RESEARCH   **Open Access**

An energy-efficient opportunistic multicast scheduling based on superposition coding for mixed traffics in wireless networks

Xing Zhang[*], Qinzi Li, Yu Huang and Wenbo Wang

Abstract

Energy efficient transmission in wireless networks is an important issue for green communications which contribute not only to goals for sustainable development, but also to the profitability of the telecommunication industry. In this article, an energy-efficient opportunistic multicast scheduling is proposed, which is based on superposition coding (SC) for multiple mixed unicast and multicast traffics transmission in wireless networks. The proposed scheme can improve the system energy efficiency while reducing traffic transmission delay and effectively guaranteeing quality of service (QoS) of the mixed unicast and multicast traffics. In addition, the scheme provides a channel gain threshold for enhanced layer transmission of the SC, which can dynamically optimize system throughput and energy-efficiency. To verify the performance advantages of our proposed scheme, this article also analyzes and compares the other two traditional schemes. Both theoretical analysis and simulation results show that the proposed scheme achieves more energy efficiency gain than traditional schemes while guaranteeing the quality of the mixed unicast and multicast traffics.

1 Introduction

The traditional protocol and algorithm design and development in wireless networks is towards maximization of the performance observed by the end-user, in terms of perceived throughput, delay, quality of service (QoS), etc. This trend does not consider the power consumed by wireless devices and networks which creates a gap between the energy a wireless network needs to operate and the battery capacity of the wireless devices. Hence, the requirement of Energy-Efficiency [1] appears as an extremely important property of new protocols for wireless networks with battery-powered mobile nodes. Moreover, Energy-Efficiency is the tool to realize the vision of green wireless networks, which are deemed important these days due to the increasing share of wireless systems of the total energy expended in communications and networking systems [2-6].

Meanwhile, for the past few decades, wireless cellular telecommunication has become an indispensable part of our modern society. More than four billion subscribers around the world depend on their mobile phones for their private and professional lives. With the development of technologies in 3G and LTE/4G telecommunication systems and smart phones such as iPhone/ Android, people's demand for multimedia communication, such as image and video, is increasingly growing, so is the quality requirements for multimedia communication in wireless systems. However, that comes at the cost of increasing energy/power consumption. Thus far, the telecommunication industry has focused on highly energy efficient transmission technologies, while achieving the goal of a high data rate and low cost, a factor that will strongly contribute to the global success of communication. Although the adoption of more efficient transmission technologies, such as OFDM, MIMO, and Turbo coding etc., the frequency efficiency has been improved greatly, the cost of data transmission in wireless telecommunication is still extremely high due to more bandwidth requirement, which is seen as the most precious resource in wireless telecommunication.

For wireless cellular communications system, according some statistics, more than 70% energy of the total wireless networks (including terminal, radio access network and core network, etc.) are consumed in the base

* Correspondence: hszhang@bupt.edu.cn
Wireless Signal Processing and Networks Lab (WSPN), Key Lab of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications (BUPT), P.O. Box 93, Beijing 100876, China
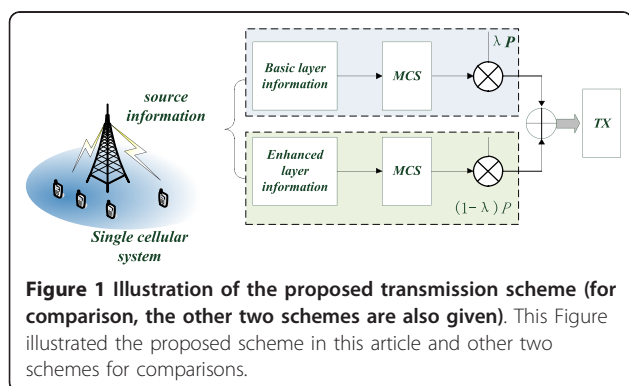
station subsystem (BSS). Thus in order to design an energy-efficiency mobile networks, one important efforts should be focused on BSS. Due to social nature and habits of human-being, users close in vicinity (e.g., the coverage of several base stations) will have similar habits, behavior and mobility rules. For a given geographical area and a certain time period, a group users/subscribers will probably request the same traffics, thus the system can uniformly transmit the traffic to all the users through multicasting instead of unicasting and more energy will be saved. In this way, the user behavior/pattern and traffic variations can be used as the basis for the design of green communication system. Motivated from this phenomenon we present the concept of *group users behavior* (GUB), which characterizes the general behavior, pattern and rules of a group of users in a certain geographical area for a certain time window, for example, user requirements, users social character, user traffics fluctuations (temporal and spatial domain), users mobility, etc. Based on GUB, energy-efficient transmission mode design can be realized.

In wireless broadcasting system data transmission cost is very low, due to that in the broadcast network the same multimedia data is broadcasted only once and all the users in a certain coverage area can receive. However, in order to guarantee the data is transmitted successfully, it needs to send multiple times. Considering the traffics/services without rigorous QoS requirements, for example, best-effort traffics or "soft-real-time (SRT)" traffics [7], such as Web page browsing, file downloading, etc. Such kind of traffics have the characteristics of low latency requirements and delay-tolerance, independent creation on demand and reception of multiple accessions. Therefore, a transmission scheme of selecting the same kind of unicast traffic in a certain time period then transmitting using broadcast/multicast was proposed in [7] (denoted as *Scheme A* thereinafter, shown in lower-left of Figure 1), which is to cache all the services in a certain time, selecting the same kind of traffics after neglecting traffics requested by fewer users,

and transmit through broadcast/multicast in each time. This scheme can save bandwidth and transmission power for nearly every kind of traffics especially when the traffic is demanded by a lot of users. Essentially, this scheme is to obtain energy efficiency at the expense of latency. As for different kinds of traffics, the scheme is supposed to transmit by time division multiplexing (TDM).

However, this scheme aims at improving system throughput and still does not consider the energy efficiency which can be optimized furthermore. In [8,9], the authors pointed out that there is no need to transmit by full power for most of users as there exist some users with good channel conditions, so the current situation leads to energy inefficiency. It proposed a scheme of superimposing multicast traffic onto unicast (which is known as basic multicast service: BMS and enhanced multicast service: EMS), which can improve system throughput efficiently. Combing the ideas above and considering the situation of SRT services in a certain time, it is believed that a superposition coding (SC) scheme for unicast and multicast (denoted as *Scheme B* thereinafter) maybe helpful. This scheme can be implemented as follows: the best-effort services are assorted by different categories first, and then be divided into hot traffics and non-hot traffics based on a certain user number threshold. As for the hot traffics are the main service objects, non-hot traffics are superimposed onto hot traffics. It can be anticipated that this scheme can improve throughput significantly.

Furthermore, there are also some inefficiencies in this scheme. Although the unicast and multicast SC scheme can transmit BMS and EMS simultaneously, the BMS and EMS have to share the transmission power, which means each of them can only use a smaller portion of the total transmission power than that in the unicast scheme. Especially for the EMS, it only gets a very smaller portion of transmission power, which cannot guarantee adequate quality for the enhanced layer. On the other hand, the data rate of BMS is determined by the instantaneous worst channel user, which leads to a very low transmit data rate and system throughput, and the power allocated to BMS is part of the full transmission power, so that the latency of this scheme is not satisfied. An opportunistic multicast scheduling (OMS) scheme was proposed in [10] whose main advantage is to improve transmission rate and reduce transmission latency. To achieve the goal of users receiving the same information, conventional multicast causes a long delay because its data rate is determined by the worst user channel condition, and conventional unicast wastes system resources because it needs too many times of transmissions. The OMS scheme determines its transmission rate according to a certain percentage of the number of

Figure 1 Illustration of the proposed transmission scheme (for comparison, the other two schemes are also given). This Figure illustrated the proposed scheme in this article and other two schemes for comparisons.

users with good channel condition, so that it can achieve the goal after a smaller times transmission than unicast scheme. As the number of subscribers receiving information each time is higher than that of unicast scheme and the data rate is improved significantly, the transmission times is reduced greatly, which in turn latency is reduced. In [11], we first studied the optimized layered multicast with SC in cellular systems.

To resolve the problems mentioned in the conventional schemes and improve system performance, this article proposes an energy-efficient traffic opportunistic multicast transmission scheduling based on SC for multiple mixed unicast and multicast traffics. The scheme cannot only reduce transmission delay, but also effectively guarantee the quality of transmission service. In addition, the scheme provides a channel gain threshold for enhanced layer, which can optimize system throughput dynamically. Therefore, the new scheme realizes the goal of guaranteeing quality and optimizing energy efficiency. The main contributions of this article can be summarized as follows:

• To improve energy efficiency and system throughput, this article considers a unicast and multicast SC scheme, which superimposes the unicast or multicast information used by fewer users onto another multicast information used by a lot of users, after dividing all of the traffics into two kinds of hot and non-hot traffics based on a certain user number threshold.
• To resolve the problem of conventional SC scheme which does not guarantee the quality of the EMS, a new scheme combining OMS and SC is proposed that not only guarantees transmission quality, but optimizes the transmission latency as well.
• A channel gain threshold for enhanced layer transmission of the SC is introduced, which can dynamically optimize system throughput and energy-efficiency.

The rest of the article is organized as follows. In Section 2, the concepts of SC scheme and OMS scheme are introduced. The system model is presented in Section 3. In Section 4, we formulate the optimization problem and certificate it through simulation whose results are provided in Section 5. Finally, conclusions and future works are summarized in Section 6.
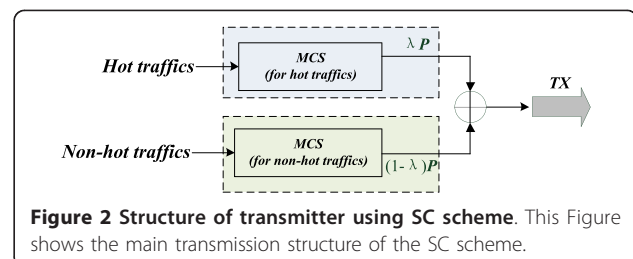
2 Background knowledge

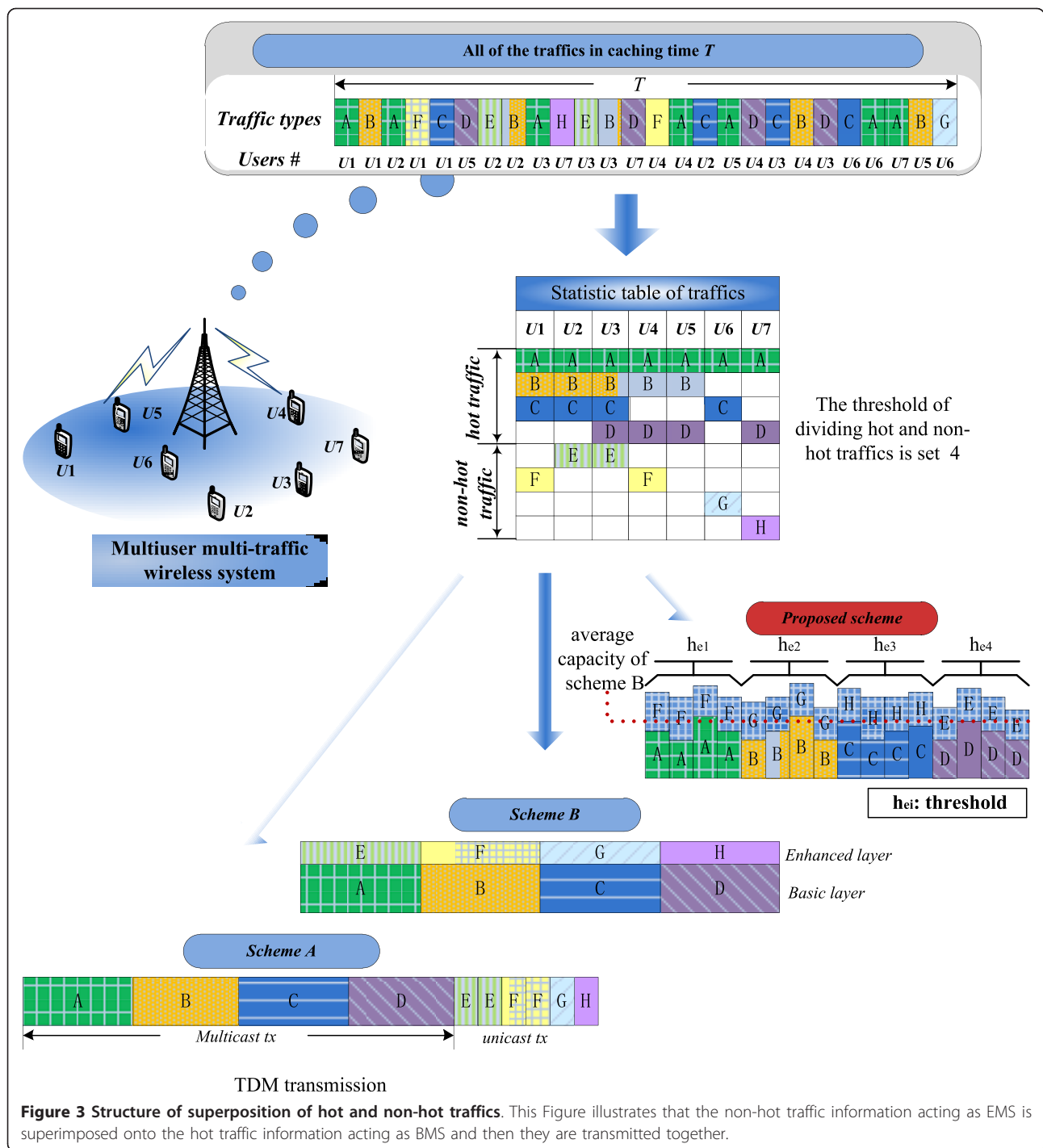### 2.1 Superposition coding and unicast and multicast superposition

Superposition coding (SC) was first proposed by Cover and Thomas [12] and its implementation in broadcast and multicast system can be found in [13]. This kind of method is to divide the source data into several streams with different rates, and send them using different level of modulation and coding schemes (MCS), where the lower layers have higher priorities which provides a basic description of the source multimedia traffic, and the higher layers have lower priorities which provides a detailed (enhanced) description for the multimedia traffic. Thus, the users even with a worse channel condition can acquire the basic information. As an example assuming a two-level SC model with two receivers, the transmission process is described as follows:

As shown in Figure 2, the basic multicast stream (BMS), which is the basic layer with higher priority, is assigned transmit power λP ($0.5 < \lambda < 1$), and the enhanced multicast stream (EMS), which is the enhanced layer with lower priority, is assigned transmit power $(1 - \lambda) P$.

In terms of different multiplexing techniques, in [14] it shows that the SC scheme [15,16] is apparently better than the orthogonalization schemes, such as time-division multiplexing (TDM) and frequency-division multiplexing (FDM). The performance gap between the SC and TDM (FDM) is more enlarged when the difference of the channel gains of the two receivers increases. For example, in the FDM scheme, two users with strong and weak channel gains orthogonally share the bandwidth, to guarantee the throughput requirement of the user with weak channel experience, the throughput of the strong user has to be reduced, which means that a considerable fraction of the bandwidth has to be allocated to the weak user to achieve an equal performance for all users, and this causes large degradation in the performance of the strong user. However, in the SC scheme, the strong user is allowed to use the full degrees of streams in the channel while the weak user can achieve the BMS at least. Therefore, SC can make full use of system resources and provide different users with different QoS requirements. Based on the above process, introducing SC into the traditional scheme B in actual system can not only be achieved but also improve system throughput. In the specific process, all of the SRT services can be first divided into two groups by a certain user number threshold, one is hot traffic group and the other is

Figure 2 Structure of transmitter using SC scheme. This Figure shows the main transmission structure of the SC scheme.

Figure 3 Structure of superposition of hot and non-hot traffics. This Figure illustrates that the non-hot traffic information acting as EMS is superimposed onto the hot traffic information acting as BMS and then they are transmitted together.

non-hot traffic group. The non-hot traffic information acting as EMS is superimposed onto the hot traffic information acting as BMS and then they are transmitted together. The process is shown in Figure 3.

After receiving the SC information, users first perform demodulation, and then decode the BMS, which is the hot traffic information, when the EMS acts as interference. With the successful decoding of BMS, after removing the hot traffic from the decoded information by interference cancelation technology, users can decode the EMS which is non-hot traffic information. Finally, subscribers attain all of the transmitted source information. Note that if the BMS cannot be decoded successfully, the EMS will also not be received.

2.2 Opportunistic multicast scheduling

Besides conventional TDM and FDM schemes, Baek et al. [17] proposed an adaptive transmission scheme for mixed multicast and unicast traffics based on SNR threshold. In such adaptive scheme, the data rate of the multicast transmission depends on the instantaneous channel of the worst user, while the unicast transmission selects a user with the instantaneous best channel, exploiting wireless channel variation and achieving a multiuser diversity gain [18,19]. However, the adaptive transmission is basically a TDM scheme, so it has the same disadvantages as conventional TDM scheme. Considering in the conventional multicast transmission scheme, the data rate is determined by the channel condition of the worst user. Although all users can receive information for one time transmission, the data rate will be significantly lower.

On the other hand, if the data rate is determined by the best channel condition among the multicast users, although the data rate is high, only the user with the best channel gain can successfully receive the information. To exploit multiuser diversity gain and multicast gain simultaneously, an opportunistic multicast scheduling (OMS) scheme was proposed in [10]. That is, rather than selecting a transmission rate depending on the worst or best channel user, the user with a median channel gain is selected which means that the OMS scheme can serve the best 50% of users in each transmission. Thus, its transmission rate is determined by the median channel condition among all the users, which will be higher than that of the conventional multicast scheme. In [13], a proportional fair (PF) scheduling scheme for multicast service was proposed. Its approach to improve system throughput is to divide a multicast group into two subgroups and transmit only for the subgroup with better channel condition. In addition, Low et al. [12,20] studied the optimal user selection ratio to minimize the multicast latency required for transmitting a finite-length fountain-encoded message to all users in a cellular network. Based on the above analysis, introducing proportional fair scheduling scheme and optimal user selection ratio scheduling into OMS scheme can not only reduce the transmission latency, but also obtain multiuser diversity gain and multicast gain. In an actual system, as for a certain message to be transmitted, it is assumed that the BS has already obtained the channel distribution information of all users in the cell according to channel feedbacks. Suppose a user selection ratio (p_u, its range is from 0 to 1), and then we can get the transmission rate based on the user distribution information and the user selection ratio. So the latency in the first transmission is $S/(BR_1)$, where S is the data length, and B is the bandwidth. R_1 is the transmission rate for the first transmission time.

Similar to the first transmission, the data rate is determined by the top p_u percent users' channel conditions in each time transmission. After all of the users have received successfully, the transmission is complete, and the total latency of the message transmission can thus be expressed as: $D = \sum_i S/BR_i$.

It can be seen that the latency is the function of user selection ratio, which can be minimized through optimizing the user selection ratio. As for the OMS, its transmission rate is determined by the median user channel condition, which is higher than that of the conventional multicast scheme. So even it transmits several times, the latency is still lower than the TDM scheme [11].

3 System model and scheme description
3.1 Scheme description
This article proposes an energy-efficient opportunistic multicast transmission scheduling based on SC for multiple mixed unicast and multicast traffics. For comparison, this article also analyzes the process and performance of other two conventional schemes in details, namely, scheme A and scheme B, which are described in Section 2. The basic process of the three schemes are illustrated in Figure 1. As shown in Figure 1, a single-cell network with a base station (BS) providing services to multiple users (denoted as $U1$, $U2$, ..., Un) with best-effort traffics/services in a certain time T. Through buffering all services requests in a certain time (determined by the latency requirement), the BS statistically analyzes the kinds of the services and determines the corresponding users of every kind of services. The BS ranks all of the services according to the user number the service is requested, and then divides them into hot services and non-hot services based on a certain user number threshold. Figure 1 illustrates the principles of the three schemes that we will explain in details. Taking Figure 1 as an example, in a certain buffering time there are 26 traffic/service requests, in which there are 8 kinds of services. Setting four as the request threshold, then A, B, C, and D are hot services; E, F, G, and H are non-hot services. Note that the number of traffic C and D are the same, and E and F are the same.

In Figure 1, Scheme A is a kind of transmission that selects the same kind of unicast traffic in a certain time then transmits them using broadcast/multicasting. Finally the scheme becomes to only multicast those traffics used by a lot of users, and those being seldom used are transmitted by unicast. Scheme B is based on the idea of applying the method of mixing unicast and multicast by SC. The process is superimposing those non-hot traffics onto those hot traffics by SC and transmitting them together. Because there is no traffic selection method, the scheme B superimposes the non-hot

services onto the hot services by the sequences in the statistic table. If the kinds of non-hot traffics are larger than that of hot traffics, the rest non-hot traffics will be transmitted by unicast. Otherwise, the rest hot traffics will be sent by traditional multicast.

In the energy-efficient opportunistic multicast transmission scheduling based on SC proposed in this article, the basic layer is transmitted according to the sequences in the traffic statistic table, as shown in Figure 1. In each transmission, the optimized user selection ratio should be determined first which can minimize the transmission delay, and then the transmission rate can be obtained based on the user selection ratio. After obtaining the user selection ratio of basic layer and the power allocation coefficient, the throughput can be maximized through choosing a certain traffic as the enhanced layer traffic whose channel gain is larger than the threshold. For example, in the "Proposed scheme" of Figure 1, during the transmission of traffic A, the optimized user selection ratio is set as 60%, which means all of the users can receive the traffic with success rate of more than 95% after 4 times of transmission (i.e., $1 - (1 - 60\%)^4 \geq 95\%$). The total transmission time during the process is considerably lower than the other two schemes. Based on the optimized user selection ratio of basic layer and the power allocation coefficient, we can get a threshold h_e which can optimize the throughput for each time transmission. During all of the potential enhanced traffics, only the channel gain of the top 60% user requesting traffic E is above the threshold, so we choose E to be superimposed onto the traffic A in this transmission. Signing the average throughput of Scheme B as a red dashed line in Figure 1, we can see that by the increase in throughput brought by the chosen enhance layer, the proposed scheme gets a significant advantage in throughput comparing to the other two schemes. The proposed scheme can not only reduce transmission delay for setting its transmission rate by a certain middle level channel condition, but also increase system throughput by choosing suitable enhanced layer traffic, and it finally achieves better performance on energy efficiency than the other two schemes.

3.2 Theoretical fundamentals

For the wireless transmission model we use the commonly used model in current literatures [10,11], here to better describe the proposed scheme we re-write some equations used in [10,11]. Assuming all users are uniformly-distributed in a single-cell scenario with cell radius of r_c. The probability distribution function (pdf) of the distance from BS to the users is written as: $f_r(r) = 2r/r_c^2, 0 < r \leq r_c$, r is the distance between BS and users. And the received SNR value z at the receiver

is expressed as $z = G \times \dfrac{\alpha^2}{r^n} \times \dfrac{P}{N}$, where P is the transmit power, N is the power of additive white Gaussian noise (AWGN), G is a parameter that reflects the effect of antenna gain and other factors. The channel gain $|h|^2 = \alpha^2/r^n$ reflects the effects of path-loss and small-scale fading, the coefficient n denotes the path loss exponent. We denote $\rho_0 = GP/N$, so the system capacity is:

$$C = B \log_2(1 + \rho_0|h|^2). \tag{1}$$

Consider a transmitter with power P and two different receivers, one with Gaussian noise power N_1 and the other with N_2. Without loss of generality, assuming $N_1 > N_2$. The model of a two-user Gaussian broadcasting channel is characterized as one transmitter sending a superimposed message $x = x_1 + x_2$ (x_1 can be regarded as basic layer and x_2 as enhanced layer) to the two receivers who will receive $y_1 = x + z_1$ and $y_2 = x + z_2$, where z_1 and z_2 are Gaussian distributed random variables with variance N_1 and N_2, respectively. The transmitter wishes to transmit two independent messages at rates R_1 and R_2, the received messages are y_1 and y_2, usually denoted as basic layer and enhanced layer, respectively. The transmit power allocated to the message x_1 is λP, and the transmit power allocated to the message x_2 is $(1 - \lambda) P$, where the λ is the power allocation ratio for the different messages.

The capacity of the basic layer and enhanced layer in the Gaussian broadcast channel are written as:

$$R_1 \leq \log_2\left(1 + \frac{\lambda P}{(1-\lambda)P + N_1}\right), \quad R_2 \leq \log_2\left(1 + \frac{(1-\lambda)P}{N_2}\right). \tag{2}$$

Accordingly, the SNR at the receiver for basic layer and enhanced layer are expressed as: $z_b = \dfrac{\alpha^2/r^n \cdot \lambda\rho_0}{\alpha^2/r^n \cdot (1-\lambda)\rho_0 + 1}$ and $z_e = \dfrac{\alpha^2}{r^n}(1-\lambda)\rho_0$, respectively. We denote the channel gain as $|h|^2 = \alpha^2/r^n = X/Y$, then its probability density function (p.d.f.) can be expressed as:

$$f(z) = \int_0^{r_c^n} \gamma f_{X,Y}(\gamma z, \gamma) d\gamma = \frac{2}{nr_c^2} z^{-2/n-1} \gamma(2/n + 1, r_c^n z).$$

The cumulative distribution function (c.d.f.) of the channel gain $|h|^2$ can be calculated as:

$$F(z) = \int_{\gamma=0}^{r_c^n} \int_{x=0}^{\gamma z} f_{X,Y}(x, \gamma) dx d\gamma = 1 - \frac{2}{nr_c^2} z^{-2/n} \gamma(2/n, r_c^n z). \tag{3}$$

where $\gamma(p, q)$ is the incomplete gamma function as $\gamma(p, q) = \int_0^q t^{p-1} \exp(-t) dt$.

As for introducing SC into the OMS scheme, let's assume the *user selection ratio* of BS transmitting BMS for the first time is $p_b(1)$, and the transmission rate is

determined by $R_b(1) = \text{Func}(\lambda, p_b(1))$ which is a function of power allocation coefficient λ and user selection ratio $p_b(1)$. Because of broadcast transmission, here the *user selection ratio* is defined as the percentage of users whose data rate requirements will be met when transmitter sets a certain transmission rate. Since that the user selection ratio of enhanced layer is also depended on the basic layer, the transmission delay of information is determined by the basic layer. The transmission delay of the first transmission can be expressed as $d(1) = S/(BR_b(1))$, where S is the basic layer information length, and B is bandwidth. Similarly, the transmission delay of the ith transmission is $d(i) = S/(BR_b(i))$. After completing m transmissions, the total transmission delay can be expressed as:

$$D = \sum_{i=1}^{m} d(i) = \sum_{i=1}^{m} \frac{S}{BR_b(i)}. \tag{4}$$

As for the throughput in the OMS scheme, it contains both BMS and EMS each time and the throughput is given as $C_i(\lambda) = R_b(i) \times p_b(i) + R_e(i) \times p_e(i)$. The total throughput is then expressed as $C = \sum_{i=1}^{m} C_i(\lambda)$.

4 Theoretical analysis
4.1 High energy-efficiency OMS based on SC for mixed traffics scheme

Let the transmitter's power be P, and $\lambda(0 \leq \lambda \leq 1)$ be the power allocation coefficient for SC that indicates how much power is allocated to a BMS, thus the transmit power allocated to a BMS is λP and the transmit power allocated to an EMS is $(1 - \lambda)P$. Take the i-th ($i = 1, 2,...$) time transmission for example, the user selection ratio is $p_b(i)$ and the transmit rate is $R_b(i)$. If the user's received data rate is less than $R_b(i)$, this transmission of this time is considered be be interrupted or outage. From (3), the outage probability of the basic layer is (noted that in the following the subscript i representing the ith transmission is neglected for simplicity):

$$P_b^{\text{out}} = F\left(\log_2(1 + \frac{\lambda \rho_0 |h|^2}{(1 - \lambda)\rho_0 |h|^2 + 1}) \leq R_b\right)$$

$$= F\left(|h|^2 \leq \frac{(2^{R_b} - 1)}{[1 - (1 - \lambda)2^{R_b}]\rho_0}\right) = 1 - p_b. \tag{5}$$

Substituting (3) into (5), we obtain the transmission rate for the basic layer:

$$R_b = \log_2\left(1 - \frac{\lambda \rho_0 \dfrac{\ln p_b}{r_c^n}}{1 - (1 - \lambda)\rho_0 \dfrac{\ln p_b}{r_c^n}}\right). \tag{6}$$

Thus, according to (4) we have the total transmission delay as follows:

$$D = \sum_{i=1}^{m} \frac{S}{B \log_2\left(1 - \dfrac{\lambda \rho_0 \ln p_b(i)/r_c^n}{1 - (1 - \lambda)\rho_0 \ln p_b(i)/r_c^n}\right)}. \tag{7}$$

When the outage probability for the service transmission is P_b^{out} and when the total transmission times m meets $\prod_{i=1}^{m}(1 - p_b(i)) \leq P_b^{\text{out}}$, thus the service completes its transmission and its outage performance can be guaranteed.

Through Lagrangian algorithm, we can get an optimal m^{opt} from (7), which can meet the demand of minimizing transmission delay. Here for the OMS based on SC, the user selection ratio for each transmission is assume to be the same for simplicity. Under this condition, then the user selection ratio can be expressed as $p_b^{\text{opt}} = 1 - (P_b^{\text{out}})^{\frac{1}{m^{\text{opt}}}}$. From (5), we can easily obtain the optimal user selection ratio p_b^{opt} which minimizing transmission delay.

As for the proposed scheme, the traffics of BMS and EMS are different, so that the scheme does not totally coherent with the characteristics of SC, which means that even though the EMS can be decoded only by the successful decoding of BMS the user selection ratio for EMS does not need to be less than that of BMS. Exactly on the opposite, in order to let the EMS be transmitted successfully for the same transmission times as the BMS, the user selection ratio of EMS should be no less than BMS. In this way the user selection ratio of the enhanced layer p_e can be expressed as:

$$p_e = 1 - F\left(|h|^2 \leq \frac{(2^{R_e} - 1)}{(1 - \lambda)\rho_0}\right) \geq p_b^{\text{opt}}. \tag{8}$$

We define system throughput as the data amount transmitted by BS in a certain time. In our proposed mixed traffics transmission scheme, at each transmission time there are both BMS and EMS services, which makes the transmission rate very high. The system throughput can be expressed as:

$$C_{\text{opt}}(\lambda) = m\left\{\log_2\left(1 + \frac{\lambda \rho_0 |h_b|^2}{(1 - \lambda)\rho_0 |h_b|^2 + 1}\right) \times p_b^{\text{opt}}\right.$$

$$\left. + \log_2\left(1 + (1 - \lambda)\rho_0 |h_e|^2\right) \times p_e\right\}. \tag{9}$$

where the channel gain for the BMS user is $|h_b|^2 = \dfrac{(2^{R_b} - 1)}{[1 - (1 - \lambda)2^{R_b}]\rho_0}$.

Therefore, our problem is to find a scheme that maximizes the average system throughput. In the process of minimizing transmission delay, the user selection ratio of BMS and the corresponding user channel gain has been determined. Through Lagrangian algorithm, the optimal user selection ratio for an EMS can be obtained from (8) and (9) as:

$$p_e^{opt} = -\frac{(1 + (1-\lambda)\rho_0|h_e|^2)\log(1 + (1-\lambda)\rho_0|h_e|^2)\ln 2}{(1-\lambda)\rho_0}$$
$$\cdot \frac{1}{\partial|h_e|^2/\partial p_e^{opt}}. \quad (10)$$

It should be noted that (10) is still not a closed-form equation. And the user selection ratio p_e^{opt} must be chosen to meet the requirement of $p_e^{opt} \geq p_b^{opt}$. Combing (3), (8), and (10) the optimal channel gain threshold $\left|h_e^{opt}\right|^2$ for choosing EMS is calculated as follows:

$$\left|h_e^{opt}\right|^2 = \left\{\frac{2}{p_e^{opt}nr_c^2}\gamma\left(2/n, r_c^n\left|h_e^{opt}\right|^2\right)\right\}^{n/2}. \quad (11)$$

Similar to (10), (11) is also not a closed-form equation, however, we can obtain $\left|h_e^{opt}\right|^2$ using iterations.

Substituting (11) into (10), we can get the optimal user selection ratio for EMS as p_e^{opt}, and then we can have the optimal system throughput in (9). Unless the p_e^{opt} channel gain of a certain kind of non-hot service is higher than the threshold in (11), this kind of service can be superimposed on hot service and be transmitted.

We define energy efficiency as the energy consumption per unit system throughput. In the proposed scheme, according to the definition, we have the energy efficiency expression in a time T:

$$EE_{opt} = \frac{P \cdot T}{C_{opt}(\lambda)}$$
$$= \frac{P \cdot T}{\log_2\left(1 + \frac{\lambda\rho_0|h_b|^2}{(1-\lambda)\rho_0|h_b|^2 + 1}\right) \times p_b^{opt} + \log_2(1 + (1-\lambda)\rho_0\left|h_e^{opt}\right|^2) \times p_e^{opt}}. \quad (12)$$

4.2 Scheme A

In scheme A of changing unicast to multicast, multicast information and unicast information are transmitted separately and they use full transmission power, respectively. Let the worst channel Gaussian noise power in broadcast users is N_1, and the Gaussian noise power of unicast user be N_2, according to delay expression, we have the transmit delay:

$$D_{schemeA} = \frac{S}{B\log_2(1 + GP|h_b|^2/N_1)} + \frac{S}{B\log_2(1 + GP|h_e|^2/N_2)}. \quad (13)$$

Comparing to the other two schemes, the transmission time of multicast and unicast is half because of its TDM transmission model. Thus for scheme A the system throughput is given as:

$$C_{schemeA}(\lambda) = m/2\{\log_2(1 + GP|h_b|^2/N_1) + \log_2(1 + GP|h_e|^2/N_2)\}. \quad (14)$$

According to the definition of energy efficiency and combining (13) and (14), the energy efficiency of this scheme can be expressed as:

$$EE_{schemeA} = \frac{PT}{C_{schemeA}(\lambda)} = \frac{2PT}{\log_2(1 + GP|h_b|^2/N_1) + \log_2(1 + GP|h_e|^2/N_2)}. \quad (15)$$

4.3 Scheme B

Scheme B is the hot over non-hot traffic SC transmission scheme, let the worst channel Gaussian noise power in basic layer multicast business be N_1, we have the transmission delay of a certain service as (16), for the delay is determined by basic layer in this scheme.

$$D_{schemeB} = \frac{S}{B\log_2\left(1 + \frac{\lambda GP|h_b|^2}{(1-\lambda)GP|h_b|^2} + N_1\right)}. \quad (16)$$

In this scheme, there are BMS and EMS in each transmission, and its transmission rate depends on the worst user channel condition of basic layer. Under a certain power allocation coefficient, the user with the worst channel condition cannot receive EMS completely which leads to outage. Considering the same transmission time as the other two schemes, the throughput in this scheme can be expresses as:

$$C_{schemeB}(\lambda) = m\left\{\log_2\left(1 + \frac{\lambda GP|h_b|^2}{(1-\lambda)GP|h_b|^2 + N_1}\right) + \log_2(1 + (1-\lambda)\rho_0|h_{es}|^2) \times (1 - P_{es}^{out})\right\}. \quad (17)$$

where P_{es}^{out} is the outage probability of EMS in this scheme, and $|h_{es}|^2$ denotes the corresponding enhanced layer user channel gain. As for the basic layer, its transmission rate is determined by the worst channel condition, so the outage probability is 0; while the outage probability of enhanced layer is depending on the worst channel condition of enhance layer users.

According to the definition of energy efficiency and combining (16) and (17), the energy efficiency of this scheme can be expressed as:

$$EE_{schemeB} = \frac{PT}{C_{schemeB}(\lambda)}$$
$$= \frac{PT}{\log_2\left(1 + \frac{\lambda GP|h_b|^2}{(1-\lambda)GP|h_b|^2 + N_1}\right) + \log_2(1 + (1-\lambda)\rho_0|h_{es}|^2) \times (1 - P_{es}^{out})}. \quad (18)$$

5 Simulation results and analysis

To verify the performance of the our proposed scheme, simulations are carried out for the three schemes with mixed traffics, detailed simulation parameters are given

in Table 1. We can see from the results that the new scheme has advantages not only on transmission delay and system throughput, but also energy efficiency.

Figure 4 shows the simulation results of the three schemes under different multicast traffic ratio, where the multicast traffic ratio is defined as the ratio of the number of all kinds of traffics divided by the number of the traffics exceeding multicast threshold. Setting system bandwidth as 10 MHz, the power allocation coefficient of BMS is 0.8 and the user selection ratio is 70% in our simulation. The transmission delay is defined as the latency needed for transmitting the same amount of information, which is inverse to the transmission rate.

From the figure on the left of Figure 4 it can be seen that when the multicast traffic ratio is 0, which indicates that all of the traffics are unicast, the transmission delay of the three schemes are the same. When the multicast traffic ratio is a little more than 0, the transmission delay of the three schemes decreases significantly and continues to decrease with the increase of multicast traffic ratio. From the figure on the right of Figure 4 it can be seen that the delay of scheme B is the highest, and our proposed scheme achieves the lowest transmission delay. That is because the SC scheme can save nearly half transmission times comparing to TDM scheme. Although scheme B can transmit BMS and EMS synchronously, the BMS and EMS have to share the transmission power, which means each of them can only use a smaller portion of power than that in the unicast scheme, so the data rate is lower which leads to a higher delay. As for the proposed scheme, its transmission rate is determined by a certain top percentage of users with good channel condition, so the transmission delay is significantly small even though it has to send several times.

Note that when the multicast traffic ratio is 100%, scheme B can use all of the transmission power because of no unicast information needed to be superimposed on those multicast traffic, so it has the same delay with the scheme A. As for the proposed scheme, it still

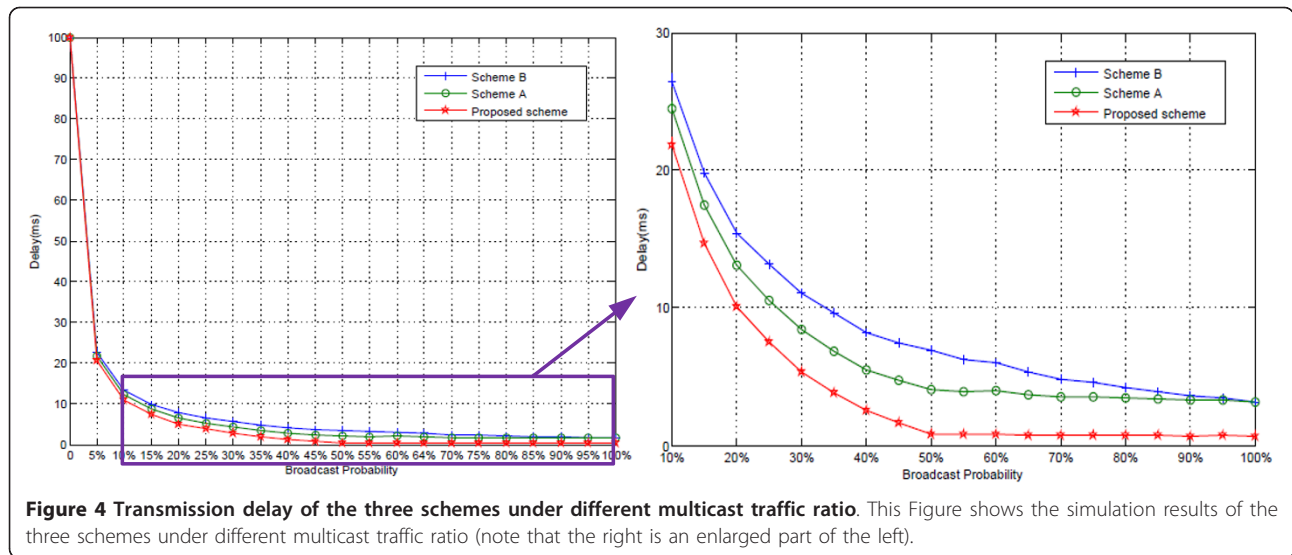transmits through several times, so the transmission delay is smaller.

It can also be seen from the figure that when the multicast traffic ratio comes to 50%, the transmission delay decreases slightly with the increase of multicast traffic ratio. This is because the main part of transmission delay in the three schemes comes from non-hot traffics. When the multicast traffic ratio reaches 50%, which means that no non-hot traffics need to be transmitted separately, the transmission delay has already been very low, so there is only slight decrease with the increase of multicast traffic ratio.

Figure 5 shows the system throughput results of the three schemes under different multicast traffic ratio. Here system throughput is defined as the amount of data transmitted by BS in a certain time period. From the results it can be seen that the throughput of the three schemes increases significantly with the increase of multicast traffic ratio. When the multicast traffic ratio is 0, i.e., all of the services are unicast, the throughput of the three schemes are the same. When the multicast traffic ratio is not 0, the throughput of the proposed scheme is far more greater than that of the other two schemes. That is because although multicast and unicast use all of the transmission power separately in scheme A, there are approximately double amount of data sent in SC scheme. As for scheme B, the transmission rate is determined by the worst channel condition and allocated a part of the transmission power, so that the increase brought by this scheme is not considerable. However, in the new scheme, the transmission rate is determined by a certain top percentage of users with good channel condition and the power allocation coefficient is determined by the result of optimizing system throughput, so the scheme makes an incredible improvement on throughput. Note that when the multicast traffic ratio is 100%, the scheme B can use all of the transmission power because of no unicast information needed to be superimposed on multicast traffic, so it has the same delay as scheme A.

Figure 6 shows the simulation result of transmission delay of the three schemes when the multicast traffic ratio is 50% which means that exactly all of the non-hot traffics can be superimposed onto hot traffics. From the figure it can be seen that when the same amount of data is transmitted, the delay needed in scheme B is the largest and the new scheme is smallest. The reason can be found from the explanation of Figure 4. Comparing with Figure 4, the value of the multicast traffic ratio in Figure 6 is approximately the best point for showing the gap between the new scheme and the other two schemes. This is because the main part of transmission delay in the three schemes comes from non-hot traffics. When the multicast traffic ratio reaches 50%, which

Table 1 Simulation parameters

Parameters	Value
System bandwidth (MHz)	10
Noise power (dBm/Hz)	-174
Pathloss model	$L = G/r^n$, r in meters, $G = 0.023568$, $n = 4$
Cell radius (m)	1000
Users number	100
Basic layer information length (bits)	1000
Enhanced layer information length (bits)	100
Time slot length (s)	0.01

Figure 4 Transmission delay of the three schemes under different multicast traffic ratio. This Figure shows the simulation results of the three schemes under different multicast traffic ratio (note that the right is an enlarged part of the left).
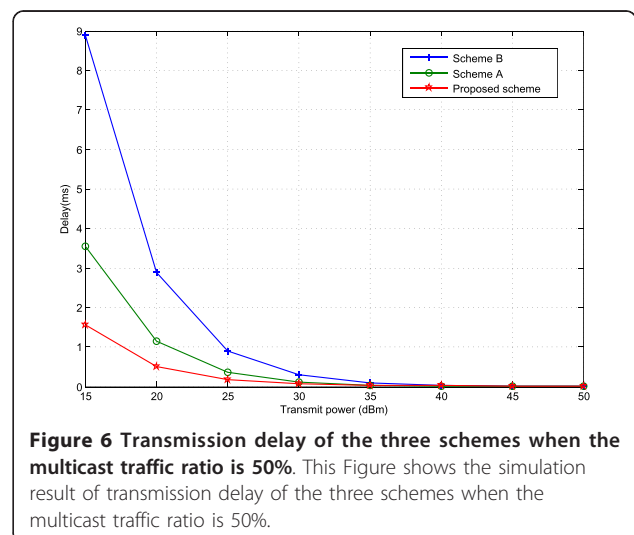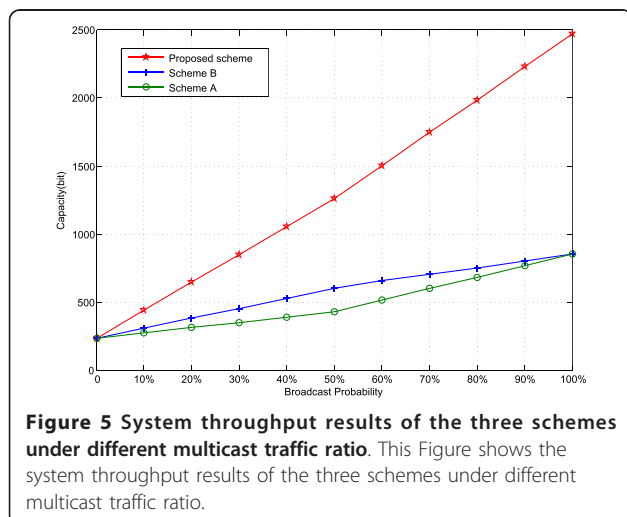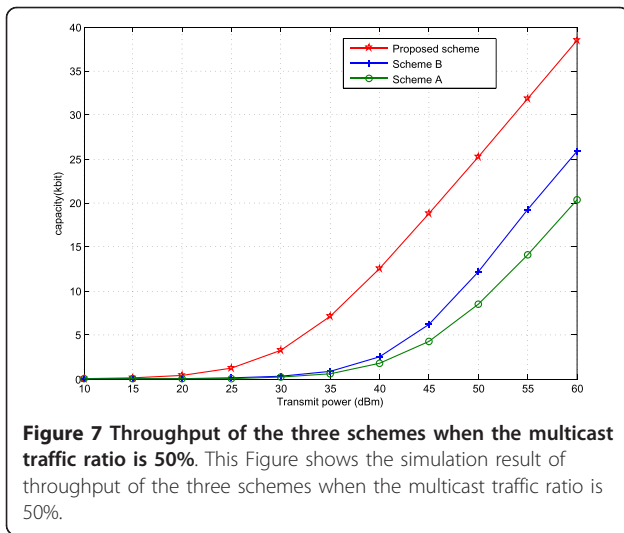
means that no non-hot traffics need to be transmitted separately, the new scheme cannot consider the non-hot traffics at all, which leads to the best performance of the new scheme.

Figure 7 shows the simulation result of throughput of the three schemes when the multicast traffic ratio is 50% which means that exactly all of the non-hot traffics can be superimposed onto the hot traffics. From the figure it can be seen that in the same transmission time the throughput of the proposed scheme is the largest, and the scheme A is the smallest. The reason can be found from the explanation of Figure 5. Comparing with Figure 4, the value of the multicast traffic ratio in Figure 7 isn't the biggest point for showing the gap between the new scheme and the other two schemes.

Figure 8 shows the simulation result of energy efficiency of the three schemes when the multicast traffic ratio is
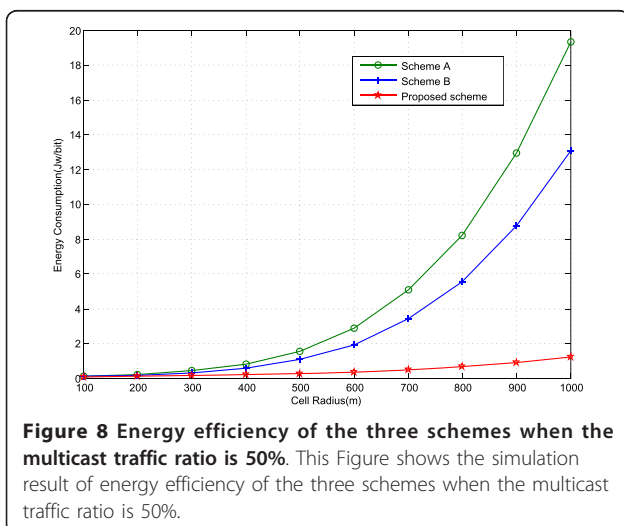
50% which means that exactly all of the non-hot traffics can be superimposed onto hot traffics. We define energy efficiency as energy consumption for unit system throughput. From the figure it can be seen that the energy consumption of the scheme A for unit system throughput is high, and the needed energy for the new scheme is much lower. Combing Figures 6 and 7, it is not difficult to understand that scheme A cannot achieve high throughput because of its TDM transmission scheme, and the scheme B cannot guarantee quality of the enhanced layer and the transmission delay for its power allocation scheme. As for the new scheme, it can guarantee QOS for several times' transmissions, and reduce the transmission delay because of its transmission rate depending on a certain top percentage user level channel condition.

Figure 5 System throughput results of the three schemes under different multicast traffic ratio. This Figure shows the system throughput results of the three schemes under different multicast traffic ratio.

Figure 6 Transmission delay of the three schemes when the multicast traffic ratio is 50%. This Figure shows the simulation result of transmission delay of the three schemes when the multicast traffic ratio is 50%.

Figure 7 Throughput of the three schemes when the multicast traffic ratio is 50%. This Figure shows the simulation result of throughput of the three schemes when the multicast traffic ratio is 50%.

6 Conclusions and future studies

This article's basic idea is from the perspective of optimizing traffic transmission through switching transmission mode from unicast to multicast. Under the target of high energy efficient traffic transmission, we proposes an energy-efficient opportunistic multicast transmission scheduling based on SC for multiple mixed unicast and multicast traffics. The proposed scheme can not only reduce transmission delay, but also effectively guarantee the quality of transmission service. In addition, the scheme provides a channel gain threshold for enhanced layer based on user selection ratio and power allocation coefficient, which can optimize system throughput dynamically. In order to prove the performance advantages of the new scheme, this article analyzes the processes and performances of other two conventional schemes particularly, through both theory derivations and simulations. The comparisons displays that our proposed scheme has advantages not only on transmission delay and system throughput, but also energy efficiency.

Energy-efficient transmission design is a key point for green wireless networks, which should be adaptive to users characteristics and traffic's QoS requirements. Current studies show that human behavior and traffic fluctuations exhibit strong regular patterns especially for social networks [21]. For wireless networks there also exist the social characteristics and many literature have studied from various perspective [22]. This article presents the preliminary study of user behavior based energy-efficient design for wireless networks. In the future studies, our study will focus on the following several aspects: (1) user/traffic-aware adaptive transmission mode selection for energy-saving; (2) energy-efficient design from user convergence's perspective; (3) optimal network configuration for Heterogeneous Networks (HetNet) considering users behavior.

Competing interests

The authors declare that they have no competing interests.

References

1. GY Li, Z Xu, C Xiong, C Yang, Energy-efficient wireless communications: tutorial, survey, and open issues. IEEE Wirel. Commun. Mag. **18**(6), 28–35 (2011)
2. Y Liu, S Xie, Y Zhang, R Yu, V Leung, Energy-Efficient Spectrum Discovery for Cognitive Radio Green Networks. *ACM/Springer Mobile Networks and Applications (MONET)*. **17**(1), 2012. accepted by special issue on "Advances in Green Mobile Networks"
3. Y Zhang, M Fujise, Energy Management in the IEEE 802.16e. MAC IEEE Commun. Lett. **10**(4), 311–313 (2006). doi:10.1109/LCOMM.2006.1613757
4. Y Zhang, Y Xiao, V Leung, Energy Management Analysis and Enhancement in IEEE 802.16e WirelessMAN. IEEE Trans. Veh. Technol. **58**(7), 3738–3752 (2009)
5. E Oh, B Krishnamachari, X Liu, Z Niu, Towards dynamic energy-efficient operation of cellular network infrastructures. IEEE Commun. Mag. **49**(6), 56–61 (2011)
6. Z Niu, Y Wu, J Gong, Z Yang, Cell zooming for Green cellular networks. IEEE Commun. Mag. **28**(11), 74–79 (2010)
7. X Zhong, M Zhao, S Zhou, X Su, J Wang, L Yang, Content aware soft real time media broadcast (CASoRT), in *IEEE Chinacom* 355–359 (2008)
8. D Kim, F Khan, Z Pi, Superposition of Broadcast and Unicast in Wireless Cellular Systems. IEEE Commun. Mag. **46**(7), 110–117 (2008)
9. PA Hoeher, T Wo, Superposition modulation: myths and facts. IEEE Commun Mag. **12**, 110–116 (2011)
10. PK Gopala, HE Gamal, Opportunistic multicasting. in *Proc IEEE 38th Asilomar Conf Signals, Systems and Computers*, Pacific Grove, USA, **1**, 845–849 (2004)
11. Y Liu, W Wang, M Peng, S Zhu, Optimized layered multicast with superposition coding in cellular systems. Wirel. Commun. Mobile Comput. (2010)
12. TM Cover, JA Thomas, in *Elements of Information Theory John Wiley and Sons Inc*, New York, **39**(1), 313–315 (1991)
13. CH Koh, YY Kim, A proportional fair scheduling for multicast services in wireless cellular networks. in *Proc IEEE VTC-Fall*, Montreal, Canada, 1–5 (2006)
14. D Tse, P Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, Cambridge, (2004)

Figure 8 Energy efficiency of the three schemes when the multicast traffic ratio is 50%. This Figure shows the simulation result of energy efficiency of the three schemes when the multicast traffic ratio is 50%.

15. EG Larsson, BR Vojcic, Cooperative Transmit Diversity based on Superposition Modulation. IEEE Commun. Lett. **9**(9), 778–780 (2005). doi:10.1109/LCOMM.2005.1506700

16. AK Goparaju, S Wei, Y Liu, On Superposition Coding based Cooperative Diversity Schemes, in *Proc Asilomar Conf Signals, Systems and Computers*, Pacific Grove, California, USA, 1046–1050 (2005)

17. SY Baek, YJ Hong, DK Sung, Adaptive transmission scheme for mixed multicast and unicast traffic in cellular systems. IEEE Trans Veh Technol. **58**(6), 2899–2908 (2009)

18. R Knopp, PA Humblet, Information capacity and power control in single-cell multiuser communications. in *Proc IEEE Int Conf Commun*, (ICC), Seattle, USA, **1**, 331–335 (1995)

19. P Viswanath, DNC Tse, R Laroia, Opportunistic beamforming using dumb antennas. IEEE Trans Inf Theory. **48**(6), 1277–1294 (2002). doi:10.1109/TIT.2002.1003822

20. S Foon, D Kim, System level performance of broadcast and unicast service overlay using superposition coding. in *Proc IEEE 18th Int Symposium on Personal, Indoor and Mobile Radio Commun*, (PIMRC07), Athens, Greece)**1-5**, 2007 ()

21. S Wasserman, K Faust, *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*, Cambridge University Press, Cambridge, (1995)

22. C Song, Z Qu, N Blumm, Albert-Laszlo Barabasi, "Limits of Predictability in Human Mobility". *Science.* **327**(5968), 1018–1021 (2010). doi:10.1126/science.1177170