

**Cross-Layer Optimisation**

**for 4G Broadband Wireless Communication Networks**

**Yang Zhao**

**Department of Electrical Engineering and Electronics**

A dissertation submitted in partial fulfilment of the requirements for the degree of BENG in *Electronic Engineering & Communications,* at the University of Liverpool

*March 2018*

Acknowledgements

This article is for my beloved parents Li Xu & Minghua Zhao and my grandmother Guiying He, who always support me whenever and wherever. Thank you for the nurture and selfless love.

I would like to express my deep gratitude to my supervisor Dr Xu Zhu, for her clear instructions, infinite patience, and immense knowledge. With her enduring guidelines, everything works out even beyond my best assumption. The door to Dr Zhu was always open when I was in trouble or proposed extra meetings. I also appreciate her suggestions and attentions for my future direction.

Besides my supervisor, I am grateful to my assessor Prof Yi Huang for his considerable encouragement and insightful comments, who always point out the places that can be improved further when I think everything is fine. The challenging questions stumped me are with answers now.

My sincere thanks also go to Dr Nan Zhou and Yanfei Qi, whose previous works helped me to build a basic system structure and understand the idea of cross-layer in detail. This project is based on their contributions.

My fellow Zimin Yang deserves appreciation as well. The discussions about the communication model that we had saved me from running into troubles more than once.

Finally, I acknowledge the IAB member Ian Macdiarmid, who managed to comprehend my project from the perspective of the real-world applications. The experience with him inspired me how to explain the project to those without relevant knowledge in communications field.

Abstract

*Focusing on the physical (PHY) layer and the medium access control (MAC) layer, this project proposed an adaptive cross-layer design for the orthogonal frequency division multiplexing (OFDM) systems with heterogeneous downlink traffic. The purpose is to realize the maximum weighted sum capacity (MWC) that balances the QoS and capacity. At the MAC layer, data scheduling is implemented. The sequence of packet transmission depends on weight, which is determined by factors as the delay, tolerance, packet size, quality of service (QoS) priority and so on. Weight is calculated for queues in modified largest weight delay first (M-LWDF) scheme, and for packets in packet dependent (PD) scheme. Once the weight is derived, it will be passed to physical layer to combine with the subcarrier & power allocation schemes to achieve flexible traffic control. It is demonstrated that with a proper packet selecting scheme, the total complexity of PD scheduling is much lower than the queue-based M-LWDF, while the system performance in terms of spectrum & energy efficiency, delay and packet drop rate is much better especially for multiuser multiservice network. Besides the conventional services, this project introduces haptic traffic to the fundamental model to ascertain the possibility of serving fundamental virtual reality (VR) and augmented reality (AR) application through existing wireless network model.*

***Index Terms****: Cross-layer design, OFDM, resource allocation, QoS, multiuser diversity, haptics.*

1. Introduction

Wireless communication has been generating fast in the past decades, but the requirements of users never settle. Apart from the substantial increase of user number, new type of services as 4K video, IoT and VR/AR are challenging the wireless network again. The proposed cross-layer design enables the adaptive resource management and active feedback control, which targets at the downlink traffic of multiuser multiservice OFDM systems.

The principle of the cross-layer design is to remove the boundaries between layers, so that the information can be shared for dynamic feedback and compensation. This project focus on the optimization of power & subcarrier allocation at the PHY layer and data scheduling at the MAC layer. As a representative sample of emerging services, haptic data is added to traditional systems to investigate the performance of different schemes for high-demand traffics. It is confirmed by relevant simulations that the optimized design can reduce the average packet delay and packet drop rate, as well as improve the spectrum & energy efficiency. Additionally, the proposed design can realize adjustable resource assignment to fulfil various requirements of different traffics. Nevertheless, the architectural costs and the unintended developing and maintaining issues cannot be neglected in real-world applications. More details will be covered in the following sections.

# Background

## 1.1.1 Open Systems Interconnection Model

In conventional Open Systems Interconnection (OSI) model, the communication system is divided into seven independent layers to standardise services and functions at layer interfaces for diverse systems [1]. This object-oriented design ensures the clear structure, simple maintenance and flexible modification, but the cost is the low information exchange efficiency between layers. Thus, a cross-layer model is proposed to provide dynamic feedback and compensation.

As the first and the lowest layer, the PHY layer focus on transmitting raw bits through channels. Information as channel response, noise level and available capacity can be obtained at this layer. Adjacent to that, the MAC layer is a sublayer of data link layer that accomplishes multiplexing and traffic control. It decides which packet or queue is served at a certain moment, thus the sequence of data transmission. The MAC layer is crucial for wireless communication since the allocation of broadcasting channels influence the user experience directly [2].

Although dissimilar networks can vary significantly in higher layers as transport layer and application layer, they comply with similar resource allocation and scheduling implementations at the PHY and MAC layer [2]. In this project, the optimisation is implemented at those two layers to provide universal support for diverse OFDM-based communication networks.

Although the OSI model is more like a theoretical reference and the networks have been dominated by the Internet protocol suite (TCP/IP) for decades, it is still beneficial to the analysis and design of typical communication systems [2]. As a consequence, the design, simulation and discussion of this project are on the basis of the OSI model.

## 1.1.2 Orthogonal Frequency Division Multiplexing

OFDM technique, as the name suggests, divides the entire channel into multiple narrow-band subchannels that orthogonal to each other. Without the need of guard band, it utilises the spectrum more efficiently compared with frequency division multiplexing (FDM). In terms of the complexity, OFDM is less intricate than time division multiplexing (TDM) and code division multiplexing (CDM), which does not demand high accuracy timing and complicated signal processing schemes.

Another advantage of OFDM is that it split the whole frequency selective channel into numerous narrow-band flat fading channels, enhancing the system stability to multipath interference. In addition, the strategy introduced guard interval (GI), in the form of cyclic prefix (CP), to avoid inter symbol interference (ISI) and inter block interference (IBI), which leads to a higher data rate. Last but not least, subcarriers enable adaptive transmission schemes, which can be applied with flexible power and diverse modulation modes. Consequently, OFDM technique is advanced in allocating resource according to varying channel situation and satisfying specific traffic requirements.

Nevertheless, the drawbacks as high sensitivity to frequency & phase offsets and high peak-to-average power ratio (PAPR) should be considered thoroughly. Dealing with interference from nearby cells can be more complex as well. The utilized system model will be discussed in the following chapter in detail.

# Previous Research

Cross-layer optimisation has been a popular topic in the past decade, partially owing to the ubiquity of OFDM. Much researches have been focusing on the interaction between the resource allocation at the PHY layer and data scheduling at the MAC layer. In [3], wireless control systems are optimised by utilising a dynamic network scheduling scheme to minimise the error in control applications. A network utility maximisation function that combines a specific utility problem with a queue management mechanism was suggested by [4] to reduce the contention and congestion in a bounded queue wireless ad-hoc network. Based on ordinary maximum signal-to-noise ratio scheduler, [5] consider an approach including dynamic control and fairness solution to maximise system throughput as well as to minimise the delay and packet loss rate by correcting the unequal spectral efficiencies caused by path loss attenuations. In [6], the balance between quality fairness and system efficiency was investigated by maximising the sum of peak signal-to-noise ratio, which provides a trade-off between conventional pure quality-fairness and maximum efficiency subcarrier allocation schemes. [7] recommends a combined design of subcarrier, bit and power allocation to improve the system capacity and decrease the packet delay, by minimising the total transmit power with fixed data rate. In [8], an integrated design based on a joint MAC-PHY layer resource allocation algorithm built a feedback loop among subcarrier allocation, power control and packet scheduling, which outperforms the tradition layer-architectural systems in power and bandwidth efficiency. Targeting at shared wireless link, [9] intended to shorten the average packet delay by a queue-based M-LWDF scheduling scheme in the MAC layer. It allocates the weight based on head-of-line (HOL) packet delay, relative data rate, and requirement for the outage probability, but the resource allocation in the PHY layer has not been covered. Sub-optimal water-filling (WF) power allocation algorithm, ascertained in [10], was combined with maximum sum capacity subcarrier distribution strategy to demonstrate its advantage with relatively low complexity. The haptic traffic was introduced to typical networks with diverse QoS designs in [11], and the result suggests that reducing delay and jitter with different QoS sets can improve user experience with haptic technology. It can further benefit distributed applications as virtual environments. As the main reference of this project, [12] proposed a packet dependent (PD) scheduling scheme, assigning different weights to varied packets in the same queue based on the delay, packet size and QoS. In such case, the system serves packets directly rather rely on queues, which enhances the network adaptability for various traffic and changing urgencies. In [13], the modified maximum weighted sum capacity was proposed based on MWC strategy, considering the weight not only for users but also for subcarriers. The principle is to maximise the summation of weighted data rates for each user and each subcarrier to increase the overall capacity.

# Objectives and Methodology

Based on the previous works, this project intends to design an adaptive cross-layer communication model between the PHY and the MAC layers for downlink multiuser multi-tasking OFDM systems with heterogeneous traffic.

By breaking the barrier between layers, the channel status information (CSI) and available capacity obtained from the PHY layer can be passed to the MAC layer for data transmission. On the other hand, the delay and weight information can be returned to the physical layer for adaptable subcarrier and power assignment. Compared with the conventional architecture with clear layers, this approach is more flexible in meeting the requirements of different services, more efficient in power and spectrum utilization, and more advanced in network performance as delay and packet loss rate.

To investigate the performance of the proposed cross-layer design for multiservice system with strict requirements, as well to examine the feasibility of utilizing the existing wireless network to support applications based on emerging services, haptic traffic was introduced to the model. Haptics refers to interactions involving touch. Through haptic technique, the applications are able to simulate the virtual objects that can be touched. It is a fundamental technology for virtual environment realization, with the application in VR and AR.

The entire project can be divided into five main parts. Above all, the wireless channel models are to be investigated. Then, the theoretical model of OFDM system will be explained, and the corresponding simulation results are to be interpreted. Note that only downlink data is considered in this project, which means all the available bandwidth and subcarriers are for downlink traffic. Subsequently, data scheduling will be carried out at the MAC layer with PD and queue-based M-LWDF respectively. Data transmission order and the weight set would be determined by the end of this stage. Next, the subcarrier allocation will be executed at the PHY layer to maximise the sum capacity and weighted capacity respectively. At the same layer, the power will be distributed according to WF and weighted water-filling (WWF) algorithms. The proper selection of the packets for the weight design can reduce the complexity of the model. MATLAB is the simulation tool employed in this project for the performance and complexity analysis.

# Industrial Relevance

The fourth generation of broadband cellular network technology (4G) has developed fast in the past decade. Compared with conventional technologies, 4G provides larger throughput, lower delay, more QoS strategies and enhanced safety. As required by [14], the user data rate should reach around 1 Gbps in low speed and 100 Mbps in high speed movement. The high data rate exacerbates fading effects, and OFDM is proposed as a proper multiplexing scheme to avoid frequency-selective fading. Therefore, the data rate can be guaranteed and the complexity in signal recovering can be reduced significantly.

Optimization in this project is realized on the basis of OFDM system. Although the title suggests the cross-layer design targets at 4G broadband network, it can be utilized in any OFDM-based broadband wireless system to provide more flexibility and enhancement on delay, packet drop rate and power & spectrum efficiencies.

With the advancement of fast Fourier transform (FFT), the complexity of OFDM implementation has been decreased significantly. The prosperity of OFDM has been witnessed in numerous famous communication systems, as wireless local area network (WLAN), asymmetric digital subscriber line (ADSL), high-definition television (HDTV) and power line communication [15]. Furthermore, an advanced OFDM scheme named OFDM-IM proposed by [16] is recommended by [17] [18] and [19] as a preferred multiple access technique in the upcoming 5G. Due to the potential difference in OFDM realization, the spectrum division thus subcarrier design may need reconsidering. Additionally, the packet size distribution should be adjusted according to the application specifications.

More flexible QoS strategy introduced by PD scheduling scheme can improve user experience as well. The simulation results demonstrate that with the proposed cross-layer design, those traffic with strict demands can be assigned with a higher priority level thus transmitted in precedence. As a consequence, the quality of relevant applications can be guaranteed. A typical instance is that users in traditional networks may experience large voice call delay and packet loss when downloading files, but PD scheduling in this cross-layer design can mitigate the circumstance substantially.

Apart from the fundamental applications investigated in the simulation, it is possible that with the further support from the applications based on this cross-layer model, users can adjust the network characteristics according to their dissimilar demands. That is to say, the developers can create various QoS priority schemes for different purpose. For those desire low delay transmission, the services rely on delay as haptic and voice should be assigned with a higher priority index. Therefore, the corresponding packets are with larger weights thus served in precedence and have lower delay. While for those want to maximize the bandwidth, a balanced design should be set up as in the following simulation. Versed users might even customize the priority levels manually for the best experience.

As can be demonstrated by the simulation result, the proposed design with PD scheduling scheme can maintain high bandwidth efficiency, even for a large number of users. This advantage makes it especially suitable for regions with numerous users as stadium, park and shopping center. Additionally, the dynamic subcarrier allocation can enhance the subchannel utilization ratio, which is significant for remote area that requires wireless signal coverage. As a consequence, the amount of base stations and repeaters can be reduced to cut down the network cost.

One of the main problems that require further attention is the fairness issue. It will be shown later that when the available resource is not enough for ideal communication, the MWSC scheme cannot guarantee the data rate of users to be in similar range. When the user number is large, it is possible that some users are not assigned with any subcarrier in a certain timeslot, whose data rate should be zero while others are possibly with stable connections.

Although the proposed model improves throughput, delay and packet drop rate, the advanced performance comes with the architectural costs. The majority of optimization design based on a single layer should be adapted before they are reused in this system. Moreover, if realized improperly, the cross-layer design can lead to spaghetti structure which means parts are mixed together. In such case, the complexity of the network can be increased and there can be unintended developing and maintaining issues. In the worst case, the combined layer may need to be redesigned rather than replacing a single layer.

1. Channel Models

Channels are the medium that convey information in terms of signals from one terminal to another. In wireless communication, impulse, intermodulation and thermal noise affect the transmission by adding meaningless components to the modulated signal. In addition, phenomenon as relative movement, crosstalk, attenuation and delay distortion increase the complexity of recovering the original message [20]. It suggests that the choose, design and simulation of the channel model is the first step of wireless communication system emulation.

# Additive White Gaussian Noise Channel

As a typical thermal noise, additive white gaussian noise (AWGN) refers to the noise that can be modeled as a zero mean Gaussian wide sense stationary (WSS) random process, with all frequency components appear with equal power.

Excluding nonlinearity, interference and fading, AWGN is regarded as the most basic channel model for wireless system design and analysis. Explaining in detail, the description of WSS requires a process to be with constant mean and a one-dimensional time-difference-dependent-only autocorrelation function. For a WSS process, power spectral density (PSD) is defined as the Fourier transform of the autocorrelation function:



Based on that, thermal noise can be modeled as a Gaussian WSS random process with the mean to be zero. For a random variable *x* with Gaussian distribution, the probability density function (PDF) is:



with *μ* being mean and *σ* being variance.

The term ‘white’ implies the channel has equal noise power for frequency from zero to infinity. By definition, AWGN is with a flat PSD:

*(W/Hz)*

while the term *N0* is the noise level determined by:

*(W/Hz)*

with *K* being Boltzmann’s constant and *T* being the temperature in Kelvins.

But in baseband communication, only a certain bandwidth is employed for data transmission, and the transmitted signal should be passed through a low-pass filter. As a consequence, the resultant noise is supposed to be bandlimited as well. Figure illustrates the details.

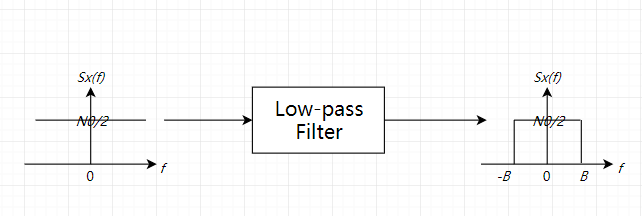


Figure White noise after low-pass filter

Moreover, ‘additive’ indicates that the received signal *y* can be obtained by simply adding AWGN *n* to the transmitted signal *x*. The relationship can be described as:



In the end, the received signal is to be processed by the decision device to recover the original message in bits. The main idea in signal detection is to compare the received signal with the threshold and utilize the result to estimate the possible input. Before the recovering, the receiver should know the expressions corresponding to bit 0 and 1. First, the impulse response of the matched filter of the optimum receiver can be formulated as:



where the *(T-t)* terms mean to sample the received signal at the moment that *t=T*. Then, the threshold can be calculated as half of the bit energy difference:



Finally, the received signal is to be passed through the matched filter and compared with the threshold. If the result is greater than the threshold, the system suggests the original bit is more likely to be 1 than 0. In such case, it is acceptable to regard the original bit as 1. Otherwise, it is reasonable to decide the initial message as 0.

Nevertheless, there can be some errors with the estimation, especially when the signal-to-noise ratio (SNR) is relatively low[[1]](#footnote-1). The rate that bit 0 transmitted incorrectly regarded as 1 plus bit 1 transmitted wrongly decided as 0 is termed as bit error rate (BER). For binary transmissions:



The term *d* is the distance of signal 0 and 1 in the signal space, while the Q-function indicates the probability that the value of a Gaussian random variable greater than the argument. When the channel noise level is high, the transmitted signal is more susceptible to be influenced, and the recovered signal can shift in a wide range. As a result, the BER is supposed to be high, as the formula suggests.

Define , for binary phase shift keying (BPSK) in AWGN, the BER can be derived as follows:



If the SNR is high (), the Q-function can be approximated as:



And the BER:



It can be concluded that in AWGN channel, BER decreases exponentially with the increase of SNR.

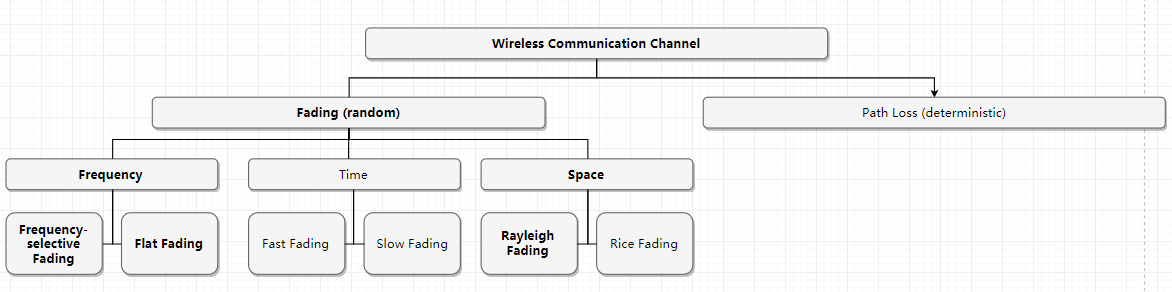
According to Shannon’s theorem, the channel capacity or the maximum data rate for ideal AWGN channel without error is:

*bps*

An interesting fact is that due to the requirement that the data rate should be less than the channel capacity, the minimum SNR can be derived as -1.6 dB for reliable communication.

# Fading Channel

In wireless communication, fading indicates the random amplitude variation of the received signal owing to the characteristics of the channel. Between the transmitter and the receiver, there can be multiple paths that the signal can traverse, with various effects as reflection (obstacle dimension *d* > wavelength *λ*), diffraction (wave propagates around edges), and scattering (numerous obstacles with *d* < *λ*) [20]. Those signals transmitted through multiple paths are with attenuation, delay and phase shift in different degrees, overlapping constructively or destructively in the receiver end.



Channel categories

It is illustrated by Figure that random fading and deterministic path loss constitute wireless communication channel, while fading can be categorised by time, space and frequency. This project is based on single-input single-output (SISO) systems with multipath small-scale flat Rayleigh fading.

## Flat Fading and Frequency-selective Fading

In terms of frequency, the fading can be categorised as flat fading and frequency-selective fading. In flat fading, all frequency components of the signal undergo fading in similar degree. In comparison, frequency-selective fading denotes that different frequency components of the message can suffer from uncorrelated fading, leading to intersymbol interference (ISI) and signal distortion.

The only criterion used for comparison is the coherence bandwidth. If the frequencies separation is less than the coherence bandwidth, the fading effects on them are supposed to be highly correlated. It is estimated that for frequency correlation higher than 0.5, the coherence bandwidth should be:

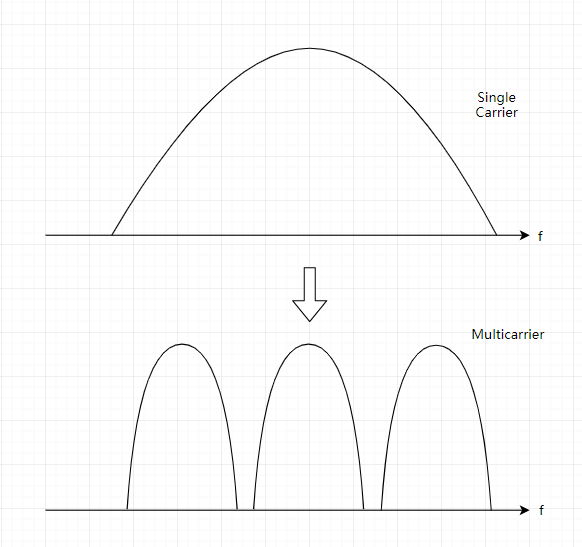


The term  is the root-mean-square (RMS) delay spread, measuring the time dispersion of multipath channels, or multipath effects. It can be expressed as:



Frequency-selective fading happens when the signal bandwidth *BS* is greater than the coherence bandwidth *BC*. In such case, the available channel cannot transmit the message without uncorrelated fading, which means the fading depends on the frequencies. On condition that, there can be some correlation among fading experienced by different frequency components. In other word, the fading affects the frequencies in similar way, so it is called flat fading.

In most cases, the wireless communication is based on flat fading channel, owing to the message undergoes little or no distortion in this model and therefore simple to recover. However, as a characteristic, the coherence bandwidth is constant for a certain channel, which means the maximum signal bandwidth should be fixed as well. One solution is utilising multicarrier to divide the broadband frequency-selective fading channel into multiple narrowband flat fading segments as Figure illustrates, then perform frequency domain equalisation (FDE).



Single carrier and multicarrier

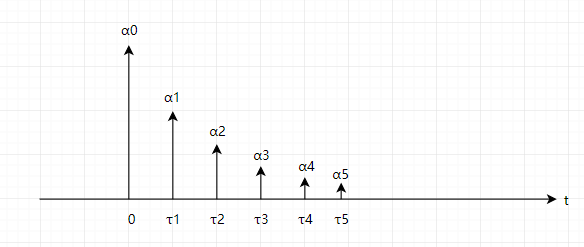
## Rayleigh Fading

As a statistical model, it is assumed in Rayleigh fading channel that the real and imaginary parts of channel response follow zero-mean Gaussian distribution, and the envelope can be expressed by Rayleigh distribution:



Ω indicates the mean power of the channel impulse response.

Rayleigh fading channel is able to model the environment with numerous scattering, which means there should be enough obstacles with dimension less than the wavelength. It also suggests there is not supposed to be line-of-sight (LoS). As a consequence, it is reasonable to utilize this model to simulate the wireless channels in urban area. Figure describes the multipath Rayleigh fading channel model. In this project, the number of path is set to 6.



Path impulse response

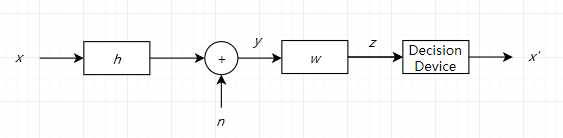
Using α to represent path gain and τ to denote delay, the channel impulse response (CIR) can be expressed as:



The relationship between transmitted signal *x*, received signal *y*, CIR *h* and noise *n* is defined as:



After obtaining the signal, the receiver passes it through a filter for signal detection, then to the decision device, as Figure suggests.



Signal recovering

The input *z* to decision device is:



To recover the original message, zero forcing (ZF) scheme set the gain of the signal detection filter to *1/h*. As a result, the input to the decision device is:



It can be seen from the equation that the noise amplitude is reduced by h times. Furthermore, define , the BER for BPSK can be expressed as:



While by definition, the BER is determined as:



Set γ0 as the mean value of SNR per bit:



In Rayleigh fading channel, the PDF can be written as:



As a consequence, the BER can be further derived as:



When the SNR is high (), an approximation can be taken as:



The result indicates that for Rayleigh fading channel, the BER decreases linearly with the drop of *1/γ0.*

On condition that the channel gain *h* is known, the channel capacity can be expressed as:



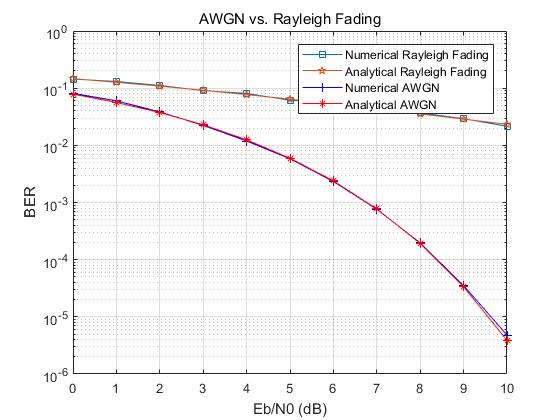
# Channel Simulation

AWGN and Rayleigh fading channels have been simulated based on the theories above. The fundamental characteristics of the channel are shown in Table.

Parameters in channel simulation

|  |  |
| --- | --- |
| Coding scheme | Bipolar non-return-to-zero (NRZ) |
| Range of SNR (in dB) | 0-10 |
| Number of subcarriers | 64 |
| Point of FFT | 64 |
| Number of paths | 6 |
| Number of blocks | 1,000 |
| Number of bits per block | 10,000 |

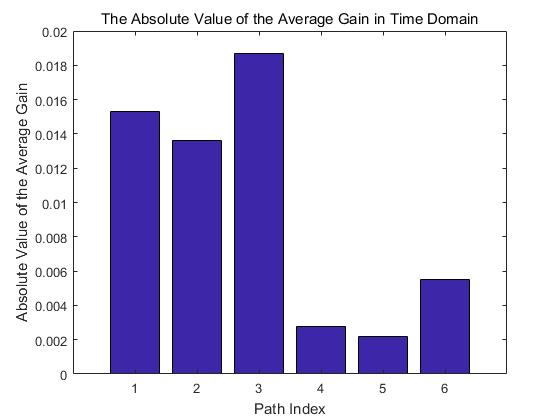
Figure intends to convey the analytical and numerical BER performance for AWGN and Rayleigh fading channels. The precise formula was used for the theoretical BER calculation of Rayleigh fading, which is also proper for low SNR cases. It can be spotted that the simulated channels match the theory perfectly, indicating the success of modeling. Furthermore, these results demonstrate convincingly that with the increasement of SNR, the BER declines exponentially in AWGN channel, while the BER is inversely proportional to SNR in Rayleigh fading channel. The result suggests that with the same SNR background, the AWGN channel has lower BER therefore better performance. Also, to achieve the same BER, Rayleigh fading channels require much higher SNR compared with AWGN channels. Nevertheless, as mentioned above, the AWGN model is the most basic one, and in most wireless communication systems the Rayleigh pattern dominates.



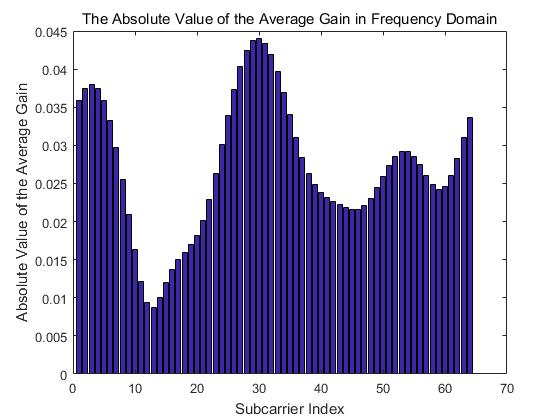
SNR vs. BER for channels

It is a consensus in wireless communication field that the maximum acceptable BER is 10-5, which corresponds to around 9.5 dB SNR. But for Rayleigh fading channel, it is estimated to meet the requirement, the SNR should be around 50 dB.

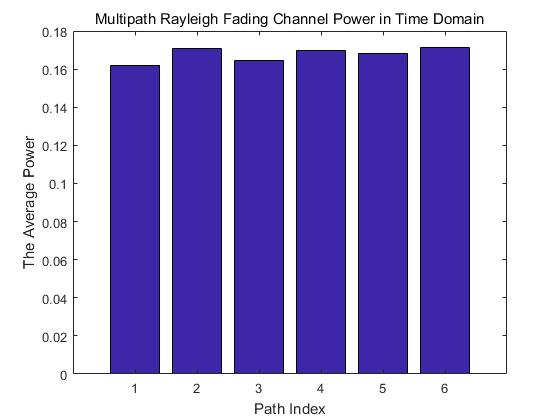
Figures a-d show the simulation results of the multipath Rayleigh fading channel.

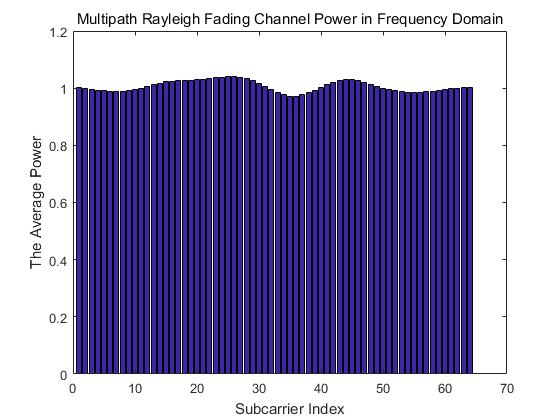


Average gain and paths



Average gain and subcarriers



Power and paths

Power and subcarriers

In a & b, it can be seen that absolute value of the average gains for paths and subcarriers are random and relatively low. With gains approaching to 0, the result suggests that random subcarrier allocation should be inefficient for users, and further schemes are demanded to utilize the subcarrier resource rationally.

In terms of power, Figures c & d illustrate the power distribution situation. In time domain, the six paths share the unit power equally, each consuming about 0.17 W. The outcome in frequency domain indicates that the average power of every subcarrier is around unit. Although the consequences are as expected, there can be some optimization algorithms that allocate more power to the subcarriers with relatively high gain to improve the system performance. Those resource allocation schemes will be compared, employed and analysed later.

1. System Model and Design

# OSI Model

The basic structure of OSI model is illustrated by Figure. Proposed in 1983 [21] and revised in 1995 [22], this model was the first attempt of layer protocols and services standardization, which describes how to interconnect the open systems that communicate with each other. Lower layer provides services to the adjacent upper layer through the inter-layer interface, and the peer entities communicates based on protocols.

OSI model structure

|  |  |
| --- | --- |
| Application Layer | |
| Presentation Layer | |
| Session Layer | |
| Transport Layer | |
| Network Layer | |
| Data Link Layer (DLL) | Logical Link Control (LLC) |
| Medium Access Control (MAC) |
| Physical Layer (PHY) | |

It can be seen that the communication system is divided into seven typical layers. Declaring the method that upper layers should access lower ones, the interfaces stipulate the parameters to be passed between adjacent layers. Services define the role of layers, while the protocols specify the rules that the peer entities within a layer should obey in data exchange [2]. The object-oriented strategy of OSI model ensures clear structure and easy modification, but the fixed pattern brings much restrictions and influences on the system performance. As a result, cross-layer optimization is proposed to increase the spectrum and power efficiency, as well as to provide dynamic feedback and compensation to improve system performance.

The PHY layer focus on transmitting raw bits through the channel. Relevant issues as CSI obtain, BER control, and power & spectrum allocation are carried out in this layer. Compared with wired channels, the BER of wireless networks are higher due to unstable time-variant factors as shadowing and multipath fading. In such case, to enlarge the throughput and shorten the delay, the CSI is supposed to be monitored for adjustment, which can be obtained through this layer.

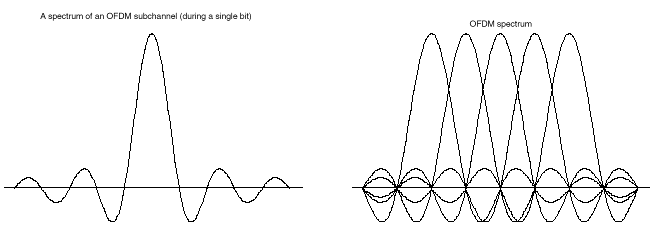
As for the MAC layer, it is a sublayer of the data link layer that deals with data scheduling and multiplexing. QoS scheme that provide different priority to varied traffics to utilize limited resources to guarantee the service quality is controlled in this layer. Typical factors that determine QoS priority are as throughput, delay tolerance, jitter, and maximum packet loss rate. With proper resource management schemes for different services, the efficiency of broadband communication system can be increased. Later in this article the QoS design for haptic data will be investigated.

The proposed optimization is realized between those two layers. In the MAC layer, the order of data transmission is determined on the basis of queues or packets. In the beginning of each time slot, the following processes are executed. First, the weight of queues or packets in buffer is calculated, relying on parameters as packet size, delay, delay tolerance, QoS priority, etc. It is those weight that directly determine the packet serving sequence. Second, the weight information is passed to the PHY layer and combined with the subcarrier and power allocation algorithms, and the maximum available data rate with weight influence can be derived. Third, it is assumed a packet corresponding to largest weight is sent, then the possible data rate is calculated and compared with the capacity. On condition that the new rate is achievable, the packet will be sent with related weight removed, buffer renewed, and data rate updated. Finally, the last step will be looped until the circumstance is not valid. According to this scheme, the packet with larger size, longer waiting time, shorter delay tolerance or higher priority level will be assigned with a larger weight, thus served earlier. With the coordinating of the two layers, it is believed that the quality of the network can be improved, and the specific requirements of various services traffic can be fulfilled.

# OFDM

In this project, the optimisation is based on the downlink of a frequency division duplexing (FDD) OFDM. FDD utilises two individual frequency bands to transmit and receive data respectively, and a guard band between uplink and downlink frequencies to avoid interference [20]. Compared with time division duplexing (TDD) where the information is transmitted and received in the same frequency band, the system design is simpler with shorter delay and less power consumption, but the cost is the guard bands as well as same amount of spectrum for message transmitting and receiving. FDD can be bandwidth efficient with symmetric traffic, but in most cases, users have more downlink data than uplink to be served. The phenomenon challenges the downlink sub-band with the demand of high spectrum efficiency, as a consequence, the optimisation is target at downlink transmission only.

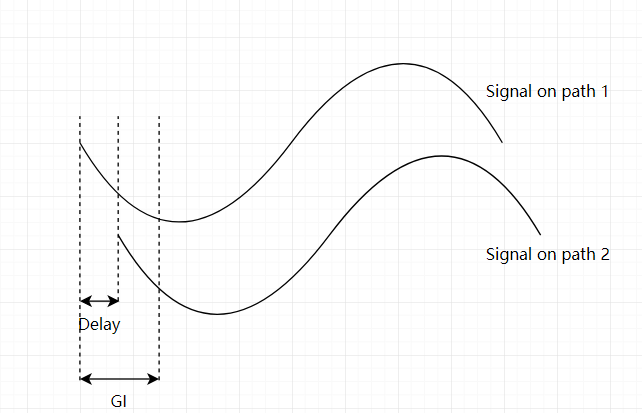
As has been discussed in the introduction part, the main advantage of OFDM are the adaptability to various channel states without the requirement of complicated equalisation. Before the transmission, the data is divided into numerous parallel data streams corresponding to subcarriers, which can be modulated with conventional modulation schemes. In comparison with single-carrier transmission, the symbol rate of carriers can be relatively low, but the overall data rate remains, avoiding frequency-selective fading and ISI. As Figure illustrates [23], the subcarriers in OFDM are in shape of sinc function.



OFDM subcarrier waveform [23]

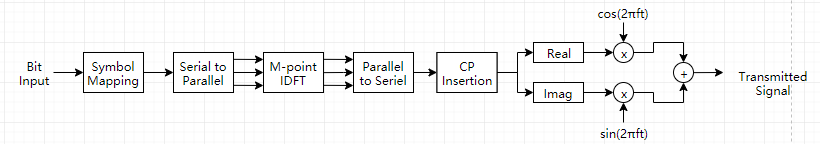
It can be seen that a single subcarrier has a peak with two symmetric tails. When a set of subcarriers share the available bandwidth, it seems that they are overlapping with the neighbors, but the digital signal modulations only happen at the frequencies corresponding to the peak, where the effect of the others can be ignored. As a consequence, there are no inter-carrier interference (ICI) and the spectrum fits more subchannels than FDM without the need of guard band, which increases the bandwidth efficiency.

To prevent IBI, in time domain, GI is required to be inserted between OFDM blocks, which is mainly in form of CP. Due to the multipath effect, there can be some delay among the signals transmitted through different paths, which may influence the next signal to be received. Therefore, the length of CP must be larger than the maximum possible delay among signal copies, as Figure indicates. The content of CP is the repetition of the last several symbols of the previous block to maintain orthogonality. However, the energy and spectrum efficiency will be reduced as well, so the length of CP is supposed to be determined to balance the trade-off.



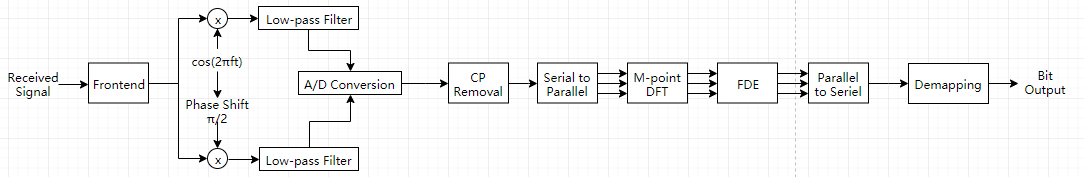
GI and signals received from different paths

The diagrams of OFDM transmitter and receiver with quadrature amplitude modulation (QAM) are shown by figures below.



OFDM transmitter

The first step of OFDM data transmission is to map the bit stream into waveform. Then, the signal will be converted to parallel parts which can be further transmitted and assigned to subcarriers. With the total data amount remains, the symbol period will be extended *m* times where *m* is the number of subcarriers. As a consequence, the frequency is only *1/m* of the original signal, experiencing flat fading, and the ISI phenomenon can be mitigated substantially. After that, the signal should be converted back into time domain for time-based CP insertion. Next, the compound signal is supposed to be separated into pure real and imaginary components for BPSK modulation. Finally, the modulated signals are to be combined together for transmission.

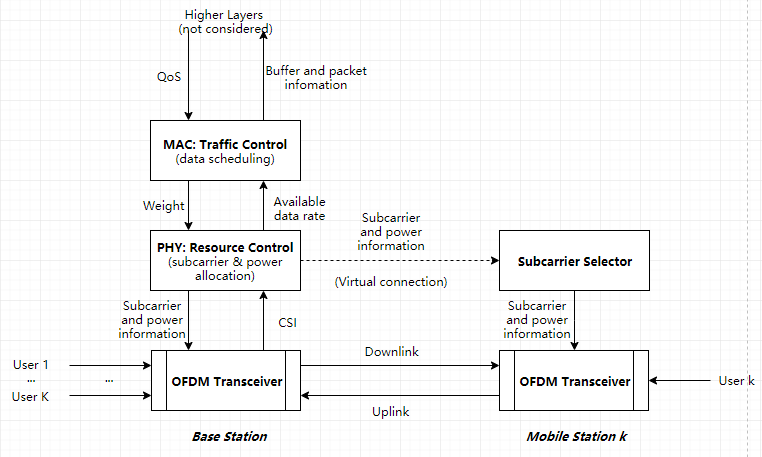


OFDM receiver

Message recovery is based on similar strategy. Divided by the front end, the real and imaginary parts are to be demodulated by multiplying with the carriers then passed through low-pass filters. After combining and analog-to-digital (A/D) conversion, the CP will be removed from the signal in time domain. Subsequently, the serial signal will be parallelized then transformed to frequency domain. FDE strategies as ZF can be carried out on single subcarriers with low complexity. Afterwards, the signal can be serialized then demapped into raw bits.

# System Model

The system structure is illustrated by Figure.



System structure of cross-layer design

Detailed explanation about the model, mechanism and process have been covered in section 3.1 and 3.2. As can be seen, it is a combination of cross-layer hierarchy and OFDM module.

One of the main advantages of this design is the spectrum and power efficiencies. From the perspective of cross-layer model, the system is to guarantee MWC for given bandwidth and power through dynamic feedback and compensation. Additionally, the weight design based on packet rather than queues can fulfil different requirements of various services flexibly, and more type of traffic can be introduced to the existing system easily, as has been implemented for haptic data. It is also demonstrated that the system performance in terms of throughput, delay and packet drop rate has been optimized. Moreover, OFDM that enables adaptive transmission tactics also ensures less IBI, ICI and ISI, which is bandwidth efficient and less complex than conventional schemes.

In contrast, the price cannot be ignored. The enhanced performance is with the cost of architectural costs. If handled improperly, it may lead to spaghetti structure, which ruins the stability of the whole network. Besides, there can be unexpected developing and maintaining issues that never occur in traditional design. In such case, the joint layer might need to be redesigned rather than a single layer. Furthermore, it is probable that other optimization strategies based on a certain layer cannot be reused on this design. More analysis and discussion based on the simulation result will be covered later.

The simulation is fulfilled on a quasi-static fading channel that on time slot basis. It is assumed that the CSI remains unchanged in every slot, and the packets should be sent at the end of slots. The setting also suggests that the weight and buffer can only be updated when slots finish. As a consequence, the achievable data rate is supposed to be constant in a certain slot. For total bandwidth *B* shared by *N* orthogonal subcarriers to serve *K* users, assuming  represents the subcarrier set allocated to user *k*  and  is the power of user *k* on subcarrier *n *, the maximum available data rate of user *k* can be derived through Shannon Theorem:



in which



is the channel-to-noise power ratio (CNR) of user *k* on subcarrier *n*.

As mentioned above, the achievable data rate is calculated in PHY layer with the weight and subcarrier & power allocation result then passed to MAC layer for data scheduling.

1. Resource Allocation at the PHY Layer

The objective of resource allocation is to utilise the limited resources effectively to enhance the quality of the wireless network. Two kinds of resources are considered in this project, namely subcarrier and power. As the most basic unit of OFDM system, subcarriers, or subchannels, can be assigned to different users for signal transmission. The purpose of subcarrier allocation is to achieve the highest bandwidth efficiency by sending as much information as possible. A subchannel can only serve one user at a time, and the CIRs of different subcarriers on dissimilar users are varied, which changes with time and space. In other words, a subchannel experiencing severe fading for this user may have the largest gain on that user compared with other subcarriers. As a consequence, to increase system capacity, it is necessary that the system utilise a proper scheme to distribute subchannels to users to maximize CNR. On the other hand, compared with allocating the same amount of power to subcarriers, the network performance can be improved with adjustable power distribution strategies. As the formula indicates, to increase the maximum data rate, those subcarriers with higher gains are supposed to receive more power, and the abandoned ones should not be considered. In the following sections, several subcarrier and power distribution schemes will be compared and combined, then a proposed one will be utilized and further blended with scheduling schemes at the MAC layer.

Please be aware that the algorithms below are based on bandwidth *B* divided into *N* independent subcarriers to serve *K* users. User index is indicated by *k*  and subcarrier index by *n* **. The subchannel set of user *k* is denoted by  subject to  and  to ensure a subcarrier serves only one user at a time and the subcarrier sets should be available. The total power transmitted from the base station is marked as *PT*, and power on subcarrier *n* for user *k* denoted by *pk,n* subject to  and  to guarantee the power on subcarriers is valid and the total subchannel power is less than or equal to the transmitted power.

# Subcarrier Allocation Schemes

## Random Subcarrier Allocation

As the most fundamental plan, random subcarrier allocation means assigning subcarriers to users randomly, with the number of subchannels fixed for every user. In such case, the network quality of individual users can vary in great degree, and there are low possibilities are that some subcarriers are unavailable with gains to be zero. Nevertheless, low as it can be, the achievable data rate for a certain user is not supposed to be zero in most times, since the number of subcarriers should be more than that of users in most cases. The formulation for subchannel set  assigned to user *k* can be expressed as:



where *n* is the subchannel and



is the size of subcarrier set.

Since this subcarrier allocation is merely determined by the number of subcarriers *N* and the user amount *K*, it does not require CSI-based feedback and control, which reduces the network complexity. Another advantage is that this approach ensures relative fairness with fixed number of subchannels assigned to users. Nevertheless, as discussed above, the cost is the capacity, which means the throughput and delay will be influenced.

## Maximum Capacity (MC) Subcarrier Allocation

Targeting at maximizing the system capacity, the MC algorithm [13] checks the CSI for all subchannels before allocating them to users. On account of the time-variant characteristic, the information should be updated frequently for better result. On the other hand, the feedback of CSI and reallocation of subchannels require a little time to be executed. As a consequence, this scheme must be carried out with proper time interval.

In MC subcarrier allocation, the first step is to check the CSI of a certain subcarrier for all users. It is realized by generating path CIRs for Rayleigh fading channel, then apply FFT with points equal to number of subcarriers on them. Next, the corresponding gains are to be compared and this subchannel will be assigned to the user with the largest gain on it. The last step is to be repeated for other subcarriers until all of them are occupied. In the beginning of every time slot, the loop above is to be accomplished with updated CSI and user information. The capacity function to be maximized is expressed as:



where *Rk* is the achievable data rate for user *k* determined by Formula. The constraints mentioned in the section 4.1 apply.

As the name suggests, MC allocation can maximise the system capacity with feedback from CSI and enlarge the throughput and shorten the delay. This strategy is suitable when there are enough data in the system to be sent, applying with either numerous users or high network traffic. However, the QoS demands are not considered in this design, which means the majority of resource can be given to low priority traffic-dominating services as downloading and streaming. Therefore, urgent packets as voice may experience longer delay and the network quality will be affected.

## Maximum Weighted Sum Capacity (MWC) Subcarrier Allocation

Contemporary communication systems should support traffic shaping to meet different requirements of various services. Therefore, proper QoS design must be reflected in the network model to improve user experience. As a consequence, the previous algorithms focusing on fairness and capacity should be adjusted to fit the actual demand. In MWC approach [13], not only CSI but also weight of users need to be updated for every time slot, which can be obtained with diverse scheduling schemes at the MAC layer. The utility function summing the weighted sum capacity of users is to be maximized as:



with *Wk* indicating the weight of user k. The constraints mentioned in the section 4.1 apply.

In contrast with MC allocation, MWC assigns subcarriers to the user with largest weighted data rate first. After obtaining enough resource for transmission, the user and corresponding subchannels will be removed from the waiting list and the steps are to be repeated until all users are satisfied. As for the algorithm implementation, rather than founding the user with best CIR on a certain subcarrier, the subchannel gains are multiplied with weight of user and to be compared then for allocation.

One of the main advantages of MWC subcarrier allocation is that it can be blended with various QoS schemes to establish a flexible network. By adjusting the weight algorithm, the traffic of different service can be controlled according to the requirements. This strategy is particularly suitable for networks with increasing type of services. Later in this article, the influence of adding haptic data with varied weight design schemes to stable communication system will be investigated.

On the contrary, there are some problems deserve attention as well. As a public concern, fairness cannot be guaranteed by this algorithm. Due to the influence of time and space, it is possible that some users are in fast and stable connection while the others are experiencing poor network. That situation can occur when the user amount is large, the subchannel gain is low, or the channel status is changing fast (e.g. poor weather or fast relative motion). In such cases, the users can be served by subcarriers with small gain or no subcarrier at all, leading to low data rate. Another defect is compared MC, the flexibility of QoS comes with the cost of some capacity. The affect will be examined and discussed in the following section.

# Simulation Result of Subcarrier Allocation

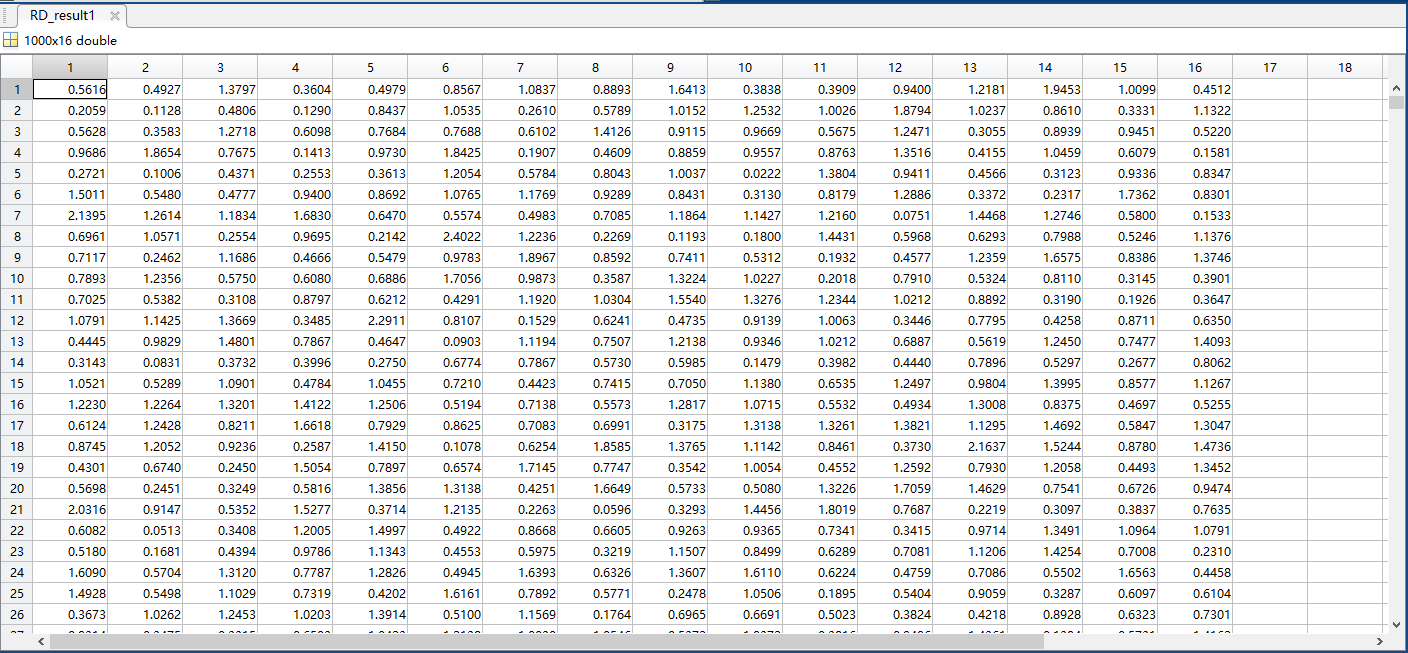
The fundamental simulation in this part is merely to compare the performance between different subcarrier strategies. Results are given in form of two matrices, with rows indicating subcarriers and columns denoting channels. For every subchannel of all channels, the first matrix holds subcarrier gains and the second one stores the user index that the subchannel is assigned to. Simulation is based on the parameters shown in Table.

Parameters of subcarrier allocation simulations

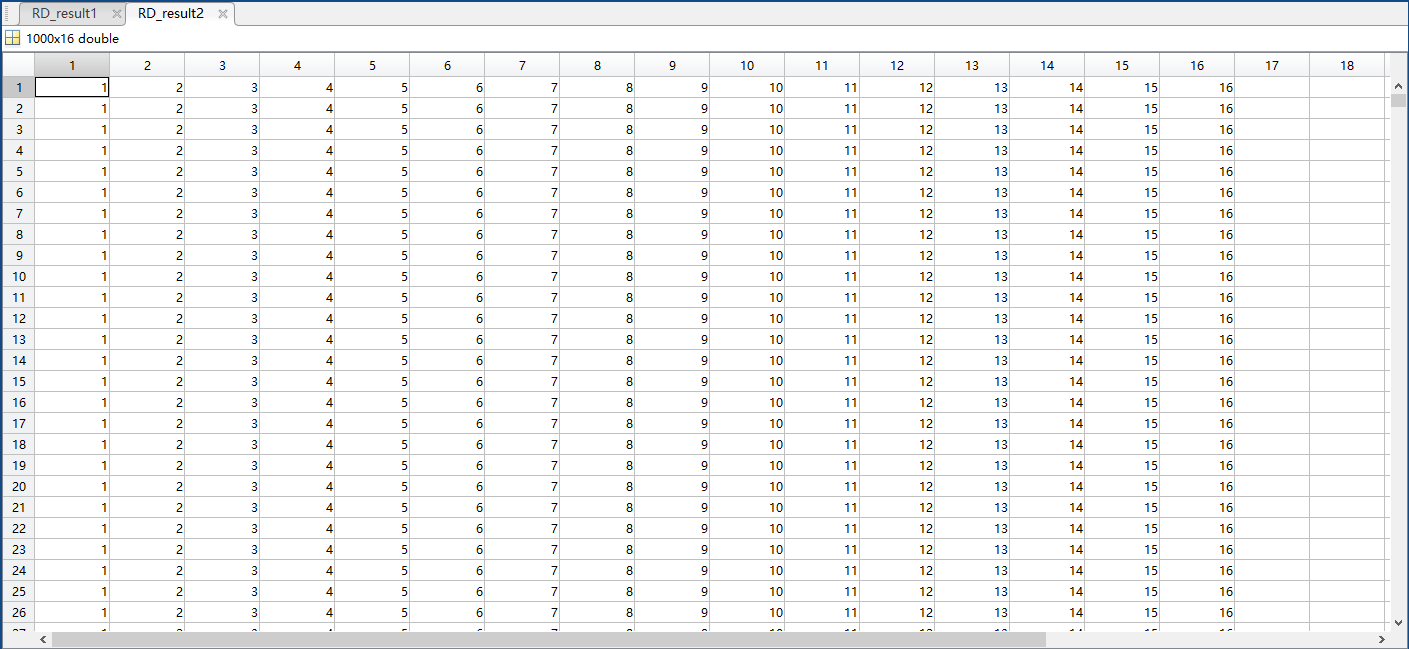
|  |  |
| --- | --- |
| SNR (dB) | 10 |
| Number of users | 16 |
| Number of subcarriers | 16 |
| Number of paths | 6 |
| Number of channels | 1000 |

## Random Subcarrier Allocation

Figure reveals the gains on each subcarrier, while Figure indicates the subcarrier allocation result. The rows indicate channels.



Subcarriers (column) and gains with random allocation

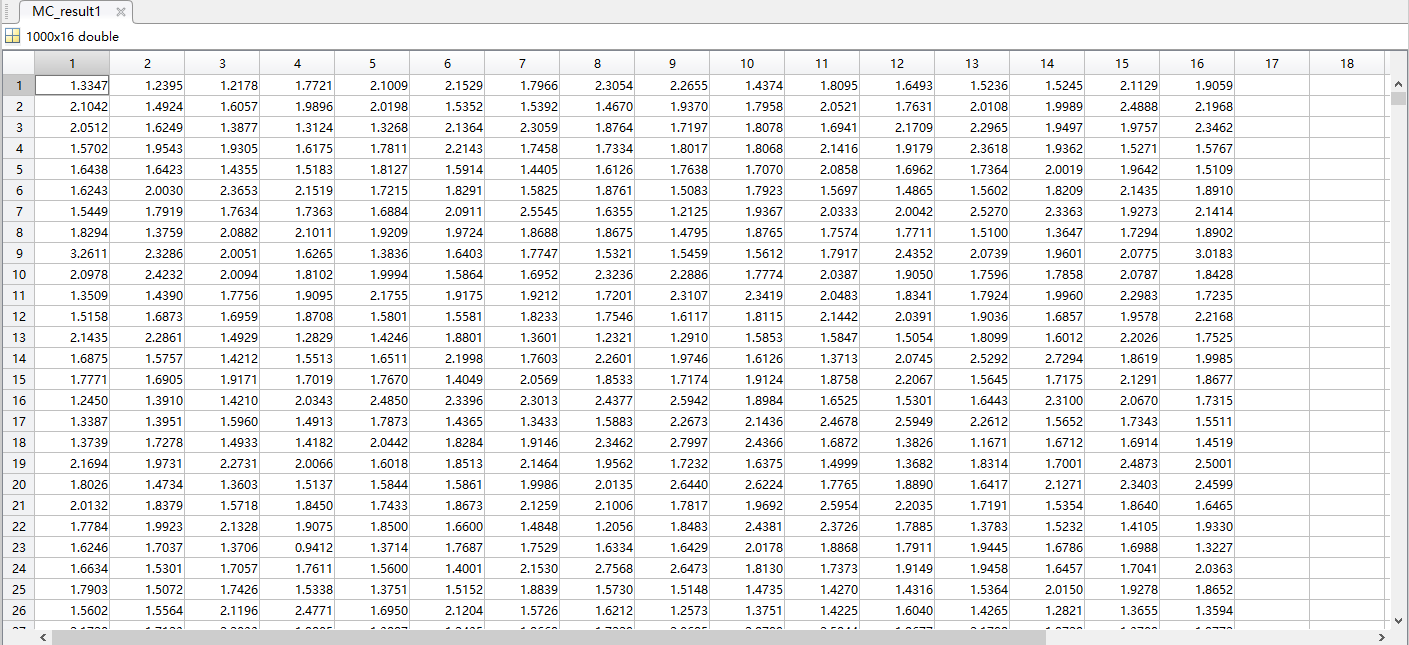


Subcarriers (column) and users with random allocation

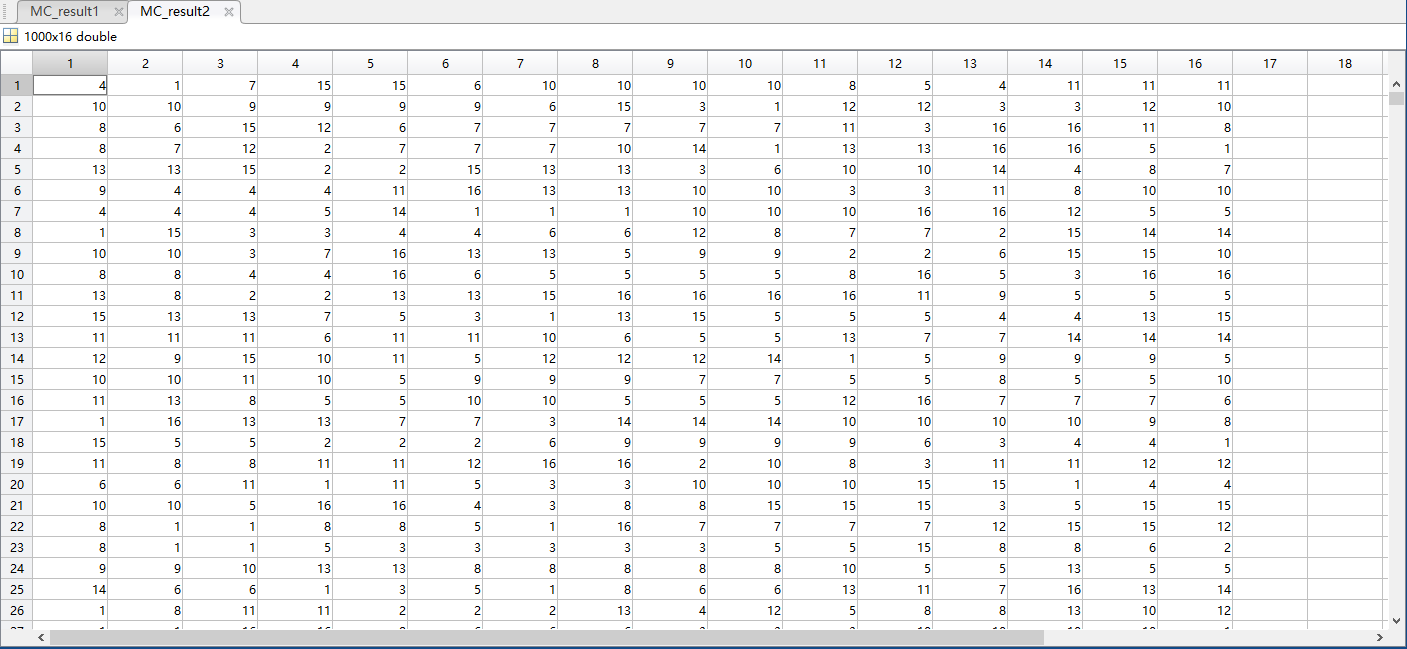
Every subcarrier has different gains on different users, but it can only serve one user at a time and the gain on corresponding user is shown. It can be concluded that the gains for the majority subchannels are less than 1 with the average value to be 0.8838, and the subcarrier allocated to users are fixed for all channels.

## MC Subcarrier Allocation

Results are shown by Figure a and b.



Subcarriers (column) and gains with MC allocation

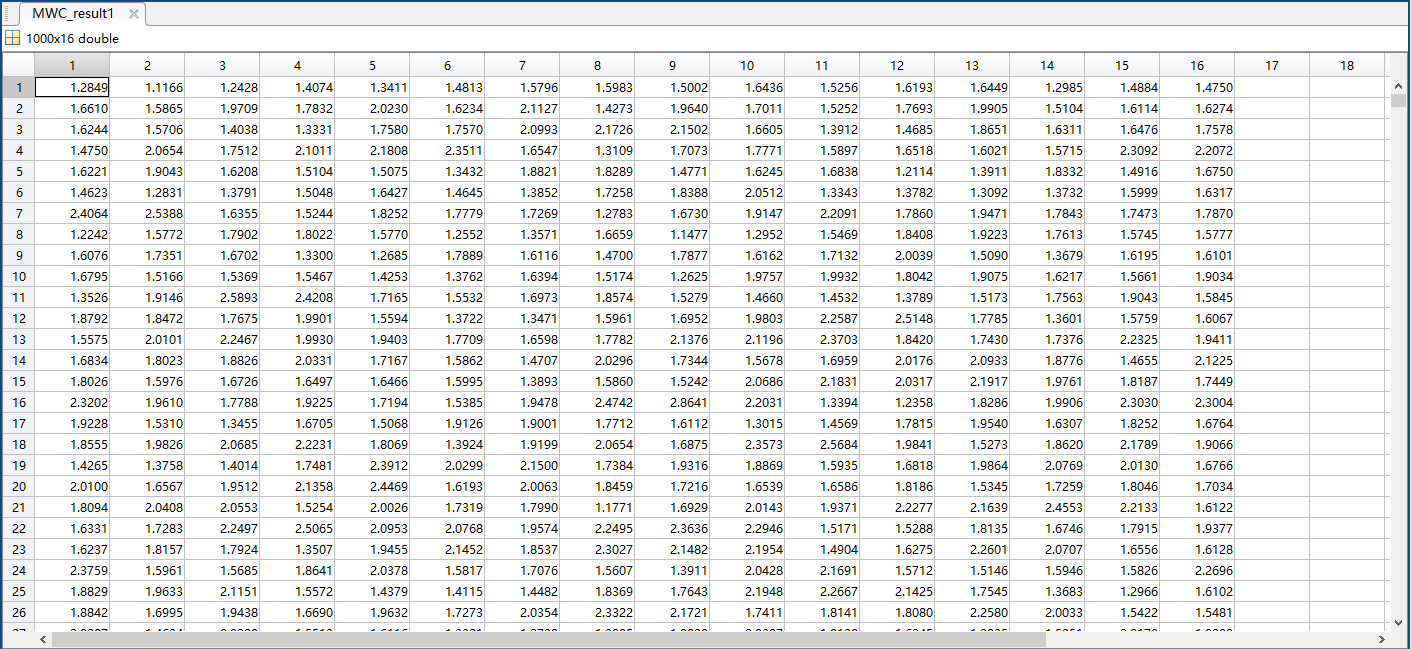


Subcarriers (column) and users with MC allocation

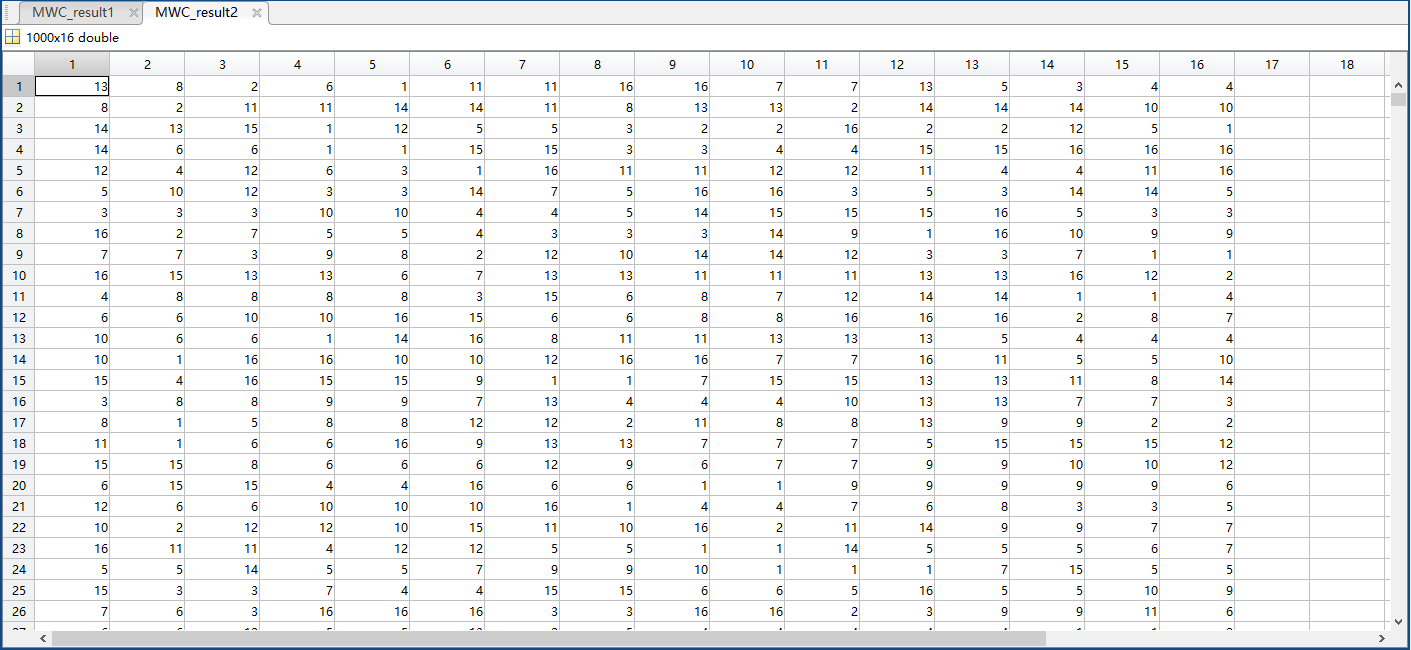
The average gain of MC allocation is 1.8075. Compared with random distribution, the gain is obviously larger, which leads to a higher data rate as the Formula suggests. As for the subcarrier allocation, it can be seen that the result is dissimilar for different channels. For instance, subchannel 1 in channel 1 is assigned to user 4 and that in channel 2 is allocated to user 10. Since the CSI is changing all the time, the subcarrier should be reallocated in every time slot. In other words, due to the time-varying characteristics, the subcarrier in the identical channel serving the same user can vary at different moments to maximise system capacity. Moreover, Figure b suggests that with this scheme, it is possible that some users are not served by any subcarrier at a certain moment as discussed above. In the simulation of this part, to reproduce the situation, the number of subcarriers and users are set to equal on purpose. Nevertheless, the phenomenon is rare in actual case where the subcarriers are far more than users in most cases. In addition, more subcarriers will be assigned to users only when the traffic cannot be handled by those already given to them. As a consequence, users can experience stable network connection provided by valid subcarriers.

## MWC Subcarrier Allocation

To compare the result with other subcarrier distribution schemes, the weight is only for reference, which is based on PD scheduling scheme that will be covered later. Figure a and b present the outcome of MWC allocation.



Subcarriers (column) and gains with MWC allocation



Subcarriers (column) and users with MWC allocation

According to the result, the average gain is 1.7674 with the effect of weight. In contrast with MC, the gain is decreased by 2.2 % hence the capacity, but traffic shaping is enabled as mentioned above, leading to a flexible network that can fulfil various requirements of different services. As a consequence, this scheme will be utilised in the cross-layer design on the basis of different data scheduling algorithms to optimise the system performance.

# Power Allocation Schemes

## Modified Equal Power Allocation

As the name suggests, equal power allocation means dividing the available power into equal amount and assigning the shares to subcarriers. The power allocated to subcarrier *n* serving user *k* can be represented as:



Equal power allocation is the most basic scheme that does not require CSI. It is assumed that all subcarriers are available and in same condition. In such case, the subcarrier with large gain will be treated identical as those in poor condition or even invalid. Simple as it is, this strategy is not efficient in power utilisation, which limits the maximum data rate as Formula indicates.

In this project, a modified version is proposed for power efficiency optimisation. The power allocated to subchannels are proportional to the square of the corresponding gains:



where *hk,n* is the gain of subcarrier *n* on user *k.*

Modified equal power allocation divide the available power into equal shares and distribute them according to CSI. As a consequence, the power efficiency can be enhanced.

## Water-filling (WF) Algorithm

It is demonstrated in [24] that iterated WF algorithm is the optimal power allocation scheme. However, the optimal solution described in [12] satisfying the restraints is with high complexity, and a suboptimal algorithm was proposed by [25] for a trade-off between complexity and performance. In the design, the suboptimal solution is employed to provide fast arrangements.

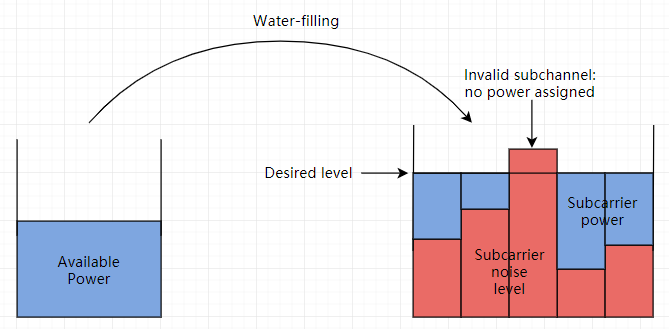


Illustration of water filling algorithm

The basic idea of water-filling scheme is presented by Figure. First of all, a desired power level determined by channel characteristics is proposed as a threshold for all subchannels. Then, the CNR of subchannels are examined to ascertain the noise levels. For a fixed modulated carrier signal power, the subcarrier noise level is high when the CNR is low. After obtaining CNR thus power levels of subchannels, power allocation is executed based on the difference between the threshold and the noise level. The principle is to ensure the sum of filled power and noise power equals the desired power level for every subcarrier. As a result, the satisfying subchannels with high CNR thus low noise power accept more power for transmission. For those invalid subchannels with noise exceed the threshold, they are abandoned in a time slot and no power will be allocated to them. In the end, the sum of power allocated to subcarrier should equal to available power. This algorithm guarantees more power to efficient subchannels, therefore higher data rates.

In algorithm realisation, it is assumed in the beginning that the subcarrier allocation is performed based on MWC scheme discussed above. Then using Karush-Kuhn-Tucker (KKT) conditions [26], the controller allocate power to subcarrier *n* of user *k* by:



with  being the user with maximum weighted data rate to be served. Subsequently, the validity of the solution is to be checked by comparing with zero. On condition that the power is zero or negative, it corresponds to the subchannel with low CNR that cannot be utilized for transmission, as the one marked as invalid subcarrier in Figure. Once the fallacious subchannel occurs, it will be removed from the available set, the entire result so far will be abandoned, and power allocation will be restarted on the basis of updated available subchannel set. The steps above will be repeated until all accessible subcarriers are assigned with valid power.

With the aid of WF algorithm, the subchannels with higher CNR can obtain more power while those in unsatisfying condition will be assigned less or no power. As a consequence, the maximum achievable data rate can be enhanced as Formula suggests.

## Weighted Water-filling (WWF) Algorithm

The relationship between WF and WWF is similar to that between MC and MWC. By introducing weight to WF algorithm, the only difference is the power allocated to subcarrier *n* on user *k* whose weight is *Wk* is now:



The expression indicates that not only the subchannels with better CNR but also the users with higher weight will be allocated more power. It is a reflect of QoS and traffic shaping on power allocation for network optimization. With various weight design satisfying diverse requirements, the power distribution scheme can be adjusted for flexible multiservice data transmission.

# Simulation Result of Power Allocation

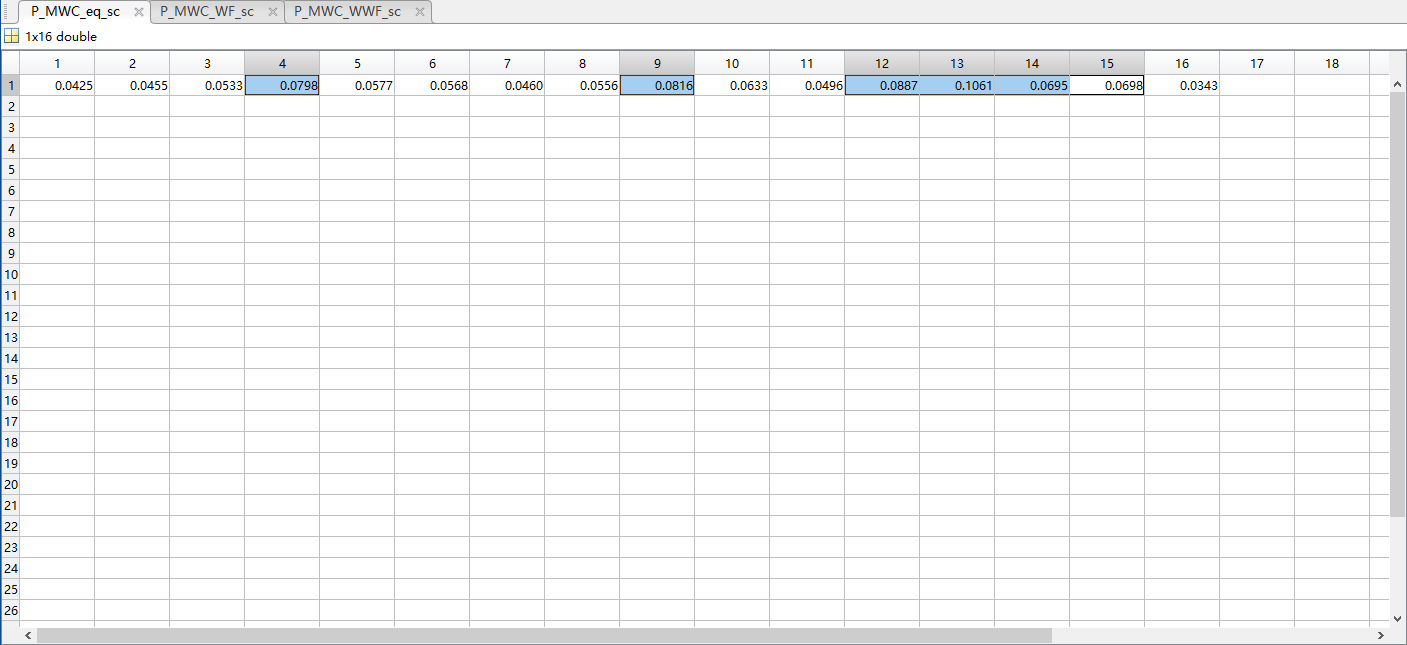
The simulation result of power allocation here is based on the parameter in Table. It is assumed that subcarrier distribution follows MWC scheme and the weight is derived from PD algorithm. Only one channel is considered to make things clear.

Parameters of power allocation simulations

|  |  |
| --- | --- |
| SNR (dB) | 10 |
| Number of users | 16 |
| Number of subcarriers | 16 |
| Number of paths | 6 |
| Number of channel | 1 |

## Modified Equal Power Allocation

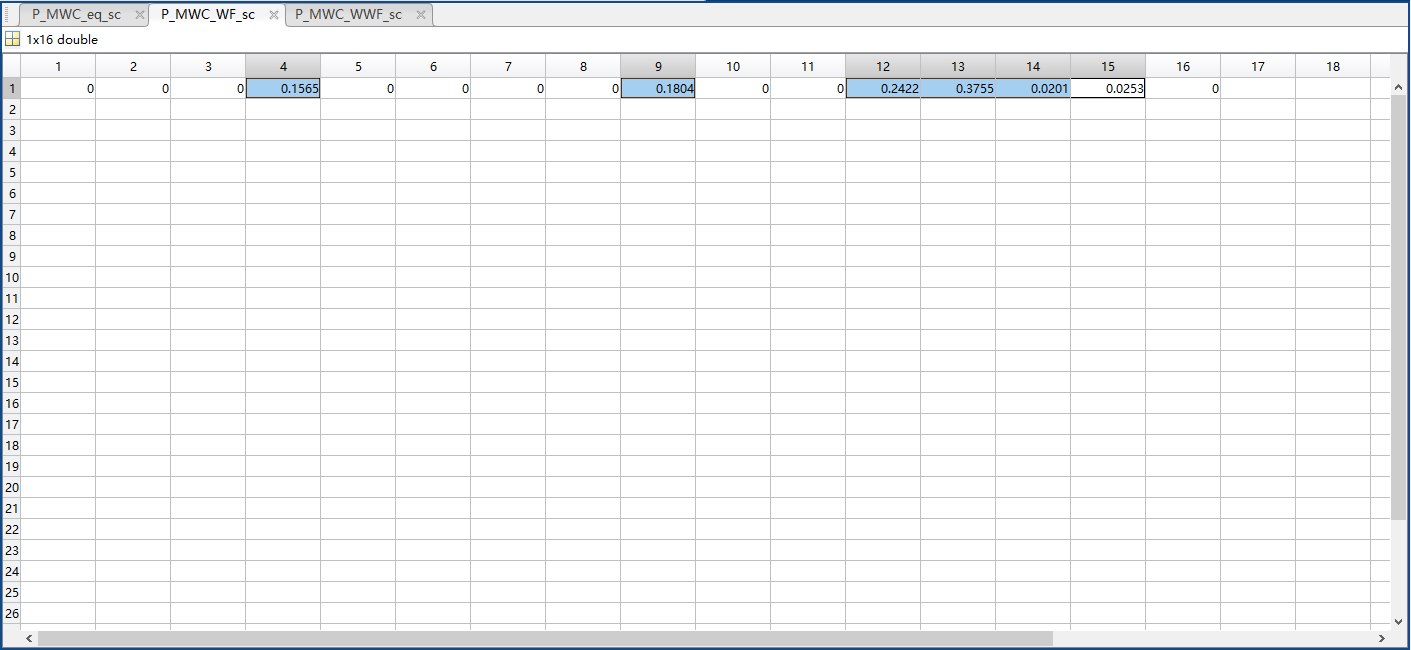
With modified equal (M-EQ) power allocation strategy, the power assigned to each subchannel are shown in Figure.



Subchannel power with M-EQ allocation

The selected subchannels 4, 9, 12, 13, 14 and 15 are assigned more power than the rest. It is due to the gains in the corresponding subcarriers are relatively large. Although CSI of the others are not as satisfying, they are still utilized for transmission with some assigned power.

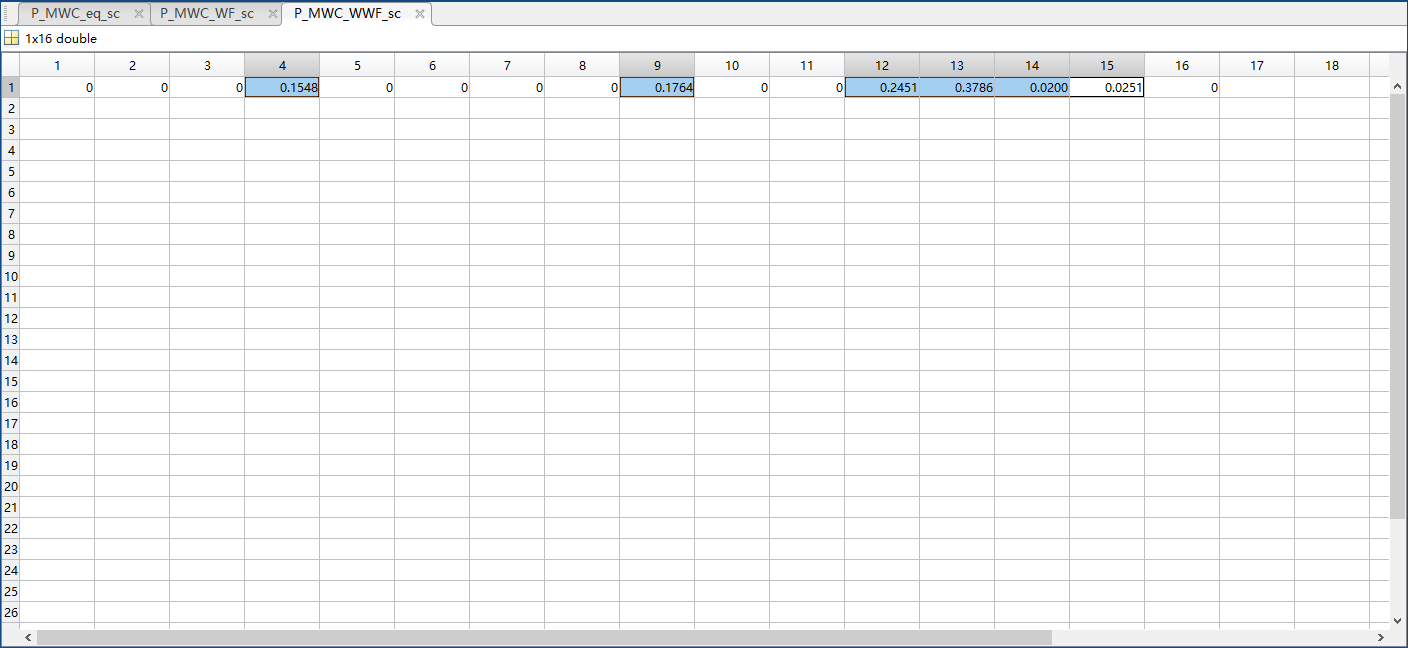
## WF Algorithm



Subchannel power with WF algorithm

Figure indicates the power allocation situation based on WF algorithm. Compared with modified equal power allocation, the total transmitted power remains unit but the subcarriers in poor condition are dropped by not assigning power to them. Therefore, those with acceptable CSI can be allocated with more power, as the selected ones. As Formula implies, the overall achievable data rate can be increased with this approach.

## WWF Algorithm



Subchannel power with WWF algorithm

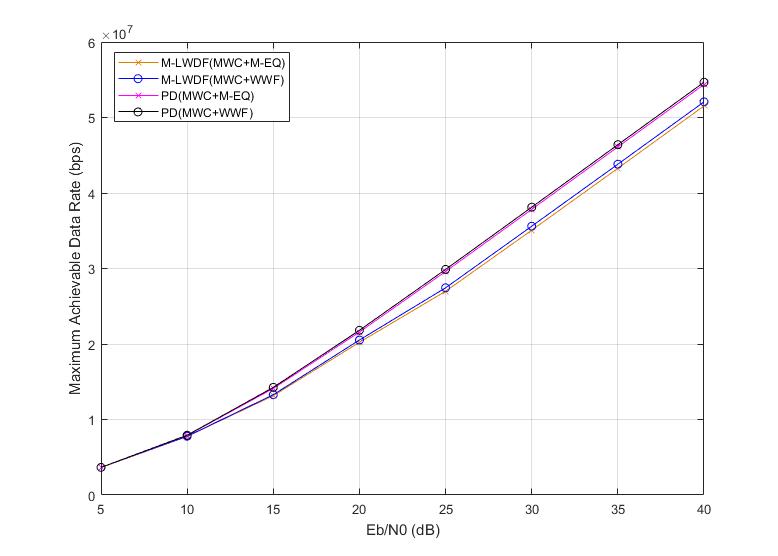
It can be seen from Figure that the power distribution is similar to that of WF. The subchannels utilized are still those but the assigned power is with tiny differences. For instance, subchannel 4 is with less power and subcarrier 12 is with more. The reason is that the weight of the user that subchannel 4 is serving is with smaller weight than the average, while the one communicating through subcarrier 12 is with higher priority. Despite the loss in capacity, it provides flexible traffic controls that can fulfil exact demands of various services.

## 4.4.4 Maximum Achievable Data Rate

On the basis of MWC subcarrier allocation and PD/M-LWDF weight design, the maximum achievable data rate of M-EQ and WWF power distribution strategies are compared in Figure. Relevant assumptions are revealed by Table.

Parameters of achievable data rate simulation

|  |  |
| --- | --- |
| SNR (dB) | 5 - 40 |
| Bandwidth | 5 MHz |
| Available power | 1 W |
| Number of subcarriers | 512 |
| Number of users | 32 |
| Number of paths | 6 |



Maximum achievable data rate for different strageties

It can be concluded that for both scheduling schemes, WWF algorithm result in more available data rate than M-EQ. In spite of the gap being narrow, the network throughput can be enhanced to some extent, and the adaptive traffic control can be enabled by WWF algorithm.

To sum up for the PHY layer, with the aid of this cross-layer model, conventional subcarrier and power allocation schemes as MC and WF can be further optimized by combining with the user status and weight information. For instance, no more subcarriers and power will be given to users whose traffic can be satisfied with existing set. Moreover, the user with high priority will be considered in precedence for subchannel assignment, and the power distributed will be more than average value. With the cross-layer design, the available resource of the wireless network is utilized efficiently and flexibly, enhancing user experience without extra cost.

1. Data Scheduling at the MAC Layer

Scheduling indicates determining the sequence of packet transmission. In contrast with the fundamental first in first out (FIFO) strategy, various scheduling schemes that send packets in specific orders to fulfil different demands of varied services are widely employed in contemporary networks. For a fixed capacity, traffics can be balanced to improve the throughput, delay and packet loss rate for some urgent services. A user watching video while downloading file tends to favor stable streaming compared with full bandwidth, and those playing games over music may desire low delay and packet drop rate. As a consequence, an adaptive scheduling approach should be able to understand the requirements in various cases and reshape the traffic to accomplish them.

To solve the problem, weight design that based on factors as delay, tolerance, packet size and importance is utilized to determine QoS priority. Packets with longer delay, stricter tolerance, larger size or more importance are assigned with a larger gain, and controllers send those packets in precedence. That is to say the transmission order depends on the packet characteristics.

Packets are arriving in queues that categorized according to the service type. For example, a multitasker is on video call while sending file, but the two services are carried by different queues. The packets from those queues arrive at buffer at the same time, but only one packet can be transmitted at a time, leading to the delay of the others. The affect is negligible when the bandwidth is sufficient. However, if the CSI is not acceptable, the transmission sequence should be considered carefully to minimum the influences.

Two strategies of weight design will be covered in this project, namely queue-based M-LWDF and packet-based PD. It is assumed that there are *K* users each with *I* queues holding *L* packets respectively. User index is indicated by *k* , queue index by *i* **, and packet index by *l*  .

# Modified Largest Weighted Delay First (M-LWDF) Scheduling

This queue-based scheduling scheme tends to serve the one with largest weighted delay with priority, thus the delay as the main consideration is kept in a bound. If *Si* and *Ui* are used to denote the head of line (HoL) packet delay and the delay tolerance for queue *I*,  is defined as the maximum allowed probability that packets are dropped  and  to be the average data rate over slots of queue *I*, then the weight of queue *i* [12] can be determined by:



It can be inferred that the queue with a larger HoL delay over delay tolerance, more demand for outage, or higher rata rate relative to the average will be assigned with a larger weight thus served in precedence.

Queue-based M-LWDF can be viewed as an extension of FIFO. At the beginning of each time slot, the weights of all queues are calculated then compared to determine which queue is to be served in this slot. This procedure is shown by Figure.

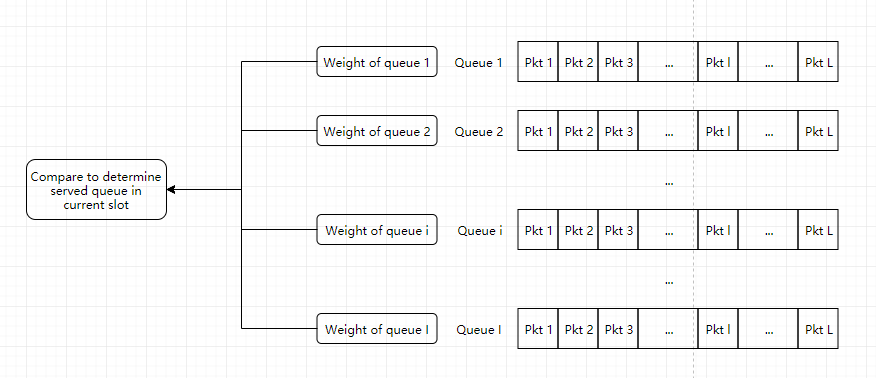


Illustration of M-LWDF principle

It suggests that although M-LWDF can adjust the traffic to guarantee throughput and keep the delay below a level, the mechanism determines that the transmission order for each queue is still FIFO. That is to say a voice packet arriving later than video packets can be served earlier, however it should always stay behind of last voice packet. As a consequence, a more flexible approach that enables urgent packet to cut in line is wanted when the urgencies of packets in the same queue vary in large scale.

# Packet Dependent (PD) Scheduling

As implied by the name, PD scheduling deals with individual packets directly, rather than relying on queues. In contrast with the conventional queue-based strategies that allocate equal weights to all packets in a queue, PD scheme assigns weights to packets based on characteristics of them, namely delay, delay tolerance, packet size and QoS priority.

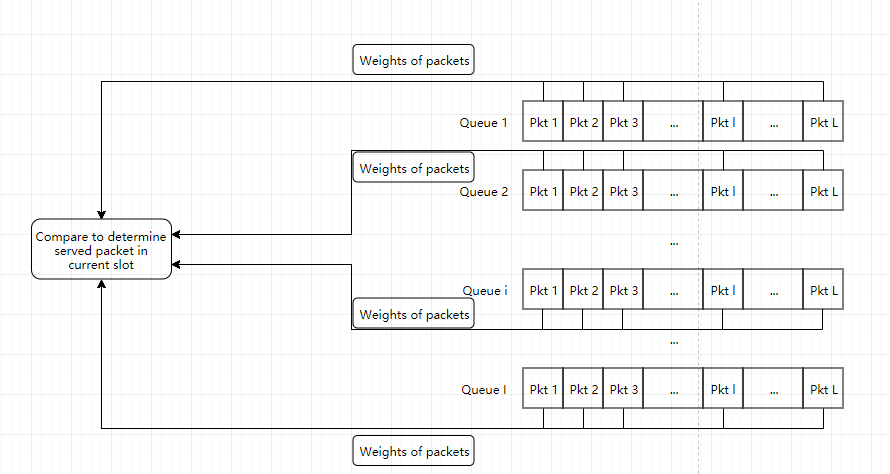


Illustration of PD principle

The principle of PD scheduling is as Figure displays. In a certain slot, at the transmission end, weight of the packets with the possibility to be transmitted are calculated. Then, weights are to be compared and the one with the largest weight will be served. The last step will be repeated until the capacity is not enough or all packets have been sent. With the aid of this approach, it can be guaranteed that those packets have been served are more urgent than those remained in the buffer. It is effective especially when the CSI is not satisfactory, or the user amount is large than expected. In such cases, there are no enough capacity and the choice of packets to serve affects user experience directly.

Denoting the delay tolerance for queue *i* of user *k* by **. Packet is this queue with waiting time larger than that will be lost. Define  as the guard interval to reduce packet loss rate. Those with delay reaching this region will be considered as urgent that need immediately serving before dropped. The current time is indicated by , and the time that packet *l* arrives is marked as . Therefore, the delay for packet *l* of queue *i* of user *k* can be determined by . Furthermore, the threshold of crucial packets can be expressed as , and the remaining time after which packet *l* will become urgent is . On condition that , namely the waiting time so far exceeds the delay threshold, the packet will be marked as urgent and thus given a precedence to serve. Letting  express the packet size and  denote the priority level, the weight of packet *l* of queue *i* of user *k* follows the equation:



The basic idea of PD weight design is illustrated by Figure.

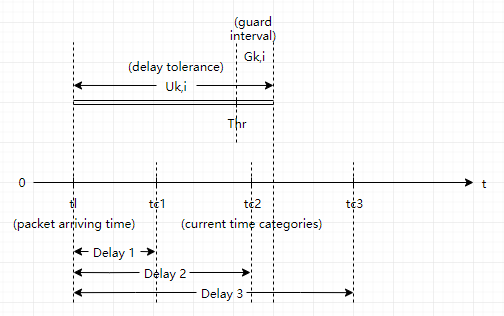


Illustration of PD weight design

For a packet arrived at , there are three possible situations of weight design based on different waiting time.

1. Non-urgency Packets 

When the time difference between packet arriving time and current time is shorter than the threshold as case 1, the weight of packet determined by Formula is relatively small. Differentiating weight with respect to remaining time to be urgent:



It suggests that when the waiting time of a certain packet is short, *i.e.*, the remaining time for it to be urgent  is large, not only the weight is small but also the impact of delay variation on the weight is slight. As a consequence, the priority level and packet size are critical factors of weight. For instance, two video packets of the same size with waiting time 3 and 4 ms are with small weights close to each other, but a haptic packet of the same size and a larger video packet that waits 3 ms can be with relative larger weights. In contrast, for those approaching threshold, the weight will increase exponentially to remind the system of their significance. That is to say the delay has major influence compared with other factors when it is close to the threshold.

1. Urgency Packets 

As the second half of Formula and part 2 of Figure indicates, if a packet is in guard interval, the weight will be fixed as the maximum value that it can reach. Compared with normal packets, those urgent tend to be assigned larger weights and transmitted in precedence to avoid packet loss. In this case, delay is not considered in weight design, which means the transmission order only depends on QoS priority level and packet size.

1. Lost Packets

For every kind of service there is a fixed tolerance. On condition that any packet waits longer than the tolerance, no matter the reason being unacceptable CSI, large traffic or network error, it will be lost and removed from the queue. Therefore, they will be dropped by the system and not assigned with valid weight.

As mentioned above, the packet-based weight design of PD scheduling enhances the system flexibility by enabling crucial packets cutting in line, but the cost is the system complexity, since numerous weights are to be updated and compared frequently. Thus, to achieve a balance between complexity and performance, only a subset of packets  are selected for weight design of queue *i* of user *k.* The subset can be sorted into those urgent in guard interval denoted by , and the rest marked by . As a consequence, the weight of user *k* can be approximated by the selected packets:

The proper choice of packets for weight estimation can reduce system complexity while maintaining accuracy. It is estimated in the previous works [12] [13] that a reasonable value of  is below 100.

# Simulation Result of Scheduling

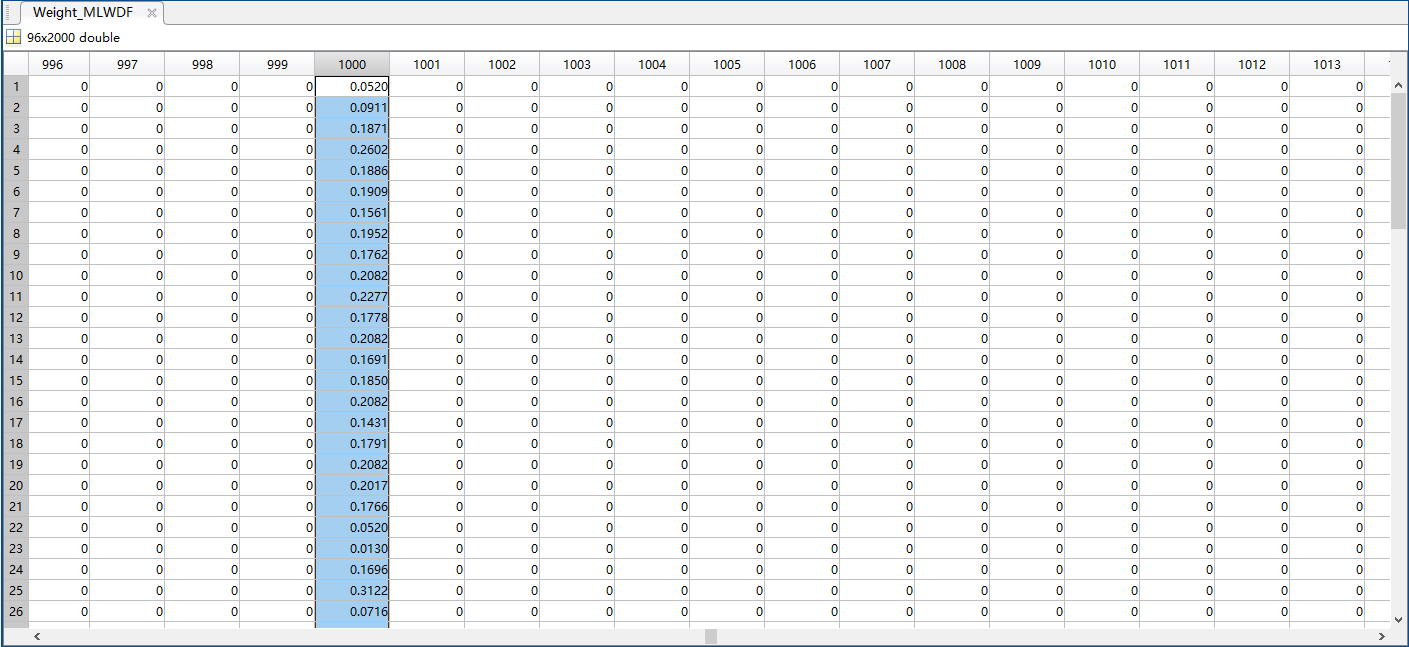
In this part, the result of M-LWDF and PD schemes are investigated to examine the characteristics of weight design. The simulations are based on parameters revealed by Table, and the weights are calculated at the 1000th time slot.

Parameters of scheduling simulation

|  |  |
| --- | --- |
| Number of users | = 512 |
| Number of queues per user | *I* = 3 |
| Number of total slots | = 2000 |
| SNR | *SNR =* 20 dB |
| Slot duration | = 2 ms |
| Packet arriving interval | = 1 ms |
| Guard interval | = 1 ms |
| Voice tolerance | = 100 ms |
| Video tolerance | = 400 ms |
| BE tolerance | = 1000 ms |
| QoS priority level of voice | = 1024 |
| QoS priority level of video | = 512 |
| QoS priority level of BE | = 1 |
| Maximum voice drop probability | = 0.05 |
| Maximum video drop probability | = 0.05 |
| Maximum BE drop probability | = 0.77 |
| Packet arriving rate of voice | = 64 Kbps |
| Packet arriving rate of video | = 120 Kbps |
| = 239 Kbps |
| = 420 Kbps |
| Packet arriving rate of BE | = 500 Kbps |

## 5.3.1 M-LWDF

Figure displays the weight design of M-LWDF.



Weight instance of M-LWDF

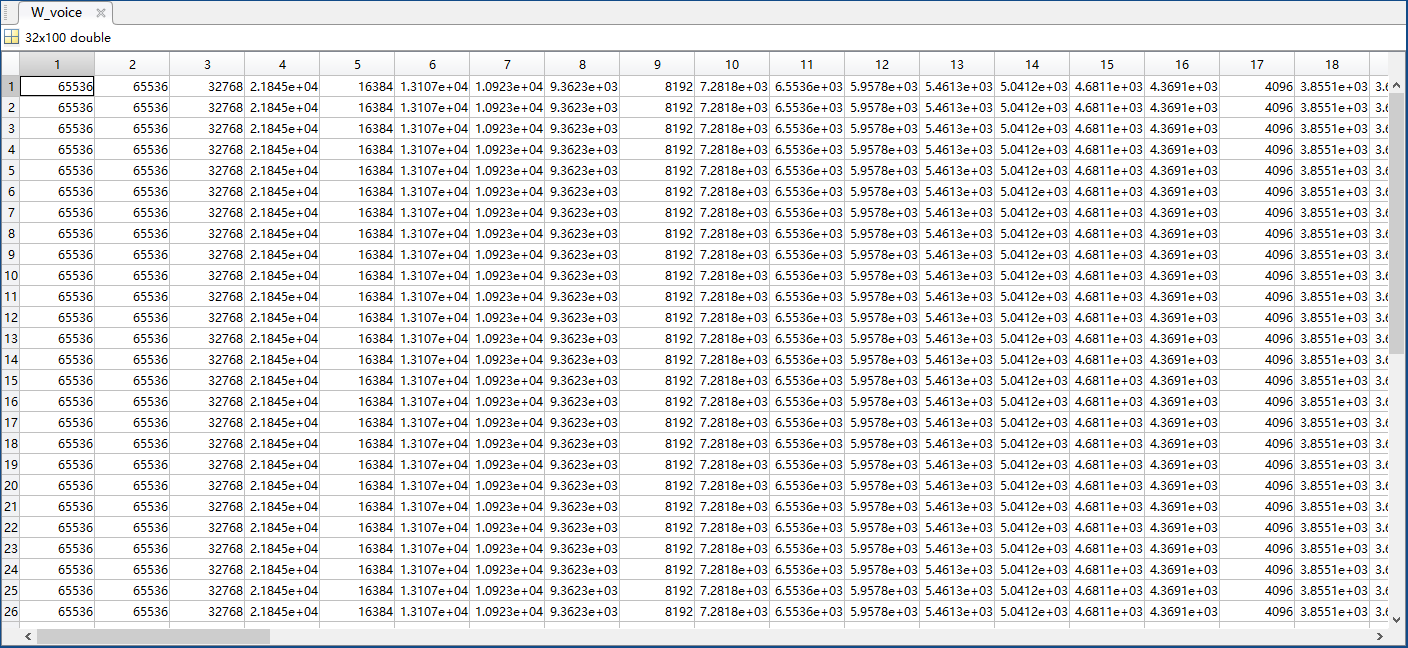
The size of matrix is since there are 2000 slots and 32 users each with 3 queues. At the 1000th slot, the weights are shown in the 1000th column. The weights of each user occupy three rows, with the first for voice, second for video and third for BE. For instance, user 2 corresponds to rows 4-6 whose weights for three queues are 0.2602, 0.1886 and 0.1909 respectively. As mentioned above, the user with maximum weight summation has the priority in transmission, and the queue with the largest weight among the three will be served first. The weight matrix will be renewed at the beginning of each time slot based on updated buffer information. After a certain user being served, the transmitted packets are removed from buffer thus the weight should be reduced, which leads to lower priority in the following slots. As a consequence, this strategy ensures the queue with largest weighted delay to be served first.

## 5.3.2 PD

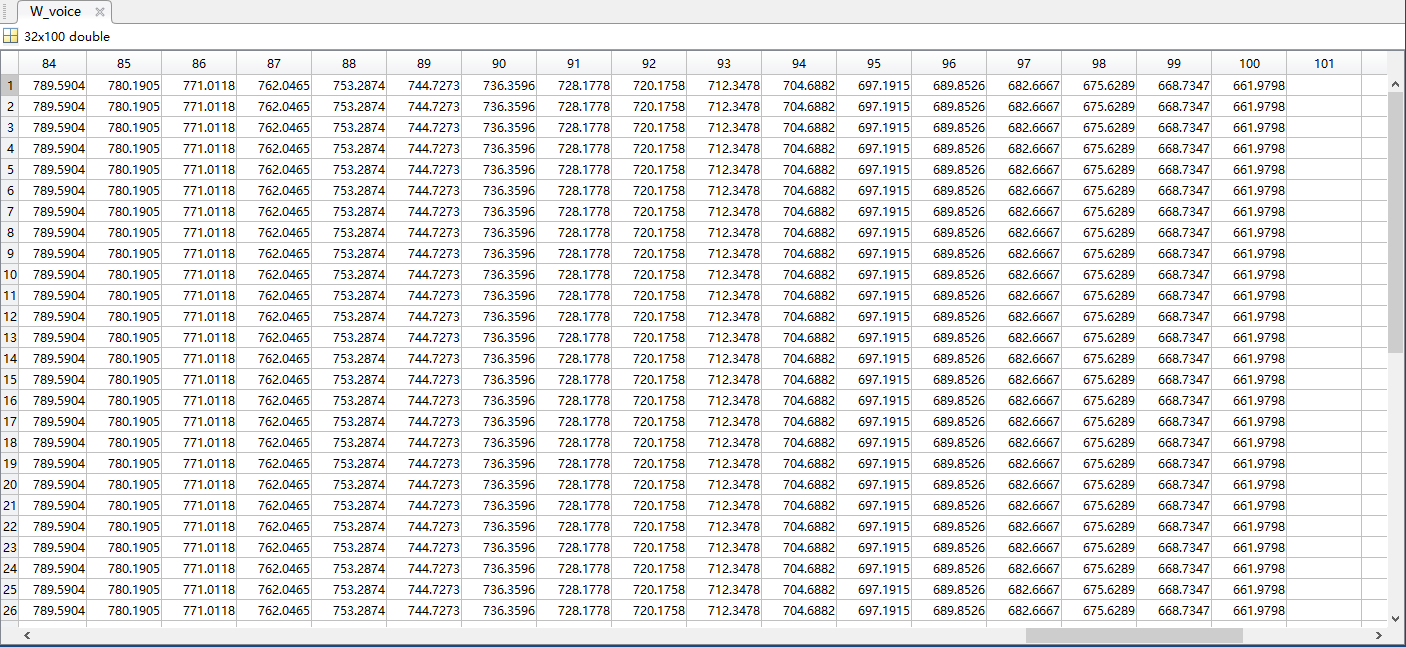
In this part, weight of all packets in buffer are calculated to check the region that maximum weight can occur. With relevant result, the subsets to be selected can be determined to reduce the system complexity as discussed in Section 5.2. Owing to the limitation of space, only the head and tail of the result matrices are presented for discussion.

Different from the previous result, the rows indicate users while the columns refer to packets. As a consequence, the sizes of weight matrices are user quantity (32) times packet quantity (100 for voice, 400 for video and 1000 for BE). The row number corresponds to user and the column index indicates the sequence of packets arriving.

The weight of voice packets is shown by Figure a and b. Voice packet size is fixed as 64 bits.



Voice packet weight instance of PD part i

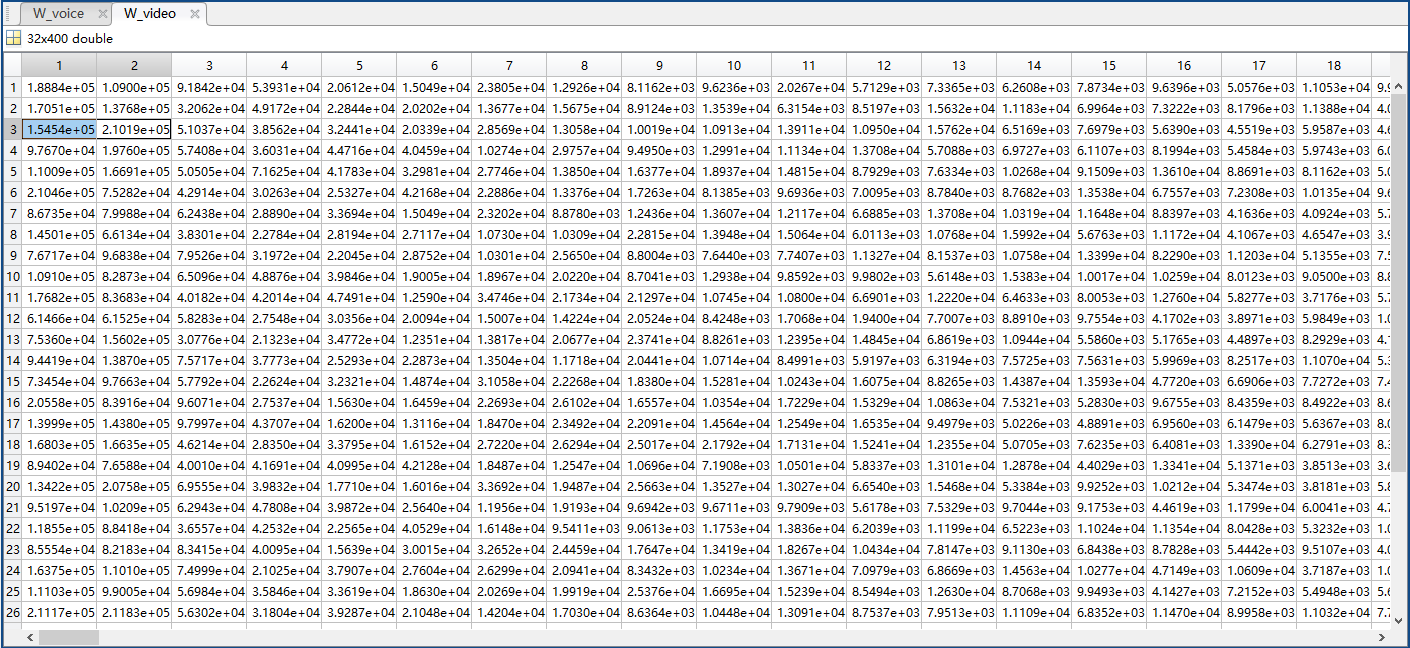


Voice packet weight instance of PD part ii

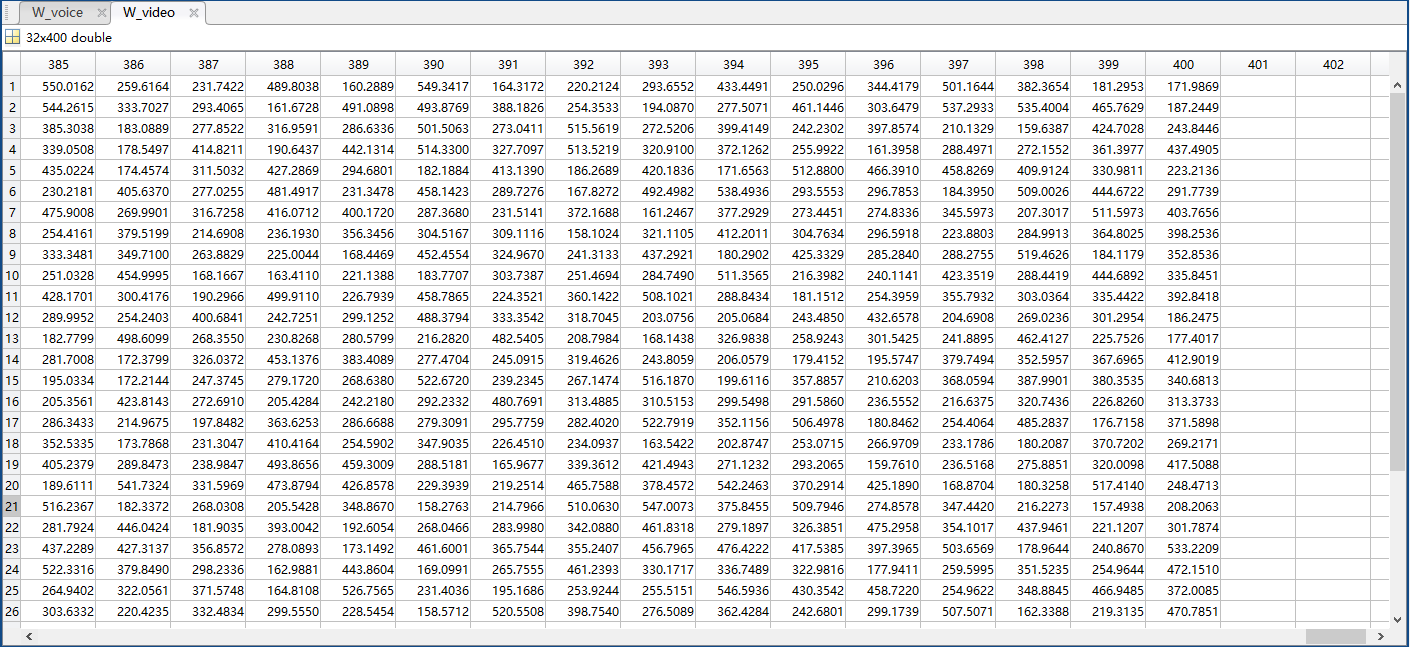
It is revealed that weight of packets arriving at the same time are equal for all users. The phenomenon is as expected from Formula, since the priority level and the remaining time to be urgent are the same for those voice packets came simultaneously. Moreover, the only changeable factor or packet size is fixed according to the characteristics of voice traffic. Therefore, the weight difference between voice packets is reflected by waiting time merely, which means those with longer delay are allocated with larger weight. Furthermore, those in guard interval are assigned the maximum possible weight as packet 1 and 2 in the Figure i.

As discussed in Section 5.2, the impact of waiting time on weight is significant when it approaches the tolerance, and relatively slight when that is short. It is demonstrated by the figure that one millisecond gap of delay result in 1 percent difference in weight value for the packets just arrived, while it leads to a double for those have waited long enough and are about to be dropped.

Result of video are illustrated by Figure a and b. The size of video traffic is determined by a truncated exponential distribution in different states, with the minimum, average and maximum value being 120, 239 and 420 bits. The duration of state depends on an exponential distribution with mean to be 160 ms [12].



Video packet weight instance of PD part i

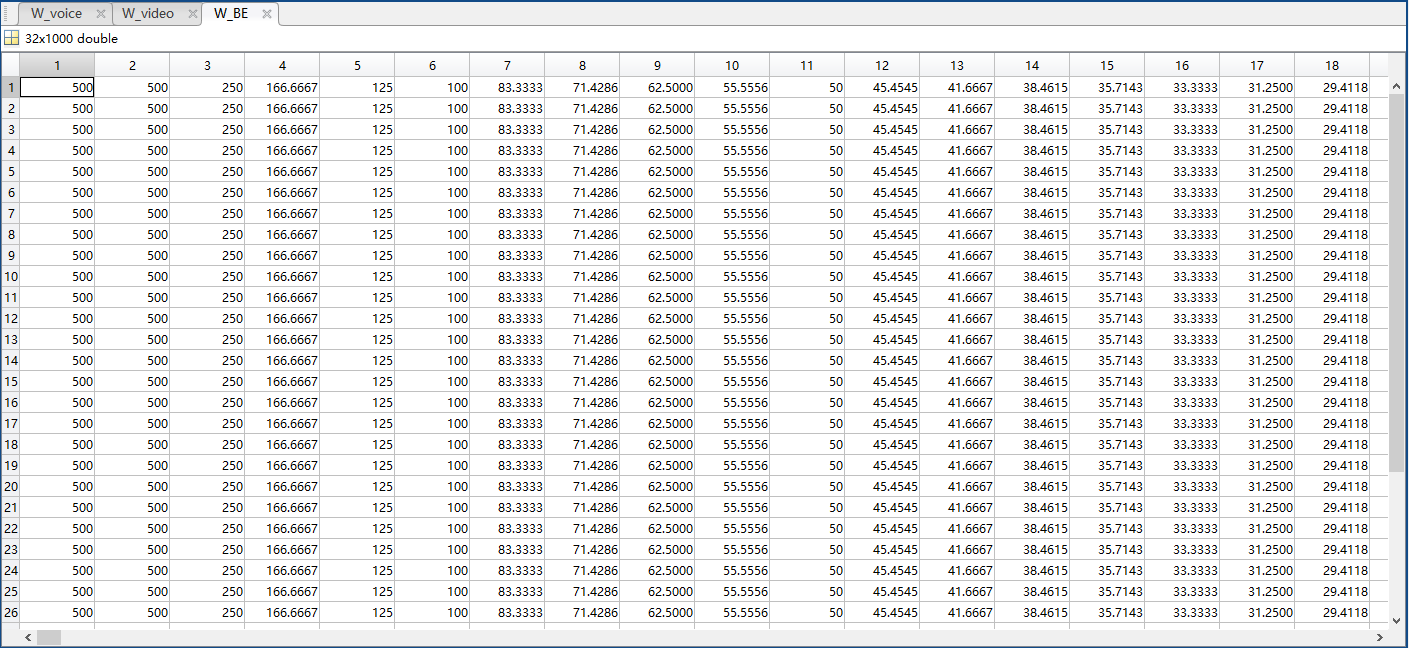


Video packet weight instance of PD part ii

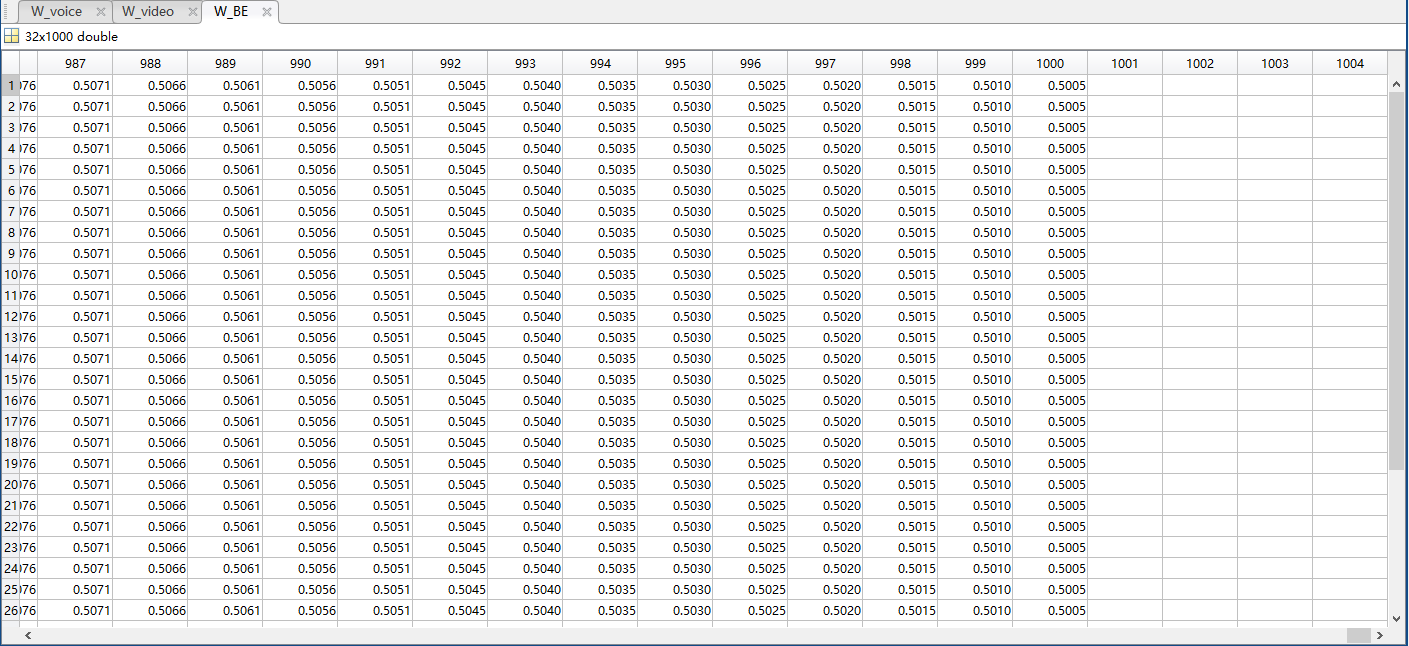
Despite the trend of assigning higher weight to the packets wait longer, there can be some special cases as the first two packet of user 3. In this instance, the packet with the second longest delay has highest priority, which will thus be sent earlier than the one with the longest waiting time. It can be explained by Formula that the influence of packet size exceeds that of delay.

Compared with voice, the overall weight of video is smaller than that of voice. It is determined by the characteristics of the service since the delay requirement of voice service is relatively higher.

Figure a and b present weight of BE packets.



BE packet weight instance of PD part i



BE packet weight instance of PD part i

Similar to video traffic, the arriving data rate thus packet size is constant. As a result, weight merely depends on remaining time for packets to be urgent, which implies that those with larger delay are assigned higher weight therefore served first. That is to say the voice and BE queues are FIFO when considered independently. Moreover, weight of BE packets is comparatively smaller than that of other services, and the most urgent BE packet should be served after the transmission of all existing voice packets. The phenomenon corresponds to the real case that user employs VoIP while downloading files, where the bandwidth should be allocated to BE only after transmitting voice packets.

According to the matrices, the video packet with largest weight tends to be found in the first 3 columns, while the packet in voice or BE queue that waits longest time should be assigned the largest weight. Nevertheless, on condition that only those packets are considered for weight calculation, then once they are served, the bandwidth will be wasted since there will be no available weight for comparison thus no more packet would be sent. On the other hand, if weight of all packets is calculated as occurred in this part at the beginning of every time slot, the complexity and response time will be increased in significant degree. As a consequence, to achieve a trade-off between performance and complexity, only some most long-awaited packets are considered in weight calculation. Due to the low-delay requirement and small size characteristics, all voice packets are selected for weight calculation since even the latest one deserves high priority to achieve real-time communication. For video service, if calculating weight for all packets, the system may not handle that much traffic at a time, and the system complexity will be affected. On the other hand, the delay and packet loss rate are also main considerations when evaluating the video stream, which requires sufficient packets be prepared for transmission. Therefore, it is determined that to take the first 75 most long-awaited packets into account. In the end, due to the limitation of bandwidth and low demand of BE traffic, only the first 50 most long-awaited packets are considered possible to be sent, therefore assigned weight through PD algorithm. The estimated packet numbers can be increased with better CSI if necessary. As discussed above, by assigning weight to the selected urgent packets that are possible to be served in the slot, the complexity is reduced with guaranteed performance.

1. Performance Analysis

On the basis of MWC subcarrier assignment and WWF power allocation schemes which have demonstrated their advantages, the performance of M-LWDF and PD scheduling are compared in terms of system throughput, packet delay, packet drop rate and outage probability. Apart from the conventional services, the influence of introducing haptic traffic to the existing stable network is investigated to check the support of current communication systems to new haptic-based applications such as AR and VR. Moreover, it can also examine whether those algorithms are able to provide a reliable service when new types of traffic with various demands appear, as happening in real cases.

The whole simulation process is divided into six parts. First, the parameters are defined in the main function and the packet generation is carried out in the beginning. Second, M-LWDF and PD scheduling are executed at the MAC layer based on the packet information and algorithms mentioned in Chapter 5. Weight can be obtained for queues and some selected packets respectively, which can be summed up to denote the priority of users. Third, a Rayleigh fading channel with six paths is developed and converted into frequency domain to create subchannels. Fourth at the PHY layer, according to the CSI and weight derived in the steps above, subcarrier and power control are realized by MWC and WWF schemes. Once they are determined, the maximum achievable data rate for users can be ensured and passed back to the MAC layer for data transmission. Controllers always select the queue or packet with the largest weight to serve when there is resource available. On condition that there are remaining subcarriers after serving the assigned users, they will be reallocated to others to maximize the resource utilization.

# System with Conventional Services

In this section, the cross-layer design is utilized in a conventional network with traffic categorized as voice, video and BE. Relevant parameters of simulated system are summarized in Table.

Parameters for conventional system simulation

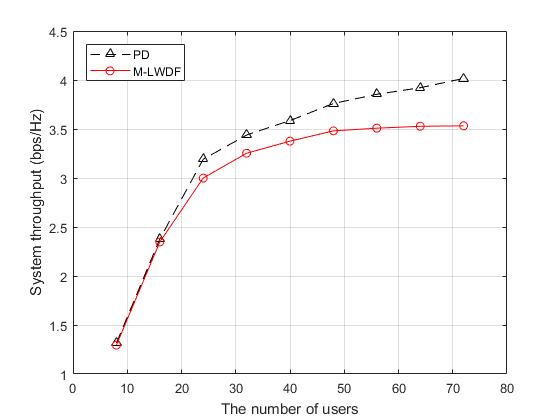
|  |  |
| --- | --- |
| Parameter | Value |
| Transmitted power | = 1 W |
| Number of paths | = 6 |
| RMS delay spread | = 0.5 μs |
| Bandwidth | *B* = 5 MHz |
| Number of subcarriers | = 512 |
| Number of slots | = 2000 |
| Slot duration | = 2 ms |
| Packet arriving interval | = 1 ms |
| Guard interval | = 1 ms |
| Number of queues per user | *I* = 3 |
| Voice tolerance | = 100 ms |
| Video tolerance | = 400 ms |
| BE tolerance | = 1000 ms |
| QoS priority level of voice | = 1024 |
| QoS priority level of video | = 512 |
| QoS priority level of BE | = 1 |
| Maximum voice drop probability | = 0.05 |
| Maximum video drop probability | = 0.05 |
| Maximum BE drop probability | = 0.77 |
| Number of voice packets selected for weight calculation | = 100 |
| Number of video packets selected for weight calculation | = 75 |
| Number of BE packets selected for weight calculation | = 50 |
| Packet arriving rate of voice | = 64 Kbps |
| Packet arriving rate of video | = 120 Kbps |
| = 239 Kbps |
| = 420 Kbps |
| Packet arriving rate of BE | = 500 Kbps |

As mentioned in Section 5.3.2, the video packet size follows a truncated exponential distribution for states with the mean of duration to be 160 ms.

For the system to be applied in real cases, there are two factors that can affect the network quality, namely the number of users and SNR. It is estimated in [12] that the number of users for each subsystem tends to be less than 70, and the range of SNR to be considered should be 5 ~ 30 dB. In the following parts, they will be analyzed independently by fixing one and varying the other. The network quality can be revealed by throughput, packet delay, packet loss rate, and outage probability of different services.

* + 1. Overall throughput

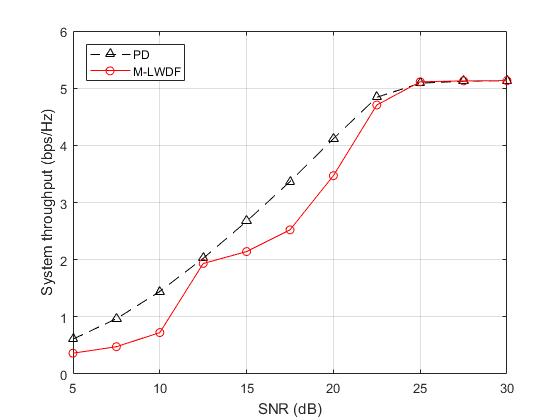
Figure indicates the influence of the user quantity on the throughput of cross-layer systems with different weight design. In this case, the SNR is set to 20 dB and the number of users varies from 8 to 72.



Impact of user number on throughput when SNR = 20 dB

It can be seen that when the user amount is less than or equal to 16, the system has enough resource for data transmission of all users. Therefore, throughput is nearly the same for M-LWDF and PD scheduling which approach the maximum possible value. For the system serving a large number of users (), PD proves its adaptive resource management capacity with relatively low complexity, as discussed in Section 5.3.2. It is due to the fact that with the growth of user amount, the diversity of packet is increased, leading to more freedom in resource allocation and thus enhanced multiuser diversity. Nevertheless, when the user number is large, the degree of weight difference among packets is raised when the user number increases. Therefore, the growing speed of throughput that can be reflected by the derivative of the curve slows down when the user amount is large. Compared with PD, the influence of this phenomenon is more severe on queue-based M-LWDF, which determines the transmission sequence by weight of HoL packet in each queue. As a consequence, the boost of multiuser diversity on the throughput of M-LWDF is less than that of PD. When the user amount is large, PD outperforms M-LWDF in system throughput.

To compare the performances of PD and M-LWDF in another dimension, the throughput variation over changing SNR is examined as well. In the simulation, it is assumed that the system is shared by 32 users with SNR in 5 – 30 dB.



Impact of SNR on throughput when user number = 32

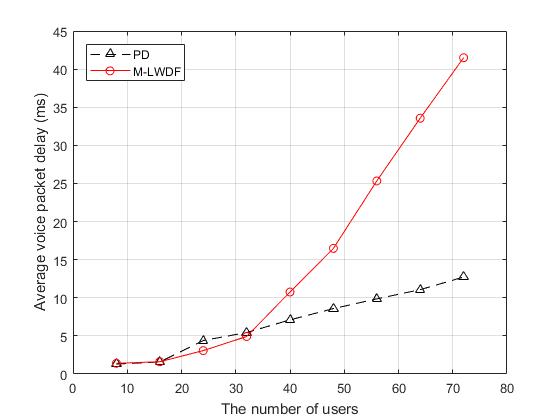
Figure reveals that when SNR is less than 25 dB, the throughput of PD is greater than that of M-LWDF, which is especially suitable for low SNR cases where the available resource is insufficient to transmit all packets. On the other hand, the importance of scheduling strategies is reduced when there is enough capacity for data traffic, as suggested by dB situations. According to the statistics, it can be calculated that at a typical case where dB, the bandwidth efficiency of PD is 24.7% higher than that of M-LWDF.

## Packet Delay

The delay of voice and video packets will be presented and discussed in this part. BE packets are not considered because there is no delay requirement on them.

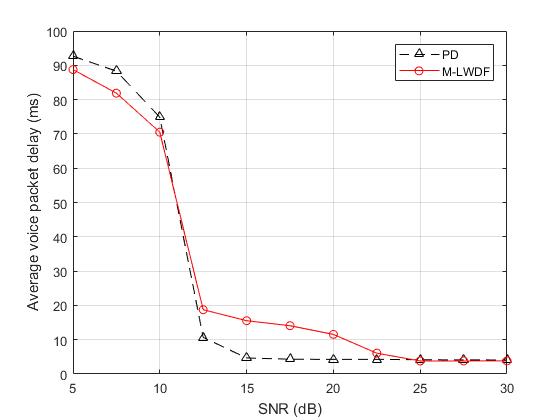
* + - 1. Voice Packet Delay

Voice traffic is regard as the one with highest priority. For different user set, the average voice packet delay is shown by Figure. It is assumed that the SNR is set to 20 dB.



Impact of user number on average voice delay when SNR = 20 dB

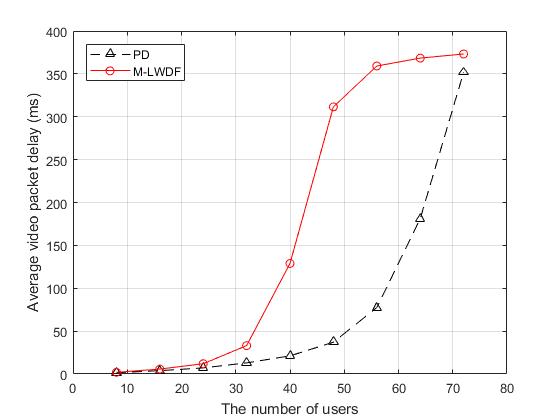
Both schemes experience little increasing of delay when the user number is varied from 8 to 32. In this region, the performance of M-LWDF is slightly better than PD. The reason is that with enough resource for transmission and zero packet loss rate, all packets tend to be with a relatively low delay. As a result, the voice packets should be served first no matter in which schemes. Therefore, the queue-based M-LWDF achieves a higher efficiency. Another possible explanation is that some trade-offs have been made between voice and video packets to reduce overall delay. In other words, the delay of voice packets can be further decreased but the cost is to increase the video packet delay. Since the network can fulfil the requirements in this region, a balanced scheme is proposed to shorten the overall delay. When the user amount is in the range , the voice packet delay of PD is much lower than that of M-LWDF. For a system shared by 72 users, the average voice delay is about 42 ms for M-LWDF, more than three times compared with PD. The reason is that for M-LWDF scheduling, a certain queue with largest weight is to be served, in which there can be some non-urgent ones that have larger size than urgent packets. Since the maximum data rate is fixed, the delay of urgent packets in other queues can be increased.



Impact of SNR on average voice delay when user number = 32

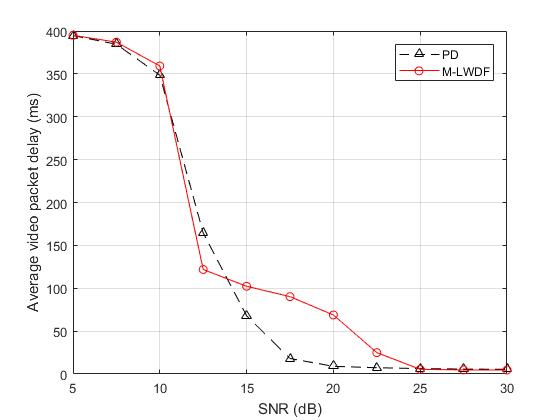
Figure shows the influence of SNR on average voice packet delay for system with 32 users. In low SNR range ( dB), M-LWDF is more advanced than PD in voice delay control. The reason is that the majority of packets are remained in buffer when SNR is not satisfying, and M-LWDF always assigns the maximum weight to voice queue. That is to say although there can be other urgent video packets, voice traffic tends to occupy more resource. Therefore, the voice delay is relatively low for M-LWDF. On the other hand, PD verifies its dominance with packet-based weight design. The mechanism avoids the non-urgent packets in the queue with priority to occupy the resource that should be utilized on more important ones. At dB, the average voice packet delay of PD is around 5 ms, only one third of that of M-LWDF.

* + - 1. Video Packet Delay



Impact of user number on average video delay when SNR = 20 dB

Figure illustrates the impact of user amount on average delay of variable bit rate (VBR) video packets with dB. It can be discovered that when the user number is less than 32, there are sufficient resource to ensure low delay for both schemes. Moreover, when the resource is insufficient to support a large number of user, the video packets are to be controlled for two strategies with delays approach the tolerance to guarantee the quality of voice service. For the user amount , PD shows its massive advantage over M-LWDF on video packet delay. The simulation result suggests that for a cross-layer system employed by 48 users, the average video packet delay is around 310 ms for M-LWDF and 45 ms for PD. On account of weight design targeting at the selected packets, PD chooses the packet to serve one by one, rather than simply transmitting those in a queue as M-LWDF does. As a consequence, it avoids unimportant packets transmission which happens in M-LWDF for BE queues frequently, allocates more resources to QoS traffic, and thus reduces the packet delay.



Impact of SNR on average video delay when user number = 32

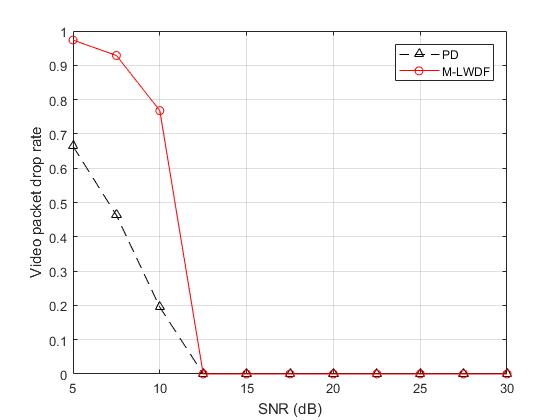
Relevant result when the number of user is fixed to 32 and the SNR is varied from 5 to 30 is presented by Figure. It can be seen that the video packet delay of PD is shorter than M-LWDF in most cases, as can be estimated based on the theory and the result above. Only at dB the video packet delay of M-LWDF is less than PD. It might be a special case, but a potential cause is that M-LWDF assigned more resource than anticipated to video queue now that the voice traffic has been satisfied as suggested in Figure 6.1.2.1-2. That is to say, the video queues are occupying some resources that should have been assigned to BE packets. This can lead to the low delay of video traffic and larger throughput as Figure 6.1.1-2 implies, but the quality of BE service will be affected. The corresponding pack drop rate of BE will be discussed in the following section. As a queue-based strategy, there can be some randomness in the performance of M-LWDF, as happened in this case. While PD that considers packets can divide the problem into smaller dimension, therefore result in more precision and stability.

## Packet Drop Rate

Drop rates for video and BE packets will be investigated in this section.

* + - 1. Video Packet Drop Rate

Figure shows the video packet drop rate for a 32-user system with varied SNR.



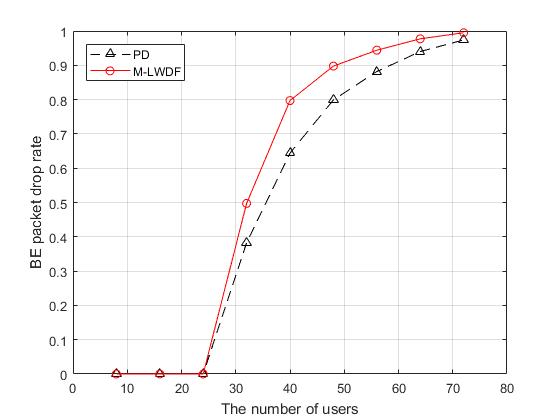
Impact of SNR on video packet drop rate when user number = 32

It can be found that video packets are possible to be lost only at the low SNR ( dB) cases. As the service with the second precedence, the quality of video data transmission can be guaranteed after ensuring reliable voice traffic communication. On condition that the capacity is insufficient, PD proves significant advantages over M-LWDF. It is because when the SNR is low, M-LWDF tends to allocate more resource to voice traffic, as mentioned in Section 6.1.2.1 and suggested by Figure 6.1.2.1-2. Thus, the voice traffic is ensured to be with low delay and packet loss rate, but the cost is the influence on video packets. In comparison, there is a balance between voice and video packets in PD weight design. Once the reliable voice service is established, PD will balance the resource allocation for voice and video traffic, rather than focusing on minimizing the voice delay only. As a consequence, at low SNR case, the performance of PD in voice delay is worse than M-LWDF, but an optimization of around 50% video packet drop rate demonstrates its rationality.

At dB the video packet drop rate is extremely low (less than 0.005) for user number , benefiting from QoS design of both strategies. Although the condition can be adjusted to present the performance, it is believed that the discussions should be based on the same parameters for clear comparison. Moreover, the result above has demonstrated the advantages of PD over M-LWDF in video packet delay for varied SNR. Therefore, the situation corresponds to different user number is not presented.

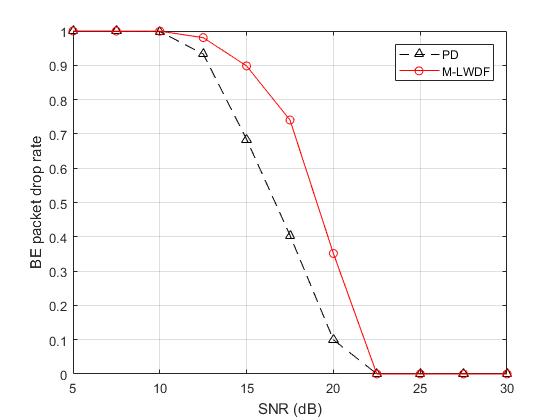
* + - 1. BE Packet Drop Rate

Compared with voice and video packets, the QoS priority of BE service is relatively low, thus BE packets should be with a higher packet drop rate. It is analyzed for different user number and SNR respectively to provide a comprehensive discussion.



Impact of user number on BE packet drop rate when SNR = 20 dB

It can be observed from Figure that both schemes can achieve zero packet drop rate when the user number is low (). On the other hand, when the system capacity is not enough, the resource will be given to voice and video traffic in precedence, causing the BE packet loss. With the linear increasing of users, the BE packet loss rate climbs sharply in the beginning (), reaching a saturation region then slows down () with value approaching 1. Moreover, PD outperforms M-LWDF when the resource is insufficient to support multiusers (). The biggest difference occurs at user number equals to 40, where the drop rate of PD is around 0.65 and that of M-LWDF is almost 0.8. There is a phenomenon named “accumulation effect” can explain this. It is assumed that there is nothing in buffer before transmission starts. Now that the capacity is not enough, there should be more packets stuck in buffer as time goes on. Thus, the packet arrives later tend to wait longer than those already there since the queue is longer. Various services are influenced to different degree based on their priority levels. When the delay of a certain packet reaches the tolerance, packet loss happens. In such case, if the transmission still depends on queues as in M-LWDF, then if the served traffic is not the most urgent queue that is about to drop packets, it will not have the chance to mitigate the circumstance. That is to say, the resource may be given to those not as urgent as long as the group has the maximum weight, and at the same time the packets in other queues approaching the deadline (BE in this case) can be lost.



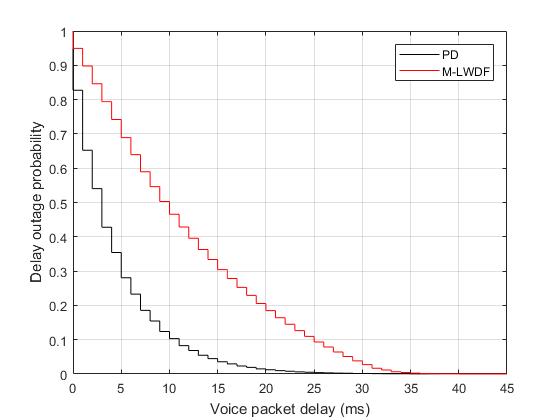
Impact of SNR on BE packet drop rate when user number = 32

According to Figure, in the range that BE packets should be selected to decrease the drop rate dB, M-LWDF is performing worse than PD. At dB which is a typical case, PD is only losing around 10% of BE packet while M-LWDF drops more than 30%. The result verifies the discussion above.

## Outage Probability

Outage probability reflects the packet delays by presenting the probability that the waiting time is longer than the argument. Please be aware that in the simulation, the minimum unit of time is millisecond, which means the delay of packets should be integer times of millisecond. As a consequence, relevant results are indicated by polylines. Furthermore, the simulation is based on a cross-layer system with 32 users at dB.

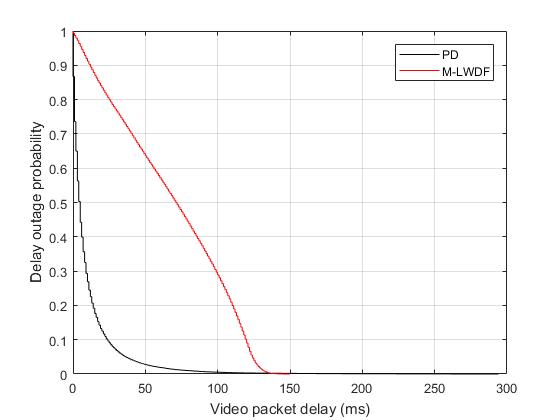
* + - 1. Voice Outage Probability



Voice traffic delay probability with 32 users when SNR = 20 dB

As indicated by Figure, the maximum possible voice delay is 25 ms for PD, which is 10 ms (28.6%) less than that of M-LWDF. Additionally, the voice packets with less than 10 ms delay takes up about 89% of the entire set, but for M-LWDF the proportion is only a half. Moreover, it is guaranteed by PD that for 95% of voice packet the delay is less than 14 ms, while the waiting time of 95% are ensured to be less than 33 ms in M-LWDF.

* + - 1. Video Outage Probability



Video traffic delay probability with 32 users when SNR = 20 dB

The performance difference in terms of video outage probability is more obvious. Figure presents that at a delay of 100 millisecond, PD has transmitted all the video packets, while M-LWDF still has around 30% to serve. PD can guarantee 95% video packets to be with less than 35 ms delay, but the promise of M-LWDF in the same case is about 125 ms. As a result, the average video packet delay of PD is far shorter than that of M-LWDF, as mentioned in Section 6.1.2.2.

# New Network with Haptic Data

As a new type of traffic, haptic data deals with the sense of touch that users can feel when using haptic service [27]. In other words, haptic technology can create the virtual objects that can be touched as if it were real. It has been widely utilized in applications related to virtual environment, such as VR and AR.

In [11], the requirements of haptic traffic are put forward. It is required that for stable haptic data transmission, the packet delay should be less than 50 milliseconds with jitter less than 2 milliseconds. Additionally, the packet loss rate is supposed to be less than 10 percent. As for the throughput, it is estimated to be 500 Kbps ~ 1 Mbps with constant packet rate.

Apart from the fundamental settings suggested by Table, relevant parameters for haptic data is shown in Table.

Supplementary parameters for haptic traffic

|  |  |
| --- | --- |
| Supplementary Parameter | Value |
| Haptic tolerance | ms |
| QoS priority level of haptic |  |
| Maximum haptic drop probability |  |
| Number of haptic packets selected for weight calculation |  |
| Packet arriving rate of haptic | Kbps |
| Bandwidth | MHz |

The tolerance of haptic packet is set to 50 milliseconds as instructed. In comparison, the maximum allowed drop probability is fixed as 0.05 rather than 0.1 to compensate the requirement of jitter. As for the QoS priority level, it is determined by tries and errors since no exact data is provided. Due to the limitation of word count, relevant comparison is not attached in this article. Considering the high demands on delay and throughput, even the latest packet arrived should be with the chance to be transmitted in no time. As a consequence, the number of packets selected for weight calculation is set to the maximum possible value namely 50. It is assumed that the packet arriving rate of haptic data is 750 Kbps, a constant value estimated by the traffic characteristics. In the end, the bandwidth is doubled since the size of data awaiting to be transferred is twice than the conventional network.

Despite users in real cases are almost unlikely to be with haptic traffic and the conventional services at the same time, it is assumed to be true in the simulation to examine the effect of introducing haptic data on other services.

The discussion in this part will mainly focus on the performance on haptic traffic since the mechanisms of PD and M-LWDF and corresponding influences on the system performance have been covered in the last section.

## Overall Throughput

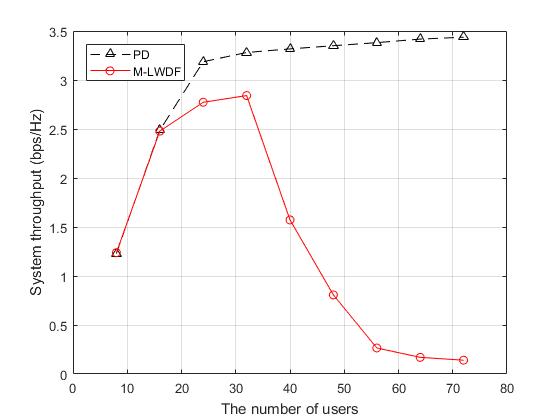


Figure illustrates the throughput at dB when the user number varies from 8 to 72. When the user amount is relatively small (), the resource is sufficient to transmit all incoming data without loss, thus the throughputs are the same. PD demonstrates its benefits over M-LWDF especially when the user number is large (). It can be seen that the bandwidth efficiency increases with the growth of user number, similar to the result in Section 6.1.1. In other words, PD is proved to be a feasible scheme for the multi-queue multiuser systems. In sharp contrast, the rise of user amount and queue quantity result in the increase of the weight gap between HoL packets, and therefore leads to the rapid decline when the number of queues is relatively large (). That is to say, the benefit that PD can gain from multiuser diversity is far more than that of M-LWDF. As a consequence, PD is proved more bandwidth efficient in multiuser multiservice system.

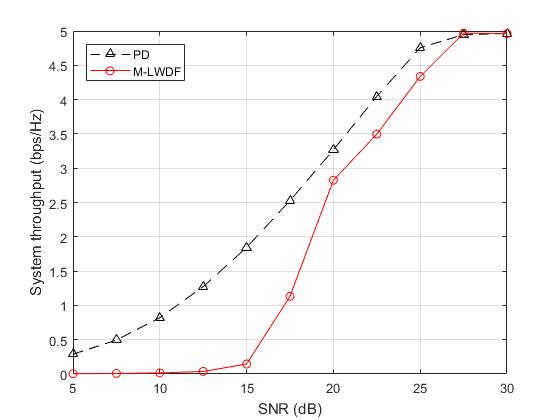


Figure verifies the conclusion from another dimension. According to it, the advantage of PD is especially obvious in low SNR cases. At dB, the bandwidth efficiency of M-LWDF still approaches zero, but for PD the value is almost 2 bps/Hz. With the improvement of SNR, the difference shrinks until the ideal case where both strategies can achieve maximum throughput.

## Delay

* + - 1. Haptic Packet Delay

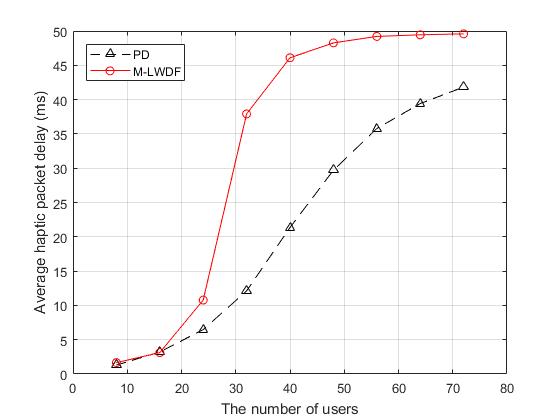
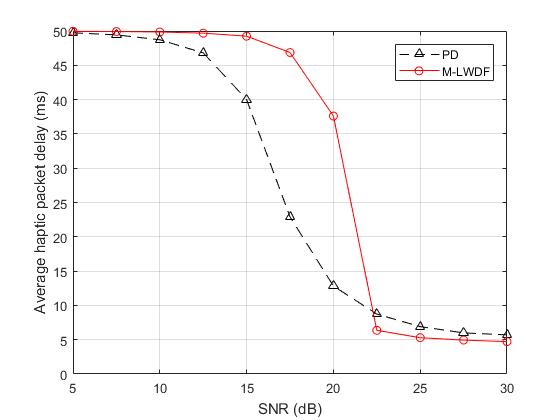
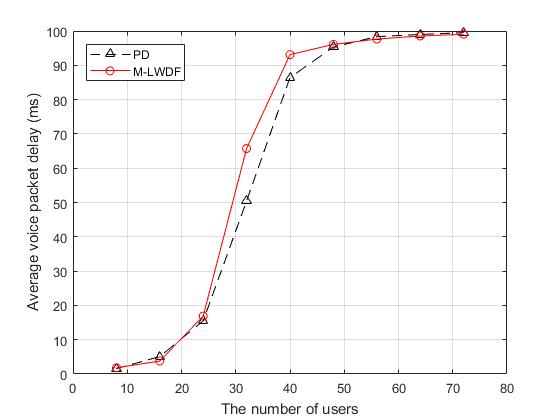


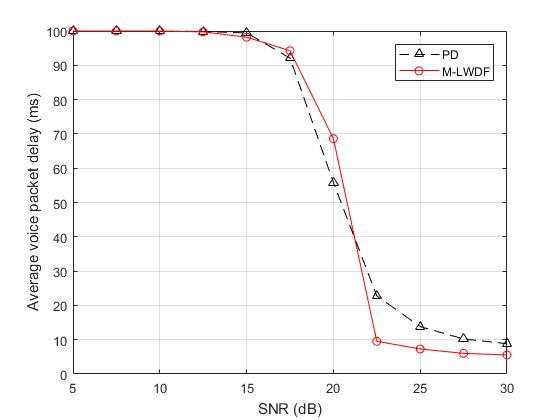
Figure shows the impact of user amount on the average haptic packet delay. As expected, PD accomplishes a lower delay than M-LWDF especially in the user range . For a typical 40-user network, the average haptic delay of PD is around 22 ms, which is only a half of that of M-LWDF.



It can be found that PD outperforms M-LWDF in a wide SNR range, from 5 dB to 20 dB. When the channel is in extremely ideal condition ( dB), M-LWDF result in a lower haptic delay. The reason is that when the resource is sufficient, the majority of the packets will not have long delay or large weight. Therefore, the system tends to transmit those packets with high priority level in precedence rather than caring too much about the delay. Thus, the queue-based scheme which serve haptic traffic in most times are making the right decision, although some bandwidth that should have been given to other services can be occupied by haptic data. As a result, the haptic service delay of M-LWDF can be shorter than that of PD.

* + - 1. Voice Packet Delay





It can be observed that the two schemes are performing similar in terms of voice traffic delay. For system with moderate number of users (), PD has around 10% lower voice delay than M-LWDF. The situation is little complicated for varied SNR. For dB where the resource is not enough, PD outperforms M-LWDF by its flexible packet-based scheduling strategy. While the system capacity is sufficient for dB, and M-LWDF wins with the reason discussed in the last section.

For PD, there is one thing interesting to mention which is the flexibility in QoS priority level determination. It is decided to be 32768 for haptic traffic as declared in the Table. The simulation results above demonstrate its rationality since the performance of haptic data transmission is enhanced significantly while the quality of other services remains acceptable with the performance close to M-LWDF. Nevertheless, the value can be adjusted for specific requirements. For instance, if the demand on haptic application is not so strict and the user want to achieve lower voice delay, then they just need to assign a lower priority level for haptic such as 16384 or less. That is to say, PD scheduling introduces sufficient flexibility while guarantees the bandwidth efficiency.

## Packet Drop Rate

* + - 1. Haptic Packet Drop Rate

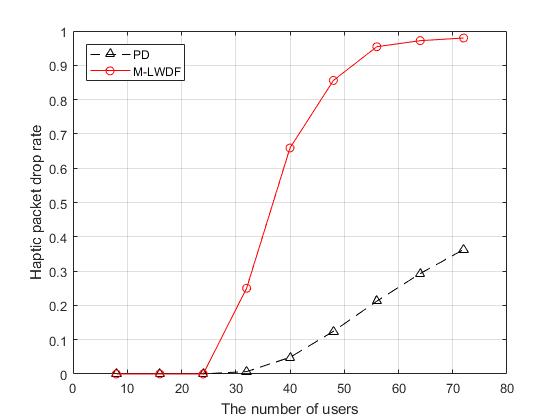


Figure suggests that although the haptic packet size is relatively large, PD is flexible enough to handle most of them even if the user amount is large. It can be observed from the figure that when the number of user exceeds a certain threshold () in this case, the accumulation effect discussed above has a great impact on M-LWDF. In other words, for multiuser multiservice system with heterogeneous traffic, the queue-based weight design of M-LWDF that rely on HoL packet result in large weight difference between queues, and thus low accuracy in choosing the right traffic to serve. When serving a certain queue, it cannot save the packets in other queues that are about to be dropped since the target cannot be changed in a whole slot. That is to say, the resource that could have save some dying packets are distributed to those not as urgent, owing to the superiority of HoL packets. In stark contrast, PD that calculate weight for packets can benefit more from multiuser diversity while avoid the accumulation effect. In a 72-user system, nearly all haptic packets are dropped with little net throughput for M-LWDF, but PD is running well with around 0.37 haptic traffic loss rate and a large bandwidth efficiecy.

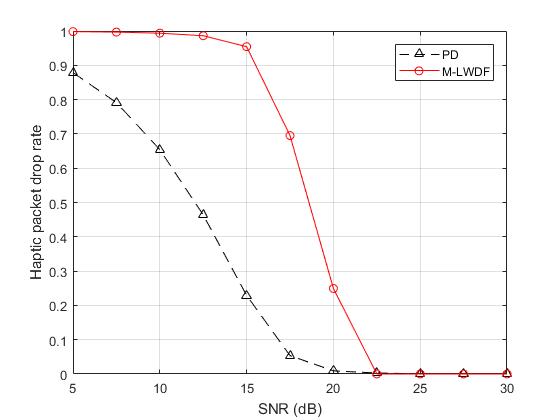
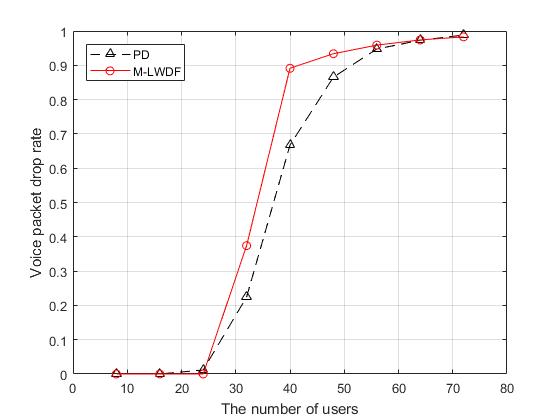
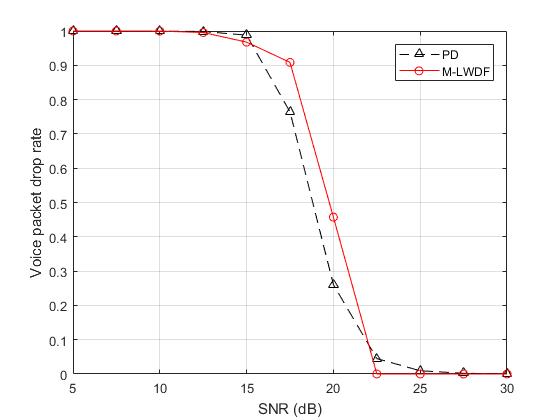


Figure verifies the advantage of PD in the haptic packet drop rate for a wide range of SNR. It can be seen that at dB, M-LWDF only transmits 5% haptic packets successfully while PD can deliever up to 78% without error. In order to achieve ideal transmission with zero drop rate, PD requires dB while M-LWDF demands 2.5 dB higher.

* + - 1. Voice Packet Drop Rate





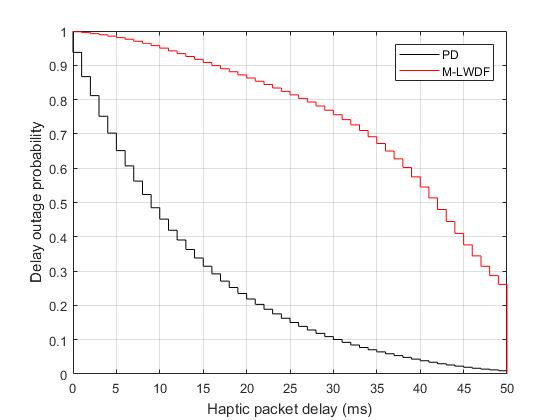
The voice packet drop rate can be estimated from the delay distribution. Figure a and b verify the result that although with little advantage, PD performs similar to M-LWDF in voice traffic loss rate, for the same reason discussed in the voice delay part (Section 6.2.2.2).

At dB, the two strategies can handle voice traffic almost without loss for user number less than 24. Then, even a small increase of user quantity can lead to sharp climb of voice packet loss rate. For a network with 56 users, the majority of bandwidth is reserved for stable haptic transmission, result in only 5% of voice packets delierved. In terms of SNR, for 32-user system, both schemes start to have valid voice transmission from dB that become ideal at 22.5 dB.

## Outage Probability

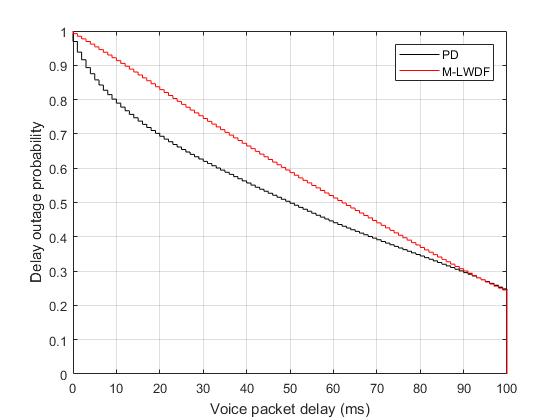
Similar to the previous case, it is assumed that the transmission happens at dB for a cross-layer network with 32 users.

* + - 1. Haptic Outage Probability



As indicated by Figure, for the provided circumstance, PD can almost send all haptic packets within the boundary while M-LWDF is able to deliver around 73% before the deadline. To transmit a half haptic data, PD only spends 9 milliseconds, which is only 21.4% of the time consumed by M-LWDF. For users that touch simulated object created by haptic applications, it is estimated that after 30 milliseconds waiting time, those in PD system are with 90% possibility to receive feedback while the chance for users employ M-LWDF scheduling is only 23%.

* + - 1. Voice Outage Probability



According to Figure, for the two schemes, there are still 25% voice packets are not delivered by the end of bound, which are to be dropped and removed from buffer. Compared with M-LWDF, those transmitted voice packets tend to be with shorter delay when utilizing PD scheduling. It takes 50 milliseconds for PD to send a half voice packet, and the time consumed by M-LWDF is 12 milliseconds (24%) more. Overall, the difference in terms of voice traffic is not as large as that of haptic traffic.

## Limitations, Comments and Future Works

Although the simulation result can reflect the real situation to some extent, there are at least three places that can be improved for more accurate outcomes. The first and most important one is the packet size distribution. In this project, the characteristics of voice, video and BE packets depend on the parameters in [12] and [13], which can be dated back to eight years ago. With the fast advancement of communication industry, services as high-resolution audio, 4K resolution video have updated the demands of public to relevant traffics. Therefore, the packet size design and weight balancing may need reconsidering when the cross-layer design is applied in specific cases. Second, it is assumed in the simulation that the haptic packet size is fixed as 750 bits. This is an estimated value based on the throughput and constant packet rate mentioned in [11], since until the article ends there are no relevant research about the haptic packet size distribution. That is to say the developer may need to estimate the packet size distribution for a specific haptic application and adjust the data generation function before utilizing this model for analysis. Third, fairness issue is not considered in this design. The resource allocation need extra care to balance the bandwidth assigned to users. Moreover, the QoS priority levels of different services can be adjusted to achieve specific requirements in various situations. Another decent idea is that to enable the discretion of QoS priority levels for users, which can be realized by predetermining fixed priority index sets. For instance, developers can create option for those favor low delay where the assigned priority index of traffics with high demand on waiting time are higher than standard.

1. Conclusion

This project focus on the cross-layer design for a multiuser multiservice OFDM system with heterogeneous downlink traffic. At the MAC layer, two scheduling schemes PD and M-LWDF indicate the importance of packets and queues by weight that associated with the traffic characteristics as delay, packet size and QoS priority level, which decides the packet transmission sequence. The weight information is passed to the PHY layer for subcarrier and power allocation that based on MWSC approach. It has been demonstrated that by combining the conventional algorithms with weight, the system can fulfil different requirements of various service, despite little decrease in overall capacity. Once the resource distribution is finished in this layer, relevant information about subchannel and power assignment will be transmitted back to the MAC layer for traffic control. As a typical type of new traffic, haptic data is introduced to the system with conventional services to check whether PD and M-LWDF are able to fulfil the diverse requirements of various services while maintain the overall performance.

The simulation results suggest that by choosing proper packets for weight calculation, PD demands a lower complexity than queue-based M-LWDF, but in most cases the performance in terms of throughput, delay and packet loss rate is better, especially for multiuser multiservice system when the resource is insufficient for lossless transmission. Overall, the objectives of learning wireless channel models, understanding basic network structure, studying OFDM system, and investigating packet transmission mechanism have been achieved.

[1] M. D. Pardue, "Fine-Tuning the OSI Model: Layer Functions and Services," in *MILCOM 1987 - IEEE Military Communications Conference - Crisis Communications: The Promise and Reality*, 1987, vol. 1, pp. 0199-0203.

[2] A. S. Tanenbaum and D. Wetherall, *Computer Networks*. Pearson Prentice Hall, 2011.

[3] W. Wang, D. Mosse, J. G. Pickel, and D. Cole, "Work-in-Progress: Cross-Layer Real-Time Scheduling for Wireless Control System," in *2017 IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)*, 2017, pp. 149-152.

[4] A. Alhosainy and T. Kunz, "Cross-layer design for IEEE 802.11 wireless ad-hoc network utility maximization with active queue management," in *2017 International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT)*, 2017, pp. 1-6.

[5] H. Hamouda, M. O. Kabaou, and M. S. Bouhlel, "A Cross-Layer Downlink Scheduling Scheme for Balancing QoS in IEEE 802.16 Broadband Wireless Access Systems," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, 2017, pp. 1-5.

[6] K. Lin and S. Dumitrescu, "Cross-Layer Resource Allocation for Scalable Video Over OFDMA Wireless Networks: Tradeoff Between Quality Fairness and Efficiency," *IEEE Transactions on Multimedia,* vol. 19, no. 7, pp. 1654-1669, 2017.

[7] F. Han, Y. Liu, and G. Zhu, "A Cross-layer Resource Allocation Scheme with Guaranteed QoS in Multiuser OFDM Systems," in *2006 International Conference on Communication Technology*, 2006, pp. 1-4.

[8] Z. Ying Jun and L. Khaled Ben, "Energy-efficient MAC-PHY resource management with guaranteed QoS in wireless OFDM networks," in *IEEE International Conference on Communications, 2005. ICC 2005. 2005*, 2005, vol. 5, pp. 3127-3131 Vol. 5.

[9] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine,* vol. 39, no. 2, pp. 150-154, 2001.

[10] S. Zukang, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Transactions on Wireless Communications,* vol. 4, no. 6, pp. 2726-2737, 2005.

[11] K. M. Yap, A. Marshall, and W. Yu, "Providing QoS for Multimodal System Traffic Flows in Distributed Haptic Virtual Environments," in *2007 First International Global Information Infrastructure Symposium*, 2007, pp. 153-158.

[12] N. Zhou, X. Zhu, Y. Huang, and H. Lin, "Low complexity cross-layer design with packet dependent scheduling for heterogeneous traffic in multiuser OFDM systems," *IEEE Transactions on Wireless Communications,* vol. 9, no. 6, pp. 1912-1923, 2010.

[13] Y. Qi, "Cross-layer Optimization for High Speed Wireless Communication Networks," MSc, Electrical Engineering & Electronics, University of Liverpool, Liverpool, 2011.

[14] T. Singh Dr. Sawtantar Singh Khurmi Amit Kumar Dr. Yunfei Liu, *IMT-Advanced: The ITU Standard For 4G Mobile Communication*. 2011.

[15] S. S. Dash, "Application of OFDM over indoor power line communication," in *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2017, pp. 240-245.

[16] E. Başar, A. Ü, E. Panayırcı, and H. V. Poor, "Orthogonal Frequency Division Multiplexing With Index Modulation," *IEEE Transactions on Signal Processing,* vol. 61, no. 22, pp. 5536-5549, 2013.

[17] M. Agiwal, A. Roy, and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," *IEEE Communications Surveys & Tutorials,* vol. 18, no. 3, pp. 1617-1655, 2016.

[18] S. Wu and M. Nekovee, "Single-cell and multi-cell performance analysis of OFDM index modulation," *IET Communications,* vol. 11, no. 7, pp. 1021-1027, 2017.

[19] E. Basar, "Index modulation techniques for 5G wireless networks," *IEEE Communications Magazine,* vol. 54, no. 7, pp. 168-175, 2016.

[20] X. Zhu, "Digital and Wireless Communications Lecture Notes," U. o. Liverpool, Ed., ed. Liverpool, 2018.

[21] J. D. Day and H. Zimmermann, "The OSI reference model," *Proceedings of the IEEE,* vol. 71, no. 12, pp. 1334-1340, 1983.

[22] J. Day, "The (un)revised OSI reference model," *SIGCOMM Comput. Commun. Rev.,* vol. 25, no. 5, pp. 39-55, 1995.

[23] D. Matic. (1999, 27/03/2018). *Mathematical description of OFDM*. Available: <http://www.wirelesscommunication.nl/reference/chaptr05/ofdm/ofdmmath.htm>

[24] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Communications, 1995. ICC '95 Seattle, 'Gateway to Globalization', 1995 IEEE International Conference on*, 1995, vol. 1, pp. 331-335 vol.1.

[25] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM wireless communications with MATLAB*. John Wiley & Sons, 2010.

[26] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[27] G. Kokkonis, K. Psannis, and M. Roumeliotis, "Real time Haptic data transferring," in *2016 Wireless Days (WD)*, 2016, pp. 1-3.

1. Please be aware that in this paper, to disambiguate, SNR refers to signal-to-noise ratio per bit, or *Eb/N0*. [↑](#footnote-ref-1)