scheduling, the persistent allocation period is $p = 4$ or 8 frames. Figure 7 shows the average throughput in relation to the scheduling period for when $\overline{\gamma} = 9$ dB. As the scheduling period increases, the average throughput increases because the MAC overhead decreases by about 38%. The signaling overhead also decreases as the scheduling period increases because the number of scheduled bursts decreases when the scheduling period increases. However, the increment in the scheduling period increases the scheduling delay. Under high loading conditions, the average throughput of dynamic scheduling is about 1.14 Mbps when $T_s = 4$ frames and about 1.61 Mbps when $T_s = 8$ frames. That is, the average throughput of the dynamic scheduling increases by about 41.2% when the scheduling period increases from 20 milliseconds to 40 milliseconds. Under high loading conditions, the average throughput of persistent scheduling is about 1.41 Mbps when $p = 4$ frames and about 1.88 Mbps when $p = 8$ frames. That is, the average throughput of the persistent scheduling increases by about 33.3%. In the simulation environments, the average probability of staying at the same MCS level is about $\xi = 0.64$ when $p = 4$ frames and $\xi = 0.54$ when $p = 8$ frames. The decrement of the value of ξ directly increases the signaling overhead. Hence, when the scheduling period increases from 20 milliseconds to 40 milliseconds, the throughput increase is smaller in persistent scheduling less than in dynamic scheduling.
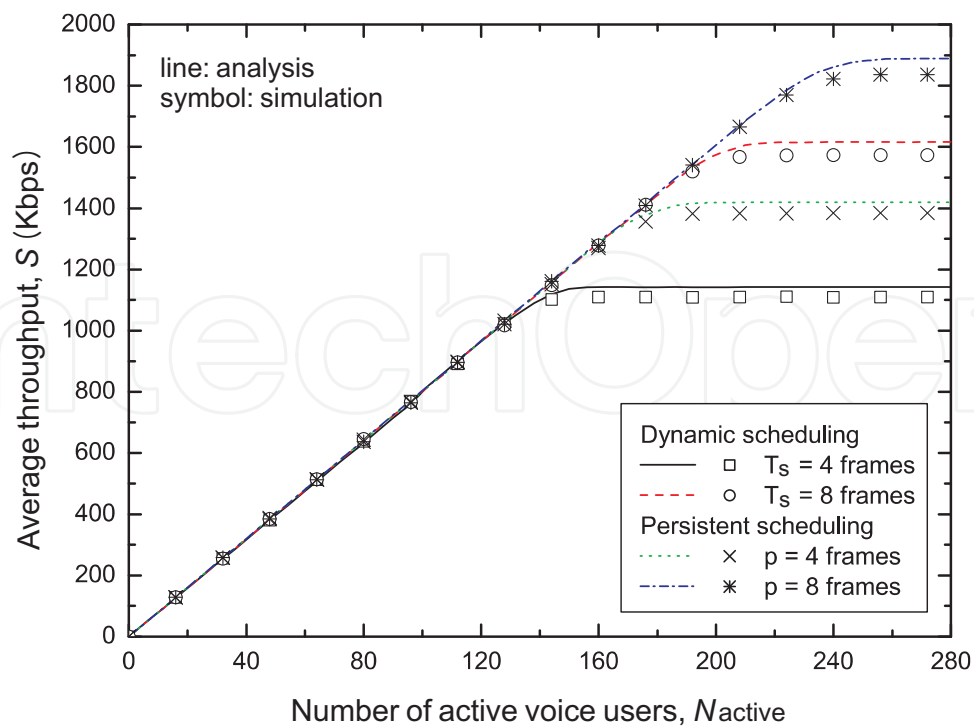
Fig. 7. Average throughput in relation to the allocation period for when $\overline{\gamma} = 9$ dB

Figure 8 shows the average signaling overhead in relation to the scheduling period for when $\overline{\gamma} = 9$ dB. Under high loading conditions, the average signaling overhead of dynamic scheduling decreases by about 23.1% as the scheduling period increases because the number of scheduled bursts decreases with the increase of the scheduling period. Similarly, under high loading conditions, the average signaling overhead of persistent scheduling decreases by about 10.5% as the scheduling period increases although the average probability of staying at the same MCS level increases with the increase of the persistent allocation period.

6. Conclusion

The chapter introduced two scheduling schemes, dynamic scheduling and persistent scheduling, for VoIP services in wireless OFDMA systems. Additionally, we developed analytical and simulation models to evaluate the performance of VoIP services in terms of the average throughput and the signaling overhead according to the scheduling schemes. The integrated voice traffic from individual users is used to construct a queueing model at the data link layer, and each VoIP packet is adaptively modulated and coded according to the wireless channel conditions at the physical layer. In VoIP services, the signaling overhead causes serious spectral inefficiency of wireless OFDMA systems. In dynamic scheduling, the signaling overhead depends on the number of scheduled VoIP packets; it also depends on the MCS-level distributions of the data bursts. However, in persistent scheduling, the signaling overhead is not dependent on the number of scheduled packets but on the number of packets whose channel states change during the allocation period. Under high loading conditions, when the average SNR is 9 dB, the average throughput is roughly 23.6% higher in persistent scheduling than in dynamic scheduling because persistent scheduling significantly reduces the signaling overhead by eliminating the notification of the resource allocation. When the allocation period is 4 frames, the signaling overhead is roughly 68.7% less in persistent scheduling than in dynamic scheduling. Hence, a reduction in the signaling overhead is
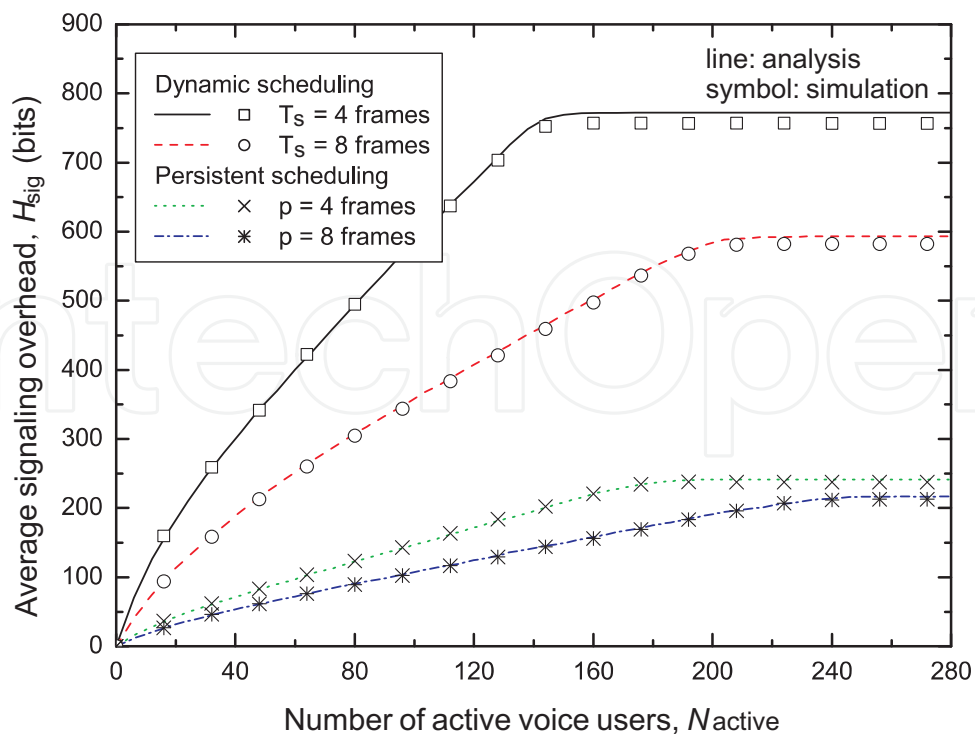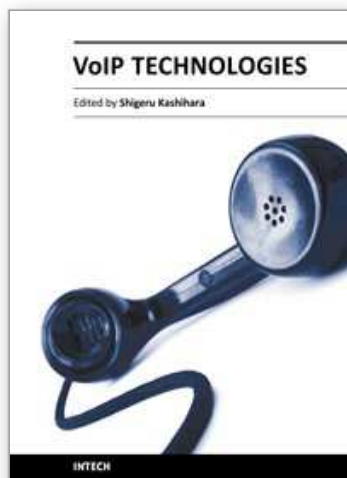
Fig. 8. Average signaling overhead in relation to the allocation period for when $\overline{\gamma} = 9$ dB

crucial for effective servicing of small packets such as VoIP packets. When the allocation period increases from 4 frames to 8 frames, the average throughput increases because the MAC overhead ratio and the signaling overhead both decrease while the scheduling delay increases. The proposed analytical model, though limited to the downlink in this study, can also be applied to the uplink.

7. References

Alouini, M.-S. & Goldsmith, A. J. (2000). Adaptive modulation over nakagami fading channels, *Wirel. Pers. Commun.* 13(1/2): 119–143.

Baiocchi, A., Melazzi, N. B., Listanti, M., Roveri, A. & Winkler, R. (1991). Loss performance analysis of an ATM multiplexer loaded with high-speed ON-OFF sources, *IEEE J. Select. Areas Commun.* 9(3): 388–393.

Ben-Shimol, Y., Chai, E. & Kitroser, I. (2006). Efficient mapping of voice calls in wireless OFDMA systems, *IEEE Commun. Lett.* 10(9): 641–643.

Bi, Q., Chen, P.-C., Yang, Y. & Q.Zhang (2006). An analysis of VoIP service using 1xEV-DO revision A system, *IEEE J. Sel. Areas Commun.* 24(1): 36–44.

Ghosh, A., Wolter, D. R., Andrews, J. G. & Che, R. (2005). Broadband wireless access with WiMax/802.16: Current performance benchmarks and future potential, *IEEE Commun. Mag.* pp. 129–136.

Gross, J., Geerdes, H.-F., Karl, H. & Wolisz, A. (2006). Performance analysis of dynamic OFDMA systems with inband signaling, *IEEE J. Sel. Areas Commun.* 24(3): 427–436.

Heffes, H. & Lucantoni, D. M. (1986). A Markov modulated characterization of packetized voice and data traffic and releated statistical multiplexer performance, *IEEE J. Select. Areas Commun.* SAC-4(6): 856–868.

Huang, J., Le-Ngoc, T. & Hayes, J. F. (1996). Broadband satcom system for multimedia services,

Proc. IEEE ICC, pp. 906–909.

IEEE (2009). IEEE standard for local and metropolitan area networks, part 16: Air interface for fixed broadband wireless access systems, *IEEE 802.16-2009 Std.* .

IEEE (2010). IEEE standard for local and metropolitan area networks, part 16: Air interface for fixed broadband wireless access systems, Advanced air interface, *IEEE 802.16m/D4 Std.* .

Jiang, D., Wang, H., Malkamaki, E. & Tuomaala, E. (2007). Principle and performance of semi-persistent scheduling for VoIP in LTE system, *Proc. IEEE WiCom*, pp. 2861–2864.

Kwon, T., Lee, H., Choi, S., Kim, J., Cho, D.-H., Cho, S., Yun, S., Park, W.-H. & Kim, K. (2005). Design and implementation of simulator based on a cross-layer protocol between MAC and PHY layers in a WiBro compatible IEEE 802.16e OFDMA system, *IEEE Commun. Mag.* pp. 136–146.

Leiba, Y., Segal, Y., Hadad, Z. & Kitroser, I. (2006). Coverage/capacity simulations for OFDMA PHY in ITU-T channel model including MRC, STC, AAS results, *IEEE C802.16e-04/16* .

Liu, Q., Zhou, S. & Giannakis, G. B. (2005). Queuing with adaptive modulation and coding over wireless links: cross-layer analysis and design, *IEEE Trans. Wireless Commun.* 4(3): 1142–1153.

McBeath, S., Smith, J., Reed, D., Bi, H., Pinckley, D., Rodriguez-Herrera, A. & O'Connor, J. (2007). Efficient signaling for VoIP in OFDMA, *Proc. IEEE WCNC*, pp. 2247–2252.

Niyato, D. & Hossain, E. (2005a). Queue-aware uplink bandwidth allocation for polling services in 802.16 broadband wireless networks, *Proc. IEEE Globecom*, pp. 3702–3706.

Niyato, D. & Hossain, E. (2005b). Queueing analysis of OFDM/TDMA systems, *Proc. IEEE Globecom*, pp. 3712–3716.

Ozer, S. Z., Papavassiliou, S. & Akansu, A. N. (2000). On performance of switching techniqueues for integrated services in CDMA wireless systems, *Proc. IEEE VTC*, pp. 1967–1973.

Razavilar, J., Liu, K. J. R. & Marcu, S. I. (2002). Jointly optimized bit-rate/delay control policy for wireless packet networks with fading channels, *IEEE Trans. Commun.* 50(3): 484–494.

Shah-Heydari, S. & Le-Ngoc, T. (1998). MMPP modeling of aggregated ATM traffic, *Proc. IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 129–132.

Shrivastava, S. & Vannithamby, R. (2009a). Group scheduling for improving VoIP capacity in IEEE 802.16e networks, *Proc. IEEE VTC*, pp. 1–5.

Shrivastava, S. & Vannithamby, R. (2009b). Performance analysis of persistent scheduling for VoIP in WiMAX networks, *Proc. WAMICON*, pp. 1–5.

So, J. (2008). Performance analysis of VoIP services in the IEEE 802.16e OFDMA system with inband signaling, *IEEE Trans. Veh. Technol.* 57(3): 359–363.

So, J. (2009). Performance analysis of a semi-fixed mapping scheme for VoIP services in wireless OFDMA systems, *Proc. International Conference on Wireless and Mobile Communications*, 13–17.

Wan, L., Ma, W. & Guo, Z. (2007). A cross-layer packet scheduling and subchannel allocation scheme in 802.16e OFDMA system, *Proc. IEEE WCNC*, pp. 1867–1872.

VoIP Technologies

Edited by Dr Shigeru Kashihara

ISBN 978-953-307-549-5

Hard cover, 336 pages

Publisher InTech

Published online 14, February, 2011

Published in print edition February, 2011

This book provides a collection of 15 excellent studies of Voice over IP (VoIP) technologies. While VoIP is undoubtedly a powerful and innovative communication tool for everyone, voice communication over the Internet is inherently less reliable than the public switched telephone network, because the Internet functions as a best-effort network without Quality of Service guarantee and voice data cannot be retransmitted. This book introduces research strategies that address various issues with the aim of enhancing VoIP quality. We hope that you will enjoy reading these diverse studies, and that the book will provide you with a lot of useful information about current VoIP technology research.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Jaewoo So (2011). Scheduling and Capacity of VoIP Services in Wireless OFDMA Systems, VoIP Technologies, Dr Shigeru Kashihara (Ed.), ISBN: 978-953-307-549-5, InTech, Available from: http://www.intechopen.com/books/voip-technologies/scheduling-and-capacity-of-voip-services-in-wireless-ofdma-systems

INTECH
open science | open minds