CHALMERS

Performance evaluation of MI-ACM algorithm with different traffic models

- *Master of Science Thesis in the Master Degree Program*

- Ben Zhou

Department of Signals and Systems
Division of Communication Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden, 2011

Master Thesis EX003/2012

Performance evaluation of MI-ACM algorithm with different traffic models
Ben Zhou

Examiner: Associate Professor Tommy Svensson

Technical report no EX003/2012
Department of Signals & Systems
Chalmers University of Technology
SE-41296 Göteborg
Sweden
Telephone +46 (0) 31-772 1000

Abstract

Link adaptation is a method of improving the wireless communication system throughput over quasi-static fading channels. However, link adaptation requires accurate channel quality information (CQI) at the transmitter side, which is difficult to obtain. Within the European WINNER project, the MI-ACM (mutual information based adaptive coding and modulation) link adaptation scheme, which is near optimum and has lower complexity, was proposed. Previous work has shown that the MI-ACM scheme could achieve better performance than link adaptation methods that used in LTE, without exact suggestions on how to choose the code word length in the physical layer design. In this thesis, the performance of the MI-ACM algorithm is evaluated with 3 different codeword lengths in the physical layer, namely K=288, K=1152 and K=2304, and I evaluate the link adaptation schemes using realistic VoIP and Video source models which are two most common used traffic models and extremely representative. Simulation results show that a shorter codeword length could obtain better performance with respect to the delay time for real time services at relatively low SINR values.

Acknowledgements

I would like to express my gratitude to everyone that helped me during the thesis project.

Tommy Svensson, my thesis supervisor and examiner. Thank you very much for your valuable advice, thoughtful critics and patient guidance during my thesis work period.

Yutao Sui, thank you for offering me the opportunity to carrying out such an interesting project. Also thanks for coordinating and assistance during the whole process. Without your help I would have not been able to finish my thesis.

Yiwen Wei, thanks for being my opponent and valuable discussion that helped me to produce a better report.

To faculty members in Signals and Systems at Chalmers University of Technology, thank you very much for your valuable suggestions and helping me with various problems.

To friends, thank you for your consistent supporting and encouragements.

Finally, I am very grateful for my family's unshakable faith in me.

Ben Zhou
Göteborg, December 2011

Acronyms

MI-ACM	Mutual Information Based Adaptive Coding and Modulation
VoIP	Voice over Internet Protocol
CQI	Channel Quality Information
SINR	Signal to Interference and Noise Ratio
UMTS	Universal Mobile Telecommunications System
HSPA	High-speed Packet Access
LTE	Long Term Evolution
WLAN	Wireless Local Area Networks
QoS	Quality of Service
FEC	Forward Error Correction
LDPC	Low-Density Parity-Check codes
HARQ	Hybrid Automatic Repeat Request
CWER	Code Word Error Rate
RB	Resource Block
UE	User Equipment
CNR	Carrier to Noise Ratio
AWGN	Additive White Gaussian Noise
BICM	Bit-Interleaved Coded Modulation
ACM	Adaptive Coding and Modulation
CBR	Constant Bit Rate
VAD	Voice Activity Detection
VBR	Variable Bit Rate
MMBP	Markov-Modulated Bernoulli Process
CG	Concept Groups
LA	Local Area
MA	Metropolitan Area
WA	Wide Area
NLOS	Non Line-of-Sight
AoA	Angle of Arrival
AoD	Angle of Departure
MIMO	Multiple-Input Multiple-Output
LS	Large Scale
BS	Base Station
MS	Mobile Station
LOS	Line-of-Sight
PL	Path-loss
CDL	Clustered Delay Line
TDD	Time Division Duplex
FDD	Frequency Division Duplex
OFDM	Orthogonal Frequency Division Multiple

PF Proportional Fair

Table of Contents

1 Introduction

Link adaptation is the state-of-art in modern wireless systems with examples of UMTS HSPA (Universal Mobile Telecommunications System, High-speed packet access), 3GPP LTE (Long Term Evolution), WLAN(Wireless Local Area Networks), to improve the throughput. In particular, as actual wireless channels are time-varying and frequency-selective, link adaptation can achieve increased spectral efficiency and provide services of different quality with better utilization of the instantaneous capacity. Link adaptation, however, requires certain degree of accuracy of channel quality information (CQI), namely the signal to interference and noise ratio (SINR), at the transmitter side. This is a challenge in real systems due to channel estimation errors, feedback delays, etc. Thus, link adaptation performance is impeded by the imperfect CQI at the transmitter side when adapting the code rates and modulation schemes for the coming transmission slots [1].

A mutual-information-based adaptive coding and modulation (MI-ACM) method was chosen as the link adaptation candidate of the European project Wireless World Initiative New Radio (WINNER) [2] MI-ACM achieves near-optimal performance with low complexity and therefore attracts considerable attention. It has been shown that the MI-ACM algorithm can achieve better performance than link adaptation methods that are used in LTE based on some system level evaluation done in [3]-[4]. However, the problem of how to choose the code word length in the physical layer design was not answered exactly. Previous work developed in [5] also showed that MI-ACM can achieve near optimum spectral efficiency with acceptable complexity. In this algorithm, it is proposed to use 3 different coded block length in the physical layer, namely K = 288, K = 1152 and K = 2304. The longer the code word length the better the bit error rate performance is, but on the other hand less users can be accommodated. The switch among different code word length may depend on the system load, traffic model, the channel condition of the user, etc.

In many system-level investigations, a full buffer traffic model is utilized which is uncomplicated with clear purpose to limit complexity and runtime of simulations. However, full buffer simulations often provide unrealistic or even rather optimistic results, because neither the impact of the traffic model nor the impact of packet handling is considered. The impact of the traffic model might be decreased queue length, decreased multi-user diversity gains because of lower number of users, and decreased link adaptation accuracy because of fixed or limited packet size, etc.

In order to better understand the MI-ACM algorithm on link-level performance, a simulation campaign is carried out to evaluate link-level results with traffic models of VoIP application and Video streaming for different code word lengths. The traffic model for the VoIP application with constant packet size is chosen for the purpose of removing the effects of varying packet sizes and reading times. User packet delay is investigated for standard

proportional fair scheduling. Comparison is performed under QoS (Quality of Service) constraints modeled by user throughput and packet delay requirements with different code word lengths.

This thesis is organized as follows. In Section 2 the MI-ACM algorithm and the puncturing algorithm are illustrated. Then in Section 3 traffic modeling for VoIP and video applications used for the evaluation is described. Further, an illustration of the channel model used in the simulations is presented in Section 4. In Section 5, a description of the assumptions and models used in the link-level simulations is provided. In Section 6, a detailed discussion of the simulation results is given. Finally, Section 7 concludes and summarizes the major findings.

2 The MI-ACM Algorithm

It is widely recognized that adaptive allocation and individual link-adaptation of time and frequency resources based on channel quality and user requirements are beneficial to exploit the capacity of broadband frequency-selective channels [5]. The most recommended way is to perform link-adaptation and resource partitioning in the minimum unit of a rectangular area in time and frequency (chunk-wise), also to reduce the signaling overhead. The physical channel structure and the partitioning of sub-symbols into chunks are shown in Figure 1. The chunk, or the Resource Block (RB), size is selected according to the coherence time and bandwidth of the channel in order to achieve an essentially flat fading within a chunk. The payload bit capacity of a chunk is generally smaller than the codeword size required for the use of efficient FEC (Forward Error Correction) coding schemes like Low-Density Parity-Check codes (LDPC) or Turbo codes. This means that FEC coding has to be performed in the form of an “outer code”, whose codeword span several chunks [5].

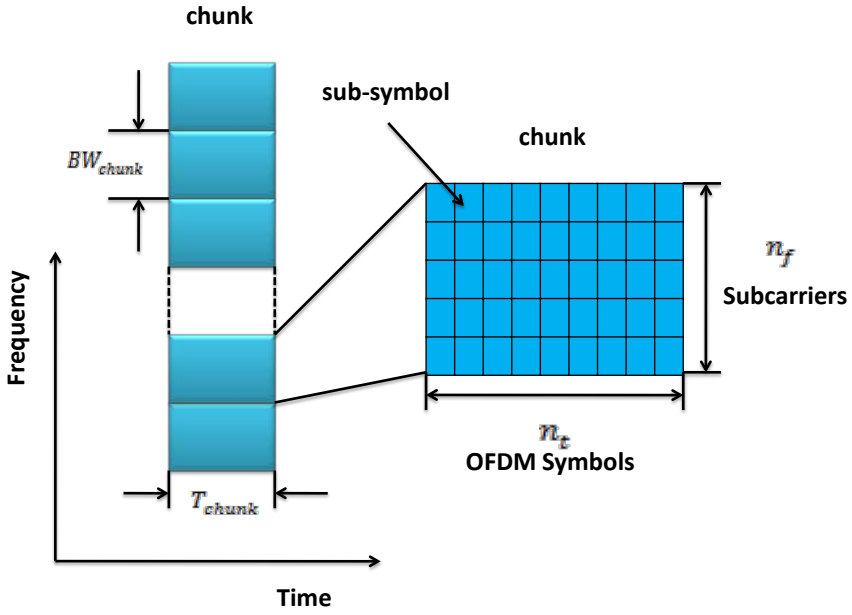


Figure 1. Physical channel structure and chunks

The combination of adaptive coding and chunk-wise link-adaptation become an open issue. The MI-ACM algorithm is built on using a mutual information based link quality metric to maximize the throughput of a type-I HARQ system with a target code word error rate (CWER) of 0.01 [1]. The MI-ACM algorithm adapts both code rate of the outer code and the modulation schemes of each RB. This makes it different from link adaptation schemes used in LTE that only outer code rate is adapted while all RBs assigned to the same UE use the same modulation scheme.

Assuming interference free conditions, the effective carrier to noise ratio (CNR) of each RB determines the modulation orders for the RB, given by

$$\mathrm{T_l} = \mathrm{bf}^{-1}\left(\frac{1}{\mathrm{n_f n_t}}\sum_{\mathrm{i=1}}^{\mathrm{n_f}}\sum_{\mathrm{t=1}}^{\mathrm{n_t}}\mathrm{f}\left(\frac{|\mathrm{H^l(i,t)^2}|}{\mathrm{N_0}}\right)\right) + (1-\mathrm{b})\min_{\mathrm{i,t}}\frac{|\mathrm{H^l(i,t)^2}|}{\mathrm{N_0}} \qquad (2\text{-}1)$$

Where 0<b<1, $\mathrm{f(x)} = \log_2(1+\mathrm{x})$, $\mathrm{n_f}$ and $\mathrm{n_t}$ are the number of sub-carriers in frequency and time domain of an RB, $\mathrm{H^l(i,t)}$ is the channel coefficient of the (i,t)th subcarrier in the *l*th RB, $\mathrm{N_0}$ is the noise power. The final CNR is a weighted sum of the average CNR and the CNR of the worst sub-carrier of that RB. According to the simulation results of [6], b = 0.6 was chosen for a good balance of the CNR value between the average and the worst corner of each RB.

After determining the modulation scheme for each RB, the next stage is to calculate the overall code rate R_{avg} based on the weighted average, given by

$$R_{avg} = \frac{\sum_{l=1}^{L} m(l) \cdot R_{local}(l)}{\sum_{l=1}^{L} m(l)} \quad (2\text{-}2)$$

where *m(l)* denotes the modulation format (i.e. the number of bits per symbol) and $R_{local}(l)$ is the local code rate of chunk *l* respectively. Both of them are obtained from the short-term chunk SINR using looking-up tables based on AWGN channel performance figures. For details on the algorithm and thresholds of choosing local rate for each RB, readers are referred to [5] and [7].

The MI-ACM algorithm can be summarized as the following steps [5]:
A set of modulation formats $\mathcal{M}$ are defined by the system and the number of bits per symbol of the *i*-th format is denoted as $\mathcal{M}(i), i = 1, \ldots |\mathcal{M}|$. The user applies the FEC scheme with (fixed) information length *K*. A set of *L* chunks would be arranged according to their CNR and $T_1 \leq T_2 \leq \cdots \leq T_L$ is assigned by the scheduler. *P* denotes the overall signal power.

1) Initialization:
   For all *l* there is power for each chunk p_l

$$p_l = P/L \quad (2\text{-}3)$$

2) For each chunk *l=1 , … , L*:
   i) Determine $m(l) \in \mathcal{M}$ with Table 1:

$$m(l) = \mathcal{M}(i) \text{ with } i = \arg\max_i \{\gamma_i < p_l T_l\} \quad (2\text{-}4)$$

   If $p_l T_l < \gamma_1$, chunk *l* is excluded from the transmission and the signal power is redistributed:

$$m(l) = 0, p_l = 0 \quad (2\text{-}5)$$

$$p_i = P/(L - l) \text{ for } i = l{+}1 \ \ldots \ L \quad (2\text{-}6)$$

   ii) Calculate the virtual local rate $R_{local}(l)$ for modulation *m(l)* using Table 2 together with linear interpolation.
3) Calculate the required code rate R_{all} as weighted average of the local rates and the number of code bits N_{CB}:

$$R_{all} = \frac{1}{\sum_{l=1}^{L} m(l)} \cdot \left[\sum_{l=1}^{L} m(l) \cdot R_{local}(l)\right] \quad (2\text{-}7)$$

$$R_{local}(l) = \check{R}\left(\check{I}_c(l)\right) = \check{R}\left(\frac{I_{m(l)}(p_l T_l)}{m(l)}\right) \quad (2\text{-}8)$$

Where $I_{m(l)}(p_l T_l)$ is the mutual information per symbol for the modulation format with m(l) bits per symbol on a bit-interleaved AWBG channel at SNR = $p_l T_l$.

$$N_{CB} = N_S \cdot \sum_{l=1}^{L} m(l) \tag{2-9}$$

Where N_S is the number of sub-symbols in a chunk.

4) Determine the number of codeword N_{CW} to be transmitted on these chunks as

$$N_{CW} = \left\lfloor \frac{N_{CB} \cdot R_{all}}{K} \right\rfloor \tag{2-10}$$

5) The actually used punctured codeword length *N* is calculated as

$$N = \left\lfloor \frac{N_{CB}}{N_{CW}} \right\rfloor \tag{2-11}$$

6) Finally, the N_{CW} codeword is interleaved jointly and loaded onto the chunks (BICM).

Table 1. Required SNR for a set of code rates for target CWER=0.01

		Code Rate R	24/48	24/44	24/40	24/36	24/32	24/30	24/28	24/26
SNR [dB]	K = 288	BPSK	-0.92	-0.27	0.50	1.34	2.47	3.25	4.37	6.41
		QPSK	2.09	2.74	3.51	4.35	5.48	6.26	7.38	9.42
		16QAM	7.56	8.36	9.27	10.40	11.06	12.70	13.90	16.18
		64QAM	12.08	13.13	14.26	15.59	17.21	18.38	19.71	22.07
	K = 1152	BPSK	-1.44	-0.85	-0.11	0.76	1.94	2.75	3.71	5.84
		QPSK	1.57	2.16	2.90	3.77	4.95	5.76	6.72	8.85
		16QAM	6.92	7.75	8.59	9.66	10.43	12.02	13.21	15.46
		64QAM	11.42	12.37	13.49	14.79	16.50	17.55	18.88	21.38

Table 2. SNR limits γ_i for the choice of the modulation format of a chunk

Modulation format		**BPSK**	**QPSK**	**16QAM**	**64QAM**
i		1	2	3	4
m [bits per symbol]		1	2	4	6
γ_i **[dB]**	**K = 288**	-0.92	2.09	7.57	13.22
	K = 1152	-1.44	1.57	6.92	12.32

This framework of MI-ACM algorithm is only one example explaining the principle of the

ACM adaptation. Loading each codeword separately onto chunks is also feasible. Given that the bandwidth assigned to a user flow is small and it is necessary to wrap the long packets over scheduling frames. Assume the scheduler would transmit one codeword and has X_t chunks to assign in the current frame *t*. The required code rate R_{all} and the number of code bits can be obtained from equation (2-7) and (2-9). However, the code rate R_{all} might not be fulfilled on these chunks, which means equation (2-10) is zero. But yet the first $N=N_{CB}$ bits of the codeword can be sent anyway in this frame according to their priority. In the next new frames, the required code rate R_{all} is recalculated taking both the X_t chunks of last frame and the X_{t+1} chunks of the current frame. The number of chunks that the codeword is loaded onto is increased until equation (2-10) is one.

3 Traffic Model

Besides the full buffer case, simulations with realistic traffic models shall be provided for assessment of the LTE performance under typical traffic and load conditions. Emphasis is put on VoIP and Video applications due to the time constrains. Voice and Video applications are the most commonly used and representative application types. In addition, traffic model for video application can be extended and applied to video conference area.

3.1 Traffic models for VoIP Applications

The conversational model based VoIP traffic proposed in [8] is used as traffic models for VoIP applications in this work. There are two previous existing methods for VoIP traffic generation. One method is to use CBR (Constant Bit Rate) traffic as simulated VoIP traffic because some popular voice codec generate constant sized frames at a constant interval. However, the actual VoIP traffic is not simply CBR but depends on the voice signal power that VAD (Voice Activity Detection) technique detects therefore the second method adopted VAD technique in which way codec only code voice signal with the power over the threshold that VAD technique recognizes as the lowest human voice level. But the problems of VAD adoption are the number and size of voice source files limitation and the lack of interactive process between two speakers in a conversation. To conclude, existing VoIP traffic generation methods lack reality and interactivity, so they can not simulate real VoIP products' traffic correctly [8].

The conversational models indicate the influence of conversational users' speech characteristics and interactive process to VoIP traffic. Therefore, it provides more realistic and interactive traffic for practical quality measurement activities according to specific voice codec and VAD (Voice Activity Detection) parameters. In order to achieve a more effective

VoIP traffic generation algorithm, the traffic characteristics of VoIP applications are needed to be analyzed. VoIP applications are affected by users' behavior, so it is necessary to take state change of user's talking and silent duration into consideration. Furthermore, VoIP traffic generation should consider the effect of interactive relationship between two users'. According to these two requirements a conversational model to simulate two speakers' conversational state change and interactive process was implemented based on [8]. It considers the influence of user behavior, VAD technique and voice codec to VoIP traffic.

3.1.1 Framework of generation algorithm

As shown in the framework in Figure 2, after setting up all input parameters, first we construct the conversational model and calculate transition probabilities to indicate two users' behavior and their interactive process. Then VAD technique is adopted and the time points to generate packets are determined based on the simulation result of the model. Finally VoIP packet size at every determined time point is decided and VoIP traffic traces are generated according to the specific codec's attributes.

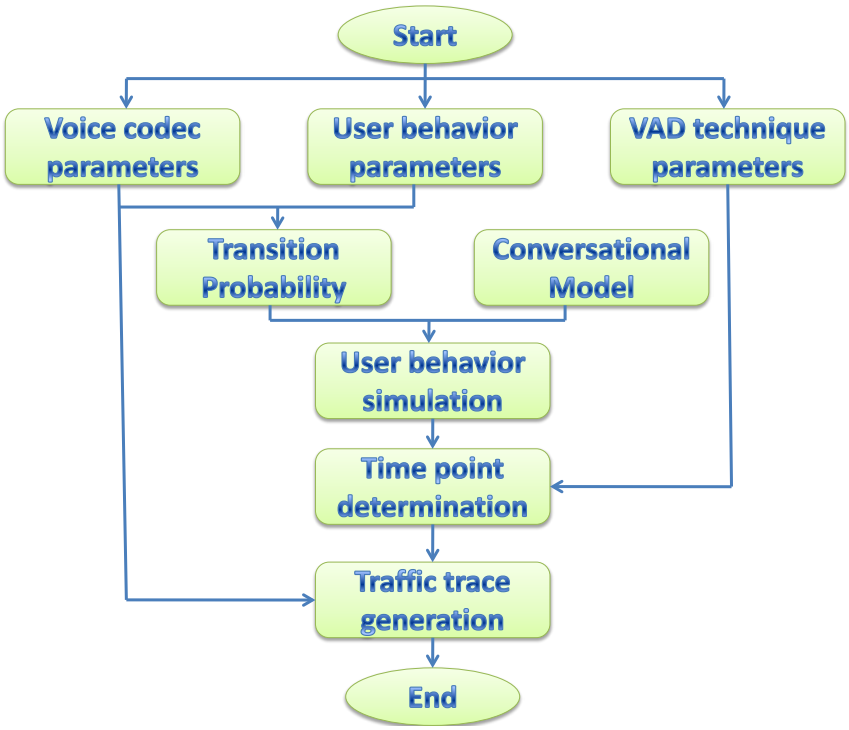


Figure 2. Framework for VoIP traffic generation algorithm

3.1.2 Model Construction

The conversational model is a discrete-time 4-state-transition model which can be seen in Figure 3. Its 4 states represent user A's talking (state 1), user B's talking (state 2), double

talking (state 3) and mutual silence (state 4). VoIP packets are generated discretely, so transition of states happens at discrete time points.

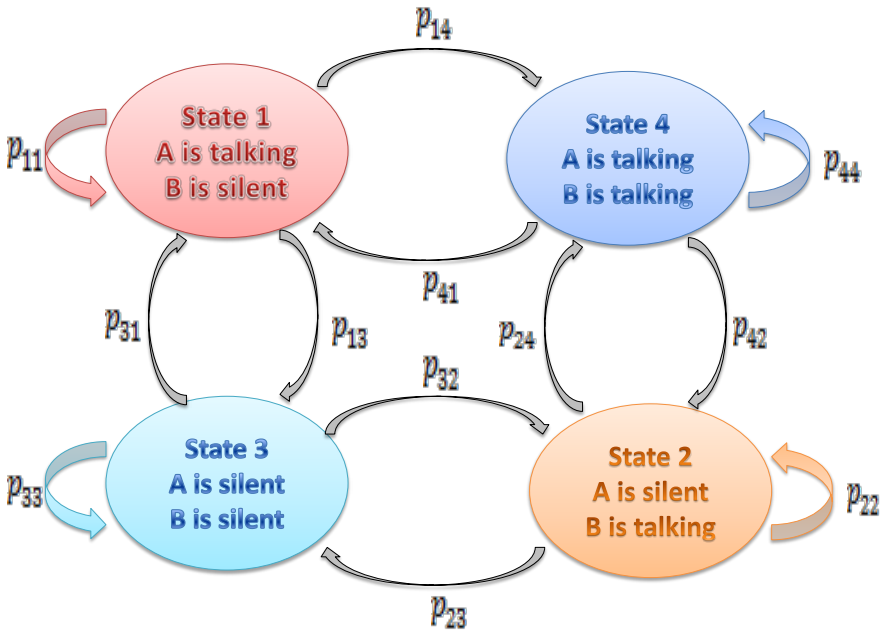


Figure 3. Discrete-time conversational model

Transition probability is the key factor of the whole conversational model. As in the conversational model of [8], state transition happens at discrete time points based on transition probability that the state *i* remains for consecutive *k* time points before transiting to another state at *k+1* time point follows a geometric distribution:

$$P_i(k) = p_{ii}^{k-1}(1 - p_{ii})$$

The duration of user behavior in discrete-time model is simulated based on the geometric distribution and the expectation of *k* is given by:

$$E_i(k) = 1/(1 - p_{ii})$$

This value should be proportional to the average duration on the state *i* from user behavior parameters that are gathered by experiments from real human conversations. According to these parameters, the transition probability for every state can be obtained by the following equations:

$$p_{11} = p_{22} = 1 - T_{interval}/T_{talk-avg} \quad (1)$$

$$p_{33} = 1 - T_{interval}/T_{stop-avg} \quad (2)$$

$$p_{44} = 1 - T_{interval}/T_{double-avg} \quad (3)$$

$T_{interval}$ represents the duration between two consecutive discrete time points and its configuration should refer to the specific voice codec. Here we choose the recommended ITU-T experiential values in the measurements, $T_{interval}$ can be configured using default value as 20 milliseconds. $T_{talk-avg}$, $T_{stop-avg}$ and $T_{double-avg}$ respectively represent the average duration of state 1 (same as state 2), state 3 and state 4. ITU-T provides experiential values of 1.004 second, 0.508 second and 0.228 second as the parameters for these three average durations which are average results from conversations in English, Italian and Japanese.

After calculation of loopback transition probabilities, it is time to consider the transition probabilities between different states. Since the probabilities should satisfy the following equation:

$$\sum_{j} p_{ij} = 1$$

Transitions from state 3 to state 1 and state 2 are symmetric, there are:

$$p_{31} = p_{32} = \ （1 - p_{33})/2 \quad (4).$$

The relation for state 4 to state 1 and state 2 is the same:

$$p_{41} = p_{42} = \ （1 - p_{44})/2 \quad (5).$$

From the symmetry of the conversational model, it is clear that the transition probability from state 1 to state 3 equals to the one from state 2 to state 3, and transition probability from state 1 to state 4 equals to the one from state 2 to state 4. Here $K_{talk-double}$ is used to represent the probability that state 1 (or state 2) transmits to state 4 in the condition that state change happens. Then the equations are:

$$p_{14} = p_{24} = (1 - p_{11}) \times K_{talk-double} \quad (6).$$

$$p_{13} = p_{23} = (1 - p_{11}) \times (1 - K_{talk-double}) \quad (7).$$

$K_{talk-double}$ is also from user behavior parameters gathered from statistics on human conversations. ITU-T suggested this parameter should be configured as 0.6 based on experiential value.

All transition probabilities required in the conversational model can be calculated by

equations (1) – (7) with which two users' behavior in a conversation can be modeled. Table 3 shows the suggested configuration of all the parameters used in this model based on the ITU-T suggested experiential values.

Table 3. Recommended parameters values from ITU-T

$T_{interval}$	$T_{talk-avg}$	$T_{stop-avg}$	$T_{double-avg}$	$K_{talk-double}$
20ms	**1004ms**	**508ms**	**228ms**	**0.6**

User's conversational interactive process was simulated by our model and is shown in Figure 4.

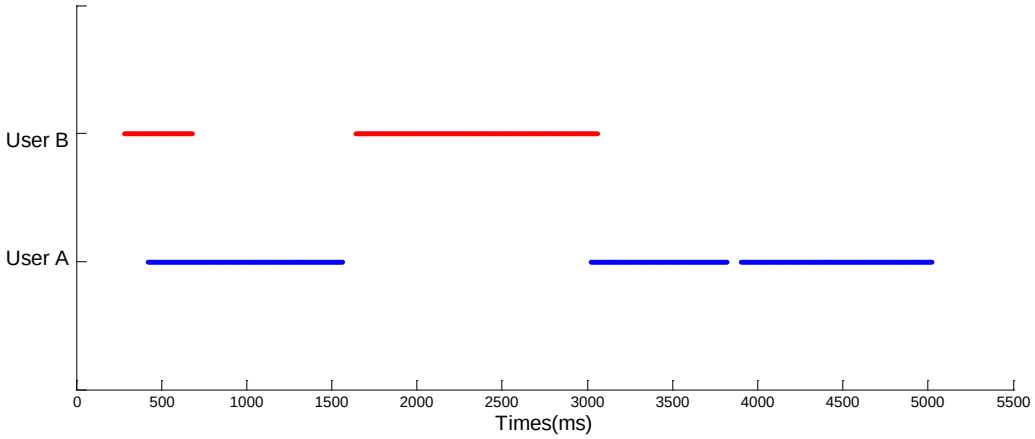


Figure 4. Simulation result by discrete-time model

Figure 4 illustrates the interactive process between two users' and each user's silent and talking duration. The marked node at each time point means that the user is talking at that specific time.

3.1.3 VAD Hangover time operation

VAD technique adopted in VoIP has a significant impact on configuration of hangover time. The concept of hangover time according to VAD is the period during which voice energy is under the threshold that VAD recognizes as the lowest human voice level. VoIP traffic is generated when voice energy is either above or under the threshold. The reason why VoIP traffic does not stop just at the hangover time starting moment is due to the existence of background sound in a VoIP call. If the traffic immediately stopped, the user on the other side would feel the abrupt disappear of background sound which would bring uncomfortable feeling and VoIP quality damage. Therefore, it is necessary for the traffic to last some time after a user stops talking. Figure 5 shows the basic function of hangover time with X-axis representing time and Y-axis representing voice energy that VAD detects [8]. The duration from T0 to T1 is the period that the voice energy is higher than the lowest human voice level (threshold) that VAD detected. Therefore, VoIP traffic is generated during this period. And the

duration from T1 to T2, in which the voice power dropped below the threshold, is called hangover time and VoIP traffic is generated as well for the consideration of VoIP quality.

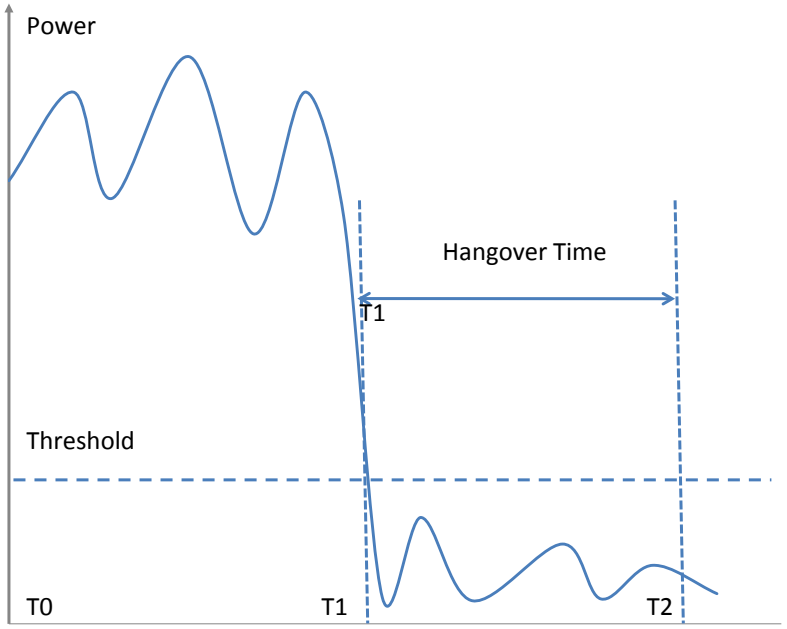


Figure 5. Usage of VAD hangover time

After the influence of VAD hangover time is added to the conversational model, the refreshed result is shown in Figure 6 and hangover time is set to be 200 milliseconds in this work. When transmitting from state 1 to state 3, or from state 4 to state 2, user A should send packets in configured hangover time; when transmitting from state 2 to state 3, or from state 4 to state 1, users B should send packets in configured hangover time. The time points marked by VAD hangover time are set to be a little higher than those marked by the conversational model. More traffic needs to be generated than as more time points are marked.

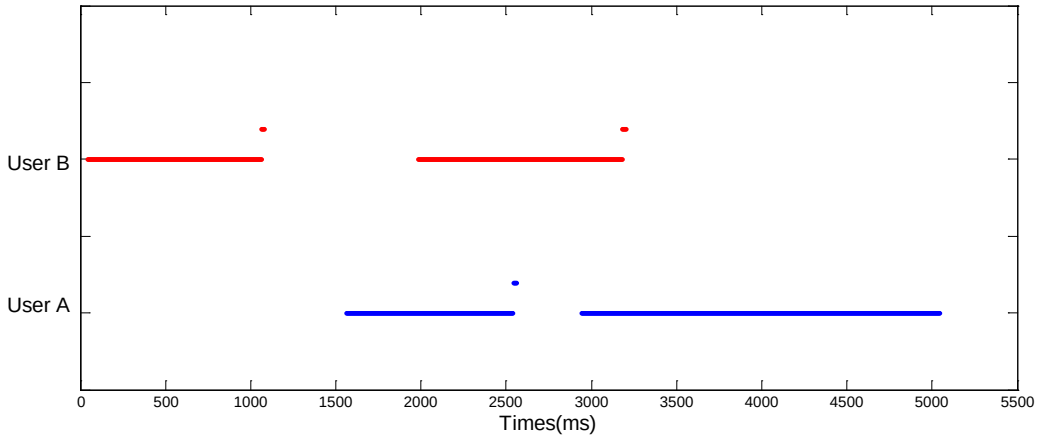


Figure 6. Simulation result with VAD hangover time

3.1.4 Traffic Trace Generation

The conversational model and hangover time operation decide whether a VoIP packet needs to be generated at each time point. Afterwards, the construction of payload format, composed by payload content and payload size, needs to be considered for each packet. The payload content can be filled by random data in measurement as routers and switches on the network do not change their operation to VoIP packets. The payload size is determined by the type of voice codec and interval between time points. It is calculated by the following equation:

$$S_{payload} = \frac{R_{codec} \times 10^3}{8} \times T_{interval}$$

$S_{payload}$ represents the calculated payload size (in bytes) and R_{codec} represents the voice codec sending rate (in Kbps) while $T_{interval}$ is the interval between time points (in seconds). The payload size could be obtained according to specific voice codec. For example, G.711 codec provides R_{codec} as 64 Kbps and $T_{interval}$ as 20 ms, therefore the payload size should be 160 bytes according to the equation. If G.729 codec (8 Kbps for R_{codec} and same time interval as G.711) is selected, the payload size should be 20 bytes.

For the convenience of measurement work, the traffic information was stored into two trace files which represent the two users' behaviors in the conversation. Table 4 indicates a segment of trace file. It contains information on packet sequence number, packet sending time and packet payload size based on which a measurement node can generate proper sized packet at proper time to simulate the VoIP application traffic model.

Table 4 Segment of a traffic trace file

Index	Sending time(ms)	Payload size(bytes)
200	4260	160
201	4280	160
202	5100	160
203	5140	160
…	…	…

3.2 Traffic Models for Video Streaming

3G wireless networks subscribers can enjoy the advantages of videophone, compressed images, data and other multimedia services besides the traditional voice service. Video based entertainment and service applications are targeted to be the driving applications for the broadband network.

Several models for VBR (Variable Bit Rate) video traffic have been proposed. A discrete-time two-state Markov chain adopted to model a multimedia traffic source is a special case of the MMBP (Markov-Modulated Bernoulli Process) which has been proposed in [11]. It is also been called the high-low (H/L) model having two nonzero-valued states corresponding to the relatively high and low activities of the multimedia traffic source. A queue-length distribution using the generating function together with matrix geometric approach was derived instead of analyzing the queue-length distribution using a probability generating function which leads to no closed-form solution. The queue length can be evaluated by decomposing matrices into spectral representation forms and considering the roots of its characteristic function. Maglaris et al [12] have proposed a model for the coding bit rate of a single video source using interframe predictive coding. Sen et al [13] proposed models for different activity levels using correlated Markov models and queuing analysis to estimate the packet loss and delay. Yegenoglu et al [14] proposed a model for VBR video using a time dependent AR (Autoregressive) model to represent data from different activity levels. However, most of the works did not consider the differences between I (Intra) and P (Predictive) frames which are that, an I frame is coded in isolation from other frames using transform coding, quantization and entropy coding while a P frame is predictively coded that a prediction is formed using a previously coded frame and only the difference between the prediction and the actual frame is coded. In this thesis work we use a traffic model for video streaming proposed by [10] which was illustrated in more details in the *Model Construction* Section.

3.2.1 Characterization of Video Sources

The performance, prognosis and evaluation of multimedia communication systems by means of traffic load modeling are important aspects in the desire to understand how the system performance, architecture and configuration will be influenced by such integration of services and the new requirements of multimedia applications [15]. The traffic load of a communication system depends closely on several attributes, such as, application environment, type of request, time interval during which load generation is observed, and chosen interfaces within the communication system. Video traffic is expected to be one of the major sources of loading of multimedia networks due to its role as main constituent of multimedia applications. Video communication requires a rather high transmission capacity due to an efficient video coding scheme that can achieve acceptable grade of quality for all images produces VBR output inherently. However, VBR video traffic streams have a complex structure as well as rather stringent real-time constraints on the networks for their transmission. Therefore the development of such source models becomes more significant for performance analyses and multimedia systems predictions.

The popular standards defining compression schemes today are ISO MPEG series and ITU H.26x series with MPEG-4 and H.263 being some of the latest versions [16]. These standards allow for three different kinds of coding schemes for a video frame, which might be Intra (***I***),

Predictive (***P***) or Bidirectionally-Predictive (***B***), in order to improve coding efficiency. An ***I*** frame is coded isolated from other frames using transform coding, quantization and entropy coding. A ***P*** frame is predictively coded which means that a prediction is formed according to the previously coded frame and only the difference between the prediction and the actual frame is coded. A ***B*** frame is predicted bidirectionally meaning that the prediction is formed according to both the previous frame and the successive frame. An ***I*** frame is often adopted to code frames efficiently corresponding to scene changes which means that successive frames are different and cannot be easily predicted. ***P*** and ***B*** frames can be used to the frames within a scene which are similar to the previous frames, and hence can be coded predictively for increased efficiency. The maximum frame sizes almost always appear with ***I***-frames while ***B***-frames are always smaller than ***I***- and ***P***-frames because ***I***-frames are coded without considering the dependency on each other frames. The fluctuation of frame size depends on scene variation in compression pattern pure intra-frames while the frame sizes in the sequences with ***P***- and ***B***-frames fluctuate to a greater extent. ***P***-frames are coded more efficiently using motion compensated prediction from a past ***I***- or ***P***-frame. The compression of ***B***-frames is similar to that of ***P***-frames where the only difference is that both past and next successive ***I***- or ***P***-frame are used as reference of motion compensation and that is also the reason that B-frames compression has the highest compression ratio.

Realistic traffic load modeling which could closely reflect the characteristic of MPEG video sources for performance evaluation of multimedia systems is the significant goal. For this purpose, a traffic load modeling from load measurements at an interface close to end users was expected. Traffic models describing the frame size generation process need first to select an appropriate distribution function that can characterize the important measures of the traffic loads. The main features of MPEG video sources are the generation of a VBR sequence (to keep the quality of the video output approximately continuous), consisting of three types of frames (***I, P*** and ***B***) and bursty compressed VBR video (due to the size of compressed frames varies from one frame to the next). Each frame to be transmitted can be treated as a single request loading the communication system. Since the interarrival time of requests is constant and therefore the most significant load attribute for modeling of MPEG video traffic is frame size which determines the amount of required system resources. A traffic load model generating frame size was proposed in LTE physical layer framework from 3GPP [10] which is approved and adopted comprehensively. In this work, this traffic load model for video application was utilized.

3.2.2 Model Construction

The traffic model for video streaming used in this work was proposed by [10]. Each frame of video data arrives at a regular time interval T determined by the numbers of frames per second. Each frame is decomposed into a fixed number of slices which are transmitted one after another as packets. The size of packet/slice is modeled to have a truncated Pareto distribution.

The video encoder introduces encoding delay intervals between the packets of a frame. These intervals are modeled by a truncated Pareto distribution as well. Table 5 illustrates statistical characterization of each parameter used in the traffic model for video application and the distributions used in the model assume a source video rate of 64 Kbps.

Table 5: Video Streaming Traffic Parameters.

Parameter	Statistical Characterization
Inter-Arrival time between the beginning of each frame	Deterministic 100 ms (based on 10 frames per second)
Number of packets (slices) in a frame	Deterministic 8 packets per frame
Packet (slice) size	Truncated Pareto Distribution Mean = 100 Bytes, Maximum = 250 Bytes (Before Truncation) PDF: $f_x = \frac{\alpha_k^\alpha}{\alpha+1}, k \le x < m$ $f_x = \left(\frac{k}{m}\right)^\alpha, x = m \qquad \alpha = 1.2, k = 53\ Bytes, m = 250\ Bytes$
Inter-Arrival time between packets (slices) in a frame	Truncated Pareto Distribution Mean = m = 6 ms, Maximum = 12.5 ms (Before Truncation) PDF: $f_x = \frac{\alpha_k^\alpha}{\alpha+1}, k \le x < m$ $f_x = \left(\frac{k}{m}\right)^\alpha, x = m \qquad \alpha = 1.2, k = 2.5\ ms, m = 12.5\ ms$

The traffic model for video application was generated according to the above parameter characterization and the feature of packet size along time was shown in Figure 7. Therefore, a certain size of packet can be generated at a proper time.

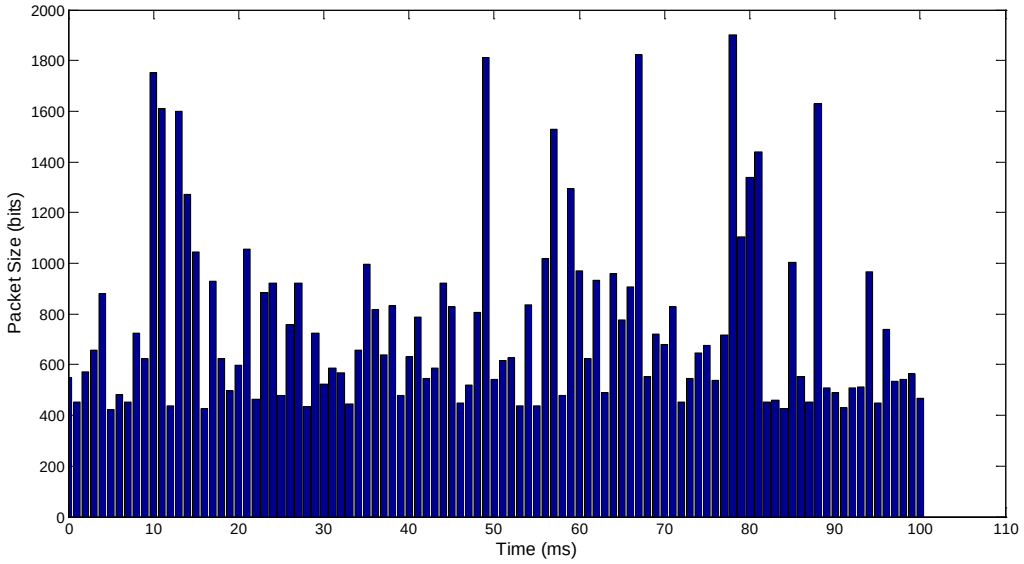


Figure 7. Simulation result of traffic model for video

3.3 Traffic model mapping

In the simulation of VoIP traffic model, the system would check if there is a packet coming or not for each TTI (20 ms). If there is a packet, then a certain sized payload of 160 bytes would be generated and thereafter transmitted. In the simulation of Video traffic model, each video frame is decomposed into 8 packets which are transmitted one after another. For the first TTI (1 ms), the size of packet is modeled under Pareto distribution which is also used for the model of delay intervals introduced by video encoder. Thereafter, the second packet comes and the same operation repeats. Obviously, the packet size is variable. If the packet size is less than the selected codeword length, then zeros need to be padded. If the packet size is bigger than the codeword length and the resource blocks assigned to this user is enough to complete the transmission, the packet would be transmitted as an integer and the remaining resource blocks would be reassigned to new users. But if the resource blocks assigned to the user is not enough to transmit the whole packet, then the packet needs to be divided into several parts and thus delay would appear. In this case, the delay could be long since when the first part of the packet arrives it has to wait until the last part arrives to make the whole packet completed.

4 Channel Model

4.1 Introduction and channel modeling approach

The propagation scenarios modeled in WINNER are shown in Table 6 and have been specified according to the requirements agreed commonly in WINNER project. Work has been divided in WINNER II between CG (Concept Groups) according to the environment they are working at. There were LA (Local Area), MA (Metropolitan Area) and WA (Wide Area) as can be seen in Table 6 in column CG.

Table 6. Propagation Scenarios in WINNER

Scenario	Definition	LOS/NLOS	Mob. Km/h	Frequency(GHz)	CG
A1 In building	Indoor office/residential	LOS/NLOS	0-5	2-6	LA
A2	Indoor to outdoor	NLOS	0-5	2-6	LA
B1 Hotspot	Typical urban micro-cell	LOS NLOS	0-70	2-6	LA MA
B2	Bad Urban micro-cell	NLOS	0-70	2-6	MA
B3 Hotspot	Large indoor hall	LOS/NLOS	0-5	2-6	LA
B4	Outdoor to indoor micro-cell	NLOS	0-5	2-6	MA
B5a Hotspot Metropol	LOS stat. feeder, rooftop to rooftop	LOS	0	2-6	MA
B5b Hotspot Metropol	LOS stat. feeder, street-level to street-level	LOS	0	2-6	MA
B5c Hotspot Metropol	LOS stat. feeder, below-rooftop to street-level	LOS	0	2-6	MA
B5d Hotspot Metropol	NLOS stat. feeder, above rooftop to street-level	NLOS	0	2-6	MA
B5f	Feeder link BS->FRS. Approximately RT to RT level	LOS/OLOS /NLOS	0	2-6	WA
C1 Metropol	Suburban	LOS/NLOS	0-120	2-6	WA
C2 Metropol	Typical urban macro-cell	LOS/NLOS	0-120	2-6	MA WA
C3	Bad Urban macro-cell	NLOS	0-70	2-6	-

C4	Outdoor to indoor macro-cell	NLOS	0-5	2-6	MA
D1 Rural	Rural macro-cell	LOS/NLOS	0-200	2-6	WA
D2	a) Moving networks: BS-MRS, rural	LOS	0-350	2-6	WA
	b) Moving networks: MRS-MS, rural	LOS/OLOS /NLOS	0-5	2-6	LA

Typical urban micro-cell with NLOS (Non Line-of-Sight) WINNER channel model (B1) is used in this work. WINNER channel model is geometry based stochastic model which enables separation of propagation parameters and antennas. The channel parameters for each snapshot are decided stochastically based on statistical distributions extracted from channel measurement. Antenna geometries and field patterns are determined by the users of the model. Channel generation is realized with geometrical principle by summing contributions of rays (plane waves) with specific small scale parameters like delay, power. AoA (Angle of Arrival) and AoD (Angle of Departure). Superposition leads to correlation between antenna elements and temporal fading with geometry dependent Doppler spectrum [9]. In the channel model, the cluster, a number of rays, is equivalent to a propagation path diffused in space, either or both in delay and angle domains. Elements of the MIMO (Multiple-Input Multiple-Output) channel (antenna arrays at both link terminals and propagation paths) are show in Figure 8.

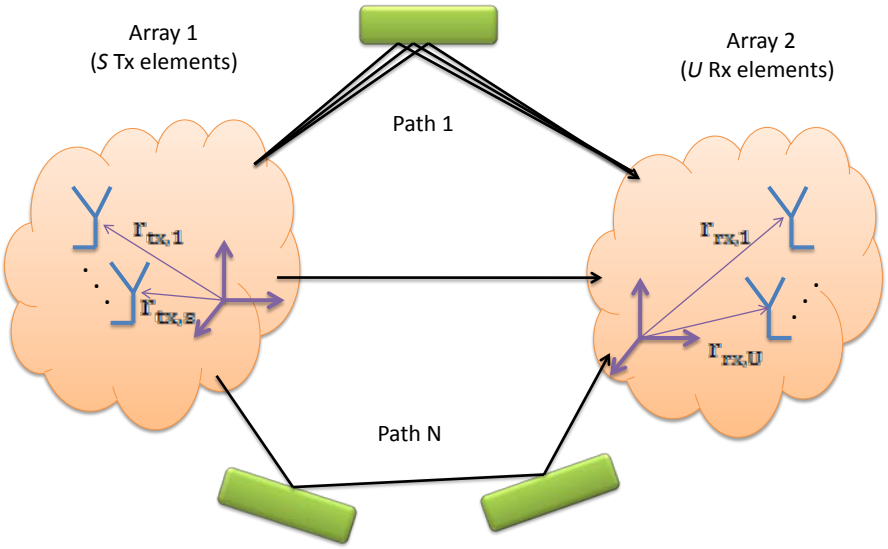


Figure 8. The MIMO channel

Transfer matrix of the MIMO channel is

$$\mathrm{H}(\mathrm{t};\tau) = \sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{H}_{\mathrm{n}}(\mathrm{t};\tau)$$

It is composed of antenna array response matrices $\boldsymbol{F}_{tx}$ for the transmitter, $\boldsymbol{F}_{rx}$ for the receiver and the propagation channel response matrix $\mathbf{h}_{\mathrm{n}}$ for the cluster *n* as follows

$$\mathbf{H}_{\mathrm{n}}(\mathrm{t};\tau) = \iint \mathbf{F}_{\mathrm{rx}}(\varphi)\mathbf{h}_{\mathrm{n}}(\mathrm{t};\tau,\phi,\varphi)\mathbf{F}_{\mathrm{tx}}^{\mathrm{T}}(\emptyset)\mathrm{d}\emptyset\mathrm{d}\varphi$$

The channel from Tx antenna element *s* to Rx element *u* for cluster *n* is

$$\begin{aligned}\mathrm{H}_{u,s,n}(\mathrm{t};\tau) = \sum_{\mathrm{m}=1}^{\mathrm{M}} &\begin{bmatrix} F_{rx,u,V}(\varphi_{n,m}) \\ F_{rx,u,H}(\varphi_{n,m}) \end{bmatrix}^{T} \begin{bmatrix} \alpha_{n,m,VV} & \alpha_{n,m,VH} \\ \alpha_{n,m,HV} & \alpha_{n,m,HH} \end{bmatrix} \begin{bmatrix} F_{tx,s,V}(\phi_{n,m}) \\ F_{tx,s,H}(\phi_{n,m}) \end{bmatrix} \\ &\times \exp\left(j2\pi\lambda_0^{-1}(\bar{\varphi}_{n,m}\cdot\bar{\gamma}_{rx,u})\right)\exp\left(j2\pi\lambda_0^{-1}(\bar{\phi}_{n,m}\cdot\bar{\gamma}_{tx,s})\right) \\ &\times \exp(j2\pi\nu_{n,m}t)\,\delta(\tau-\tau_{n,m})\end{aligned}$$

All the small scale parameters mentioned above are time variant if the radio channel is modeled as dynamic. Table 7 illustrates the corresponding meaning of all the parameters used in the above equation.

Table 7. MIMO channel parameters

$\boldsymbol{F_{rx,u,V}}$	antenna element *u* field patterns for vertical polarizations
$\boldsymbol{F_{rx,u,H}}$	antenna element *u* field patterns for horizontal polarizations
$\boldsymbol{\alpha_{n,m,VV}}$	complex gains for vertical-to-vertical polarizations for ray *n,m*
$\boldsymbol{\alpha_{n,m,VH}}$	complex gains for horizontal-to-vertical polarizations for ray *n,m*
$\boldsymbol{\lambda_0}$	wave length of carrier frequency
$\boldsymbol{\bar{\phi}_{n,m}}$	AoD unit vector
$\boldsymbol{\bar{\varphi}_{n,m}}$	AoA unit vector
$\boldsymbol{\bar{\gamma}_{tx,s}}$	Location vectors of element *s*
$\boldsymbol{\bar{\gamma}_{rx,u}}$	Location vectors of element *u*
$\boldsymbol{\nu_{n,m}}$	Doppler frequency component of ray *n,m*

4.2 WINNER generic channel model

WINNER generic model is a system level model that is able to indicate arbitrary number of propagation environment realizations for either single or multiple radio links for any defined

scenarios for desired antenna configurations with one mathematical framework by different parameter sets. There are two (or three) levels of randomness in generic model. The first random level is the LS (large scale) parameters like shadow fading, delay and angular spreads which are chosen randomly from distribution functions. The next level of randomness is the small scale parameters like delays, powers, directions of arrival and departure which are determined randomly based on distribution functions and random LS parameters (second moments). Geometric setup is therefore fixed in this step and only free variables are the random initial phases of the scatters. An unlimited number of different model realizations can be generated by selecting (randomly) different initial phases. The model is completely determined when the initial phases are fixed as well.

Modeled parameters used in the channel model are listed and illustrated below in Table 8. LS parameters are set first since they are considered as an average over a typical channel segment (distance of some tens of wave-lengths). The first three large scale parameters are used to control the distributions of delay and angular parameters.

Table 8. Parameters in WINNER II channel models

Large Scale Parameters	Support Parameters
Delay Spread and Distribution	*Scaling Parameter for Delay Distribution*
	Cross-Polarisation Power Ratios
Angle of Departure Spread and Distribution	*Number of Clusters*
	Cluster Angel Spread of Departure
Angle of Arrival Spread and Distribution	*Cluster Angel Spread of Arrival*
	Per Cluster Shadowing
Shadow Fading Standard Deviation	*Auto-Correlations of the LS Parameters*
	Cross-Correlations of the LS Parameters
Ricean K-factor	*Number of Rays per Cluster*

Different radio-propagation environment would cause different radio-channel characteristics. WINNER models are using propagation parameters obtained from channel measurements in different environments. For each environment scenario, measured data is analyzed and complemented with results from literature to obtain specific parameters. Thereafter, the same generic channel model is used for all scenarios by just using different values of channel parameters.

4.3 B1-Urban micro-cell

Urban micro-cell with NLOS channel model (B1) is used as channel model in this thesis work. As illustrated in [9], the height of both the antenna at BS (Base Station) and at the MS

(Mobile Station) is assumed to be much lower than the tops of surrounding buildings in urban micro-cell scenario. Both antennas are assumed to be outdoors in an area that streets layout is in Manhattan-like style. The streets in coverage area with LOS (Line-of-Sight) from all locations to the BS are classified as “the main street”, with the possible exception that the LOS is blocked by vehicles on the street temporarily. Streets intersecting the main street are referred to as perpendicular streets while those parallel to the main street are referred to as parallel streets. Cell shapes are determined by the surrounding buildings and energy transmission in the NLOS streets is considered as propagation around corners, through buildings and between them.

Measurements for urban micro-cellular scenario were taken in Helsinki city center at 53.GHz center frequency. The used chip rate was either 60 MHz or 100MHz.

4.3.1 B1 - Path-loss and shadow fading

The PL (Path-loss) model for B1 urban microcells of NLOS in regular street grid environment introduced in [9] is given as:

$$PL_{NLOS} = 0.096 * d_1[m] + 65 + (28 - 0.024 * d_1)\log_{10}(d_2[m])$$

$$10\text{m} < d_1 550m, \frac{w}{2} < d_2 < 450m, h_{BS} = 8m, h_{MS} = 2m$$

Where d_1 and d_2 are shown in Figure 9. It is shown that path loss is extremely close to $20\log_{10}(f)$ frequency dependency in range 0.5~15GHz (f is the target frequency in GHz). Corner loss characteristics, transition from LOS to NLOS, were examined as a function of f in range 3~15 GHz. With BS height increases path loss dependency on f is slightly greater due to wavelength difference. However, different frequency scaling for NLOS models would cause difficulties in model continuity. Therefore same frequency scaling $20\log_{10}(f)$ would be preferred. In the channel model in this work, only small scale path-loss was considered and yet the large scale path-loss was contained in SINR.

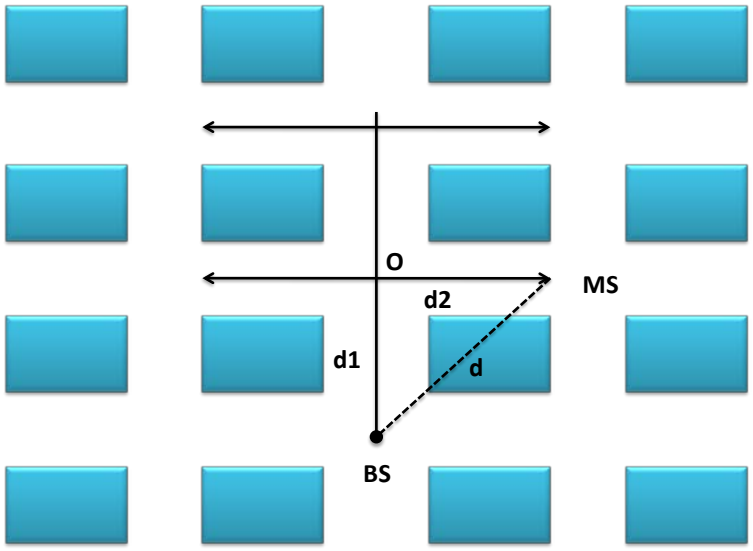


Figure 9: Layout of regular street grid

4.3.2 B1 – CDL Model

A need for reduced-complexity channel models is identified to be used in rapid simulations. Multipath AoD and AoA information is inherent for the determination of tap fading characteristics. Therefore, a reduced complexity models referred to as CDL (Clustered Delay Line) models was reported. Detailed knowledge of CDL models can be found in [9].

The parameters of the CDL model were extracted from measurements with chip frequency of 60 MHz at frequency range of 5.3 GHz.

Table 9: Scenario B1: NLOS CDL Model

Cluster#	Delay [ns]			Power [dB]		
1	0			-1.0		
2	90	95	100	-0.3	-5.2	-7.0
3	100	105	110	-3.9	-6.1	-7.9
4	115			-8.1		
5	230			-8.6		
6	240			-11.7		
7	245			-12.0		
8	285			-12.9		
9	390			-19.6		

10	430	-23.9
11	460	-22.1
12	505	-25.6
13	515	-23.3
14	595	-32.2
15	600	-31.7
16	615	-29.9

5 Link-level Model and Assumptions

The impact of traffic modeling design on packet delay and rate for different codeword length was investigated. In order to keep consistency with the original MI-ACM algorithm the simulation setup is based on the WINNER II Microcellular TDD mode of 1:1 asymmetry [13]. This can be easily extended to the FDD case, by using the corresponding bandwidth. We only consider the case of downlink transmission with a carrier frequency of 5 GHz. Different information block size, i.e. K = 288, 1152, 2304, are used for the MI-ACM algorithm. The FFT bandwidth is 100 MHz within which the signal bandwidth is 89.84 MHz. The subcarrier distance is 48.828 MHz and the useful symbol duration is 20.48 μs . Each RB contains 15 OFDM symbols and 8 subcarriers and is 0.3456 *ms* long in the time domain, which includes a duplex guard time of 8.4 μs and 390.62 kHz in the frequency domain. The scheduling method used here is the classical Proportional Fair (PF) scheduling algorithm and a total transmission time of 3 seconds is simulated. We drop 50 users to the system at a time.

The maximum tolerable packet delay for VoIP is set to 50 milliseconds, which means that if the scheduling delay time of VoIP application exceeds the maximum tolerable value, the user would be dropped out and thereafter, the corresponding buffer would be empty and thus a new user would be added in. Thus, there are always 50 users in the system. The QoS (Quality of Service) for traffic model of VoIP application is evaluated mainly by the packet delay and number of satisfied users been dropped since VoIP application requires strictly low time-to-live delay. On the other hands, video streaming is a kind of elastic service, and usually there is a quite large buffer in both of the transmitter and receiver side. Thus, we only consider the rate constraint as its QoS parameter. We also present the maximum and average delay that a video streaming user may experience during the transmission in the following section.

6 Simulation Results

For the traffic model for VoIP application, We setup the simulation for the VOIP application but the simulation results indicate that all 50 users could obtain satisfaction with the stringent

delay (50ms) requirement. According to the VoIP conversational model in Figure 3, the possibility of a user would generate any packet during each time interval can be calculated. Whether a user would generate any packet depends on whether the user is talking or silent. Taking User A as an example and assuming its previous state was State 1, then in this time interval the possibility of User A being in talking state is:

$$(P11+P14)\times(1/4);$$

1/4 is the possibility of User A being in State 1 during previous time interval. Filling the other three assumptions complete, the following equations could be obtained:
When User A's previous state was State 2, the possibility of User A talking in this time interval is:

$$P24\times(1/4);$$

When User A's previous state was State 3, the possibility of User A talking in this time interval is:

$$P31\times(1/4);$$

When User A's previous state was State 4, the possibility of User A talking in this time interval is:

$$(P44+P41)\times(1/4);$$

Therefore, the possibility of User A talking in any time interval is the summation of the four equations:

$$(P11+P14)\times(1/4)+P24\times(1/4)+P31\times(1/4)+(P44+P41)\times(1/4)=0.5;$$

The value of the possibilities (*P11, P14, P24, P31, P44, P41)* are indicated in *Chapter 3.1.2*. This result also means the possibility of a user having packet sending out in any time interval is 0.5.

Therefore, the possibility of each user staying in the talking state is 0.5 which is equal to the possibility of user staying in the silent state. In order to obtain more precise result, the calculation of how many users asking for service at the same time with the highest possibility and its value is processed. Assuming there are only 3 users, and then the possibility of 2 of them talking at the same time is:

$$0.5\times(1-0.5)\times(1-0.5)\times C_3^2;$$

It can also be written as:

$$0.5^3 \times C_3^2;$$

Therefore, for 50 users, the possibility of *n (≤50)* users talking at the same time is:

$$p(n) = 0.5^n \times C_{50}^n;$$

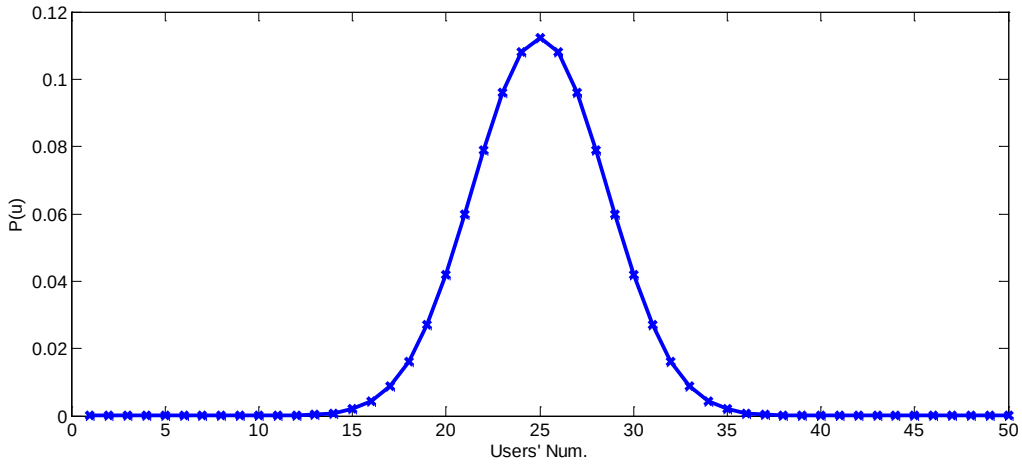


Figure 10. Possibility of *n* users talking together

From Figure 10 it can be seen apparently that 25 users talking at the same time taking the highest possibility point but only with around 10% which is such a small value. Each user is independent, for example the possibility of all the 50 users being in the same talking state is 0.5^50 which is extremely small and thus it would not happen basically, Therefore, the reason why all the users are satisfied is that not all the 50 users would generate packets and need service at the same time instant, plus the user packet is small, according to the traffic model, the channel bandwidth is adequately enough with 10MHz and the users could be satisfied even with low SNR. Since our simulation was processed on link-level model, 50 users are the maximum number of users that the system could handle based on the computational capability we have. System-level simulation would be necessarily required if more users needed to be tested.

For the traffic model for video streaming application, it can be seen clearly from Figure 11-2 and 11-3 that the delay time increases when the codeword length grows especially while the users' mean rate remain almost the same for three different codeword lengths with fairly low SINR values. The average delay (seen in Figure 11-2) for the longest codeword length (K=2304) is around 600ms within the low SINR region. With the SINR values increasing, users with shorter codeword length could obtain better performance with respect to users' mean rate. Since PF scheduling was used in the simulation, even in the condition that the delay value is large, the scheduler would still keep the rate almost remain the same when the buffer is long enough (can be seen as in Figure 11-1).

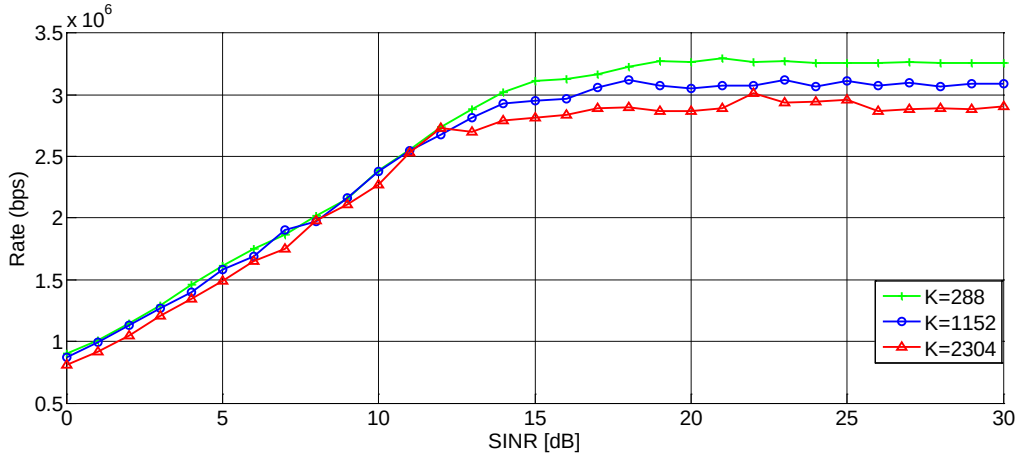


Figure 11-1. Users' Mean Rate of Different Codeword Length

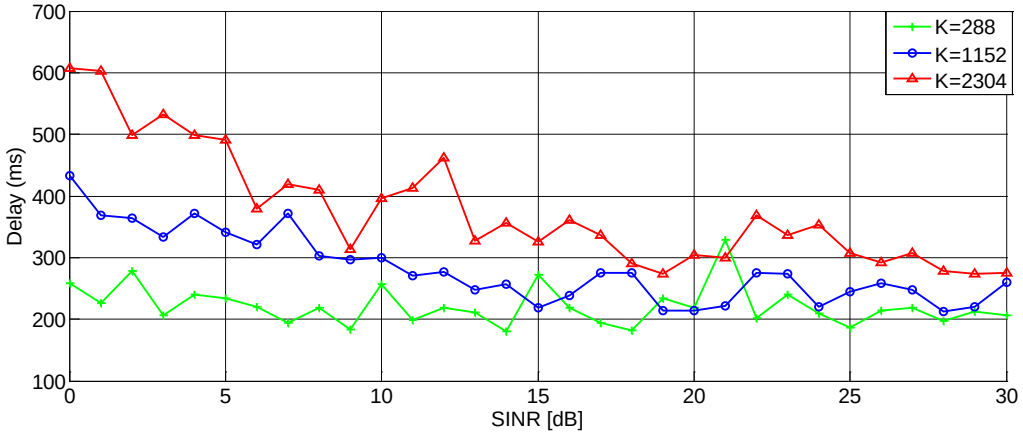


Figure 11-2. Mean Delay of Different Codeword Length

In Figure 11-3 and Figure 11-4, the worst scenario of all users' performance is shown in the aspect of delay time and its corresponding minimum rate. The maximum delay time could be over 2 s in the worst case scenario with the bad channel conditions. The reason why the worst case scenario exists in this way is that the packet in video streaming could be large or small, when the packet is large occasionally and the channel condition is not so good, the delay time would therefore increase dramatically. Another reason could be that while the codeword length increases (K=2304), it would thus take more resource blocks, in addition, if the SINR values are low, the delay time would of course rise up. In Figure 11-1 and 11-4, it can be seen that even the lowest rate at the beginning with small SINR value is still bigger than the minimum rate constraint (64 Kbps).

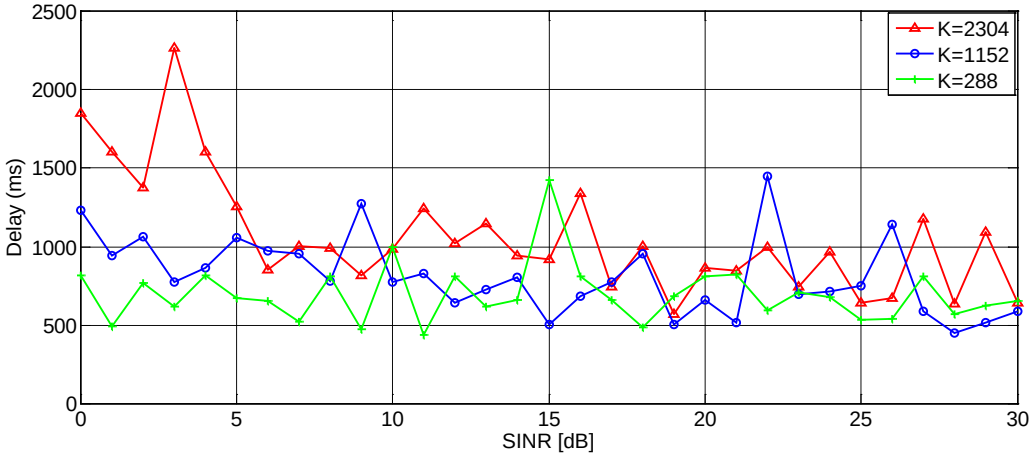


Figure 11-3. Maximum Delay of the worst user

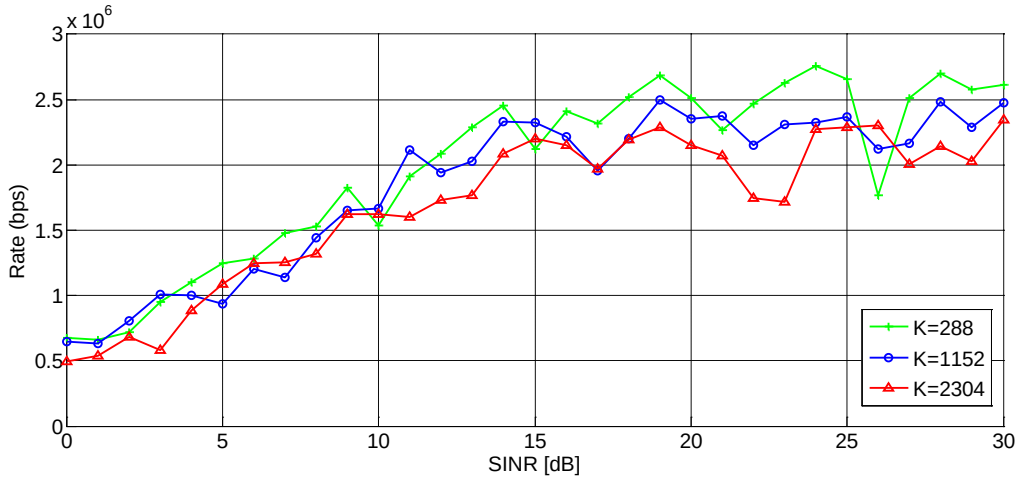


Figure 11-4. Minimum Rate of the worst user

For different codeword length, zero padding is needed when the packet is small. Therefore, when codeword length is of large value as well as the packet is small, it would yield loss and thereafter the delay time would be longer in the condition of low SINR values which can be seen in Figure 11-2.

Due to the packet size is variable, relatively big-sized packet appears occasionally. Thus, several scheduling round might be needed especially when the channel conditions are not so good (even when the SINR is of high values, this condition still exists) which would lead to even longer delay time. However, this situation happens rarely which can be seen in all the CDF and HIST figures. But for the consideration of QoS in this condition, then the buffer size needs to be enlarged according to the reference of maximum delay of the worst case shown in Figure 11-3.

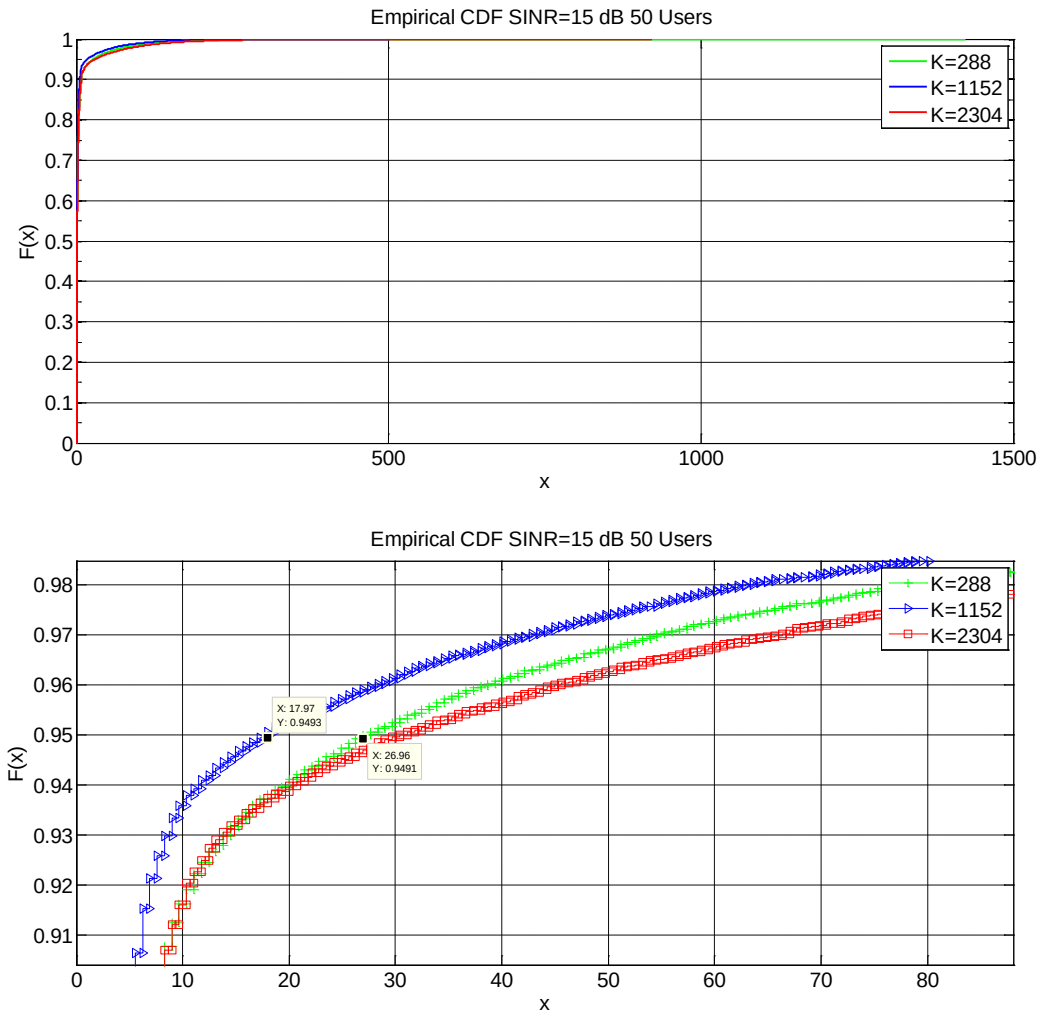
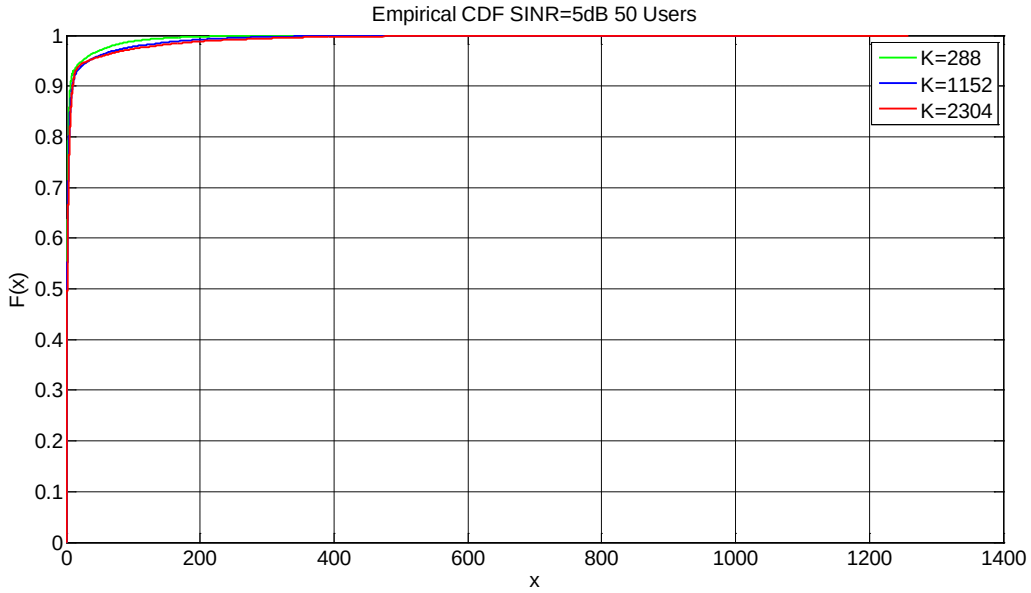


Figure 11-5. CDF of delay time of 50 users with SINR = 15 and the enlarged segment



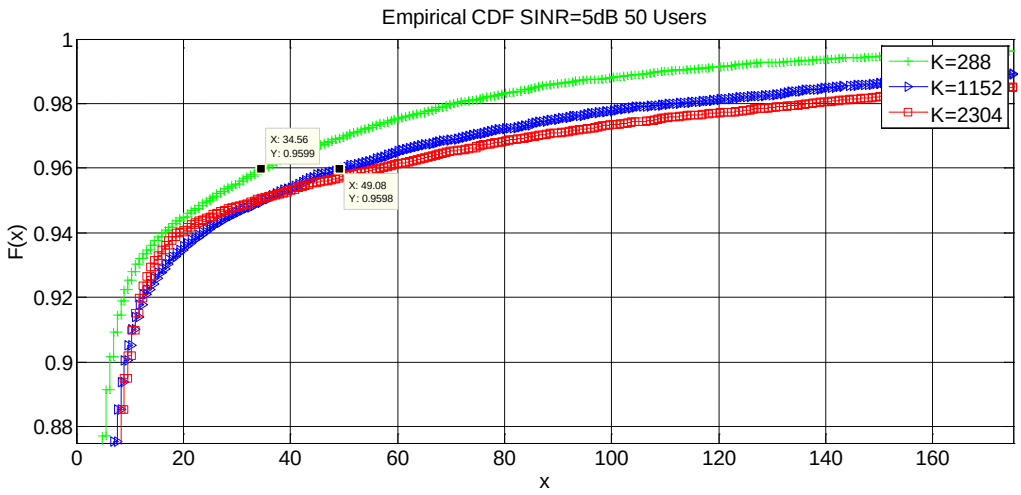


Figure 11-6. CDF of delay time of 50 users with SINR = 5 and the enlarged segment

The figures (Figure 11-5 and Figure 11-6) of CDF of delay time show that with 50 users, when the SINR values are relatively high with 15 dB in Figure 11-4, around 95% users could obtain delay time less than 27.66ms with codeword length K=288 while same amount of users obtain delay time less than 17.97ms with codeword length K=1152. And when the SINR values decrease, the performance with codeword length K=288 and K=1152 exchange to the opposite position which can be seen in Figure 11-6. It is obvious that when codeword length K=288 is used, less delay time could be needed to satisfy the same amount of users.

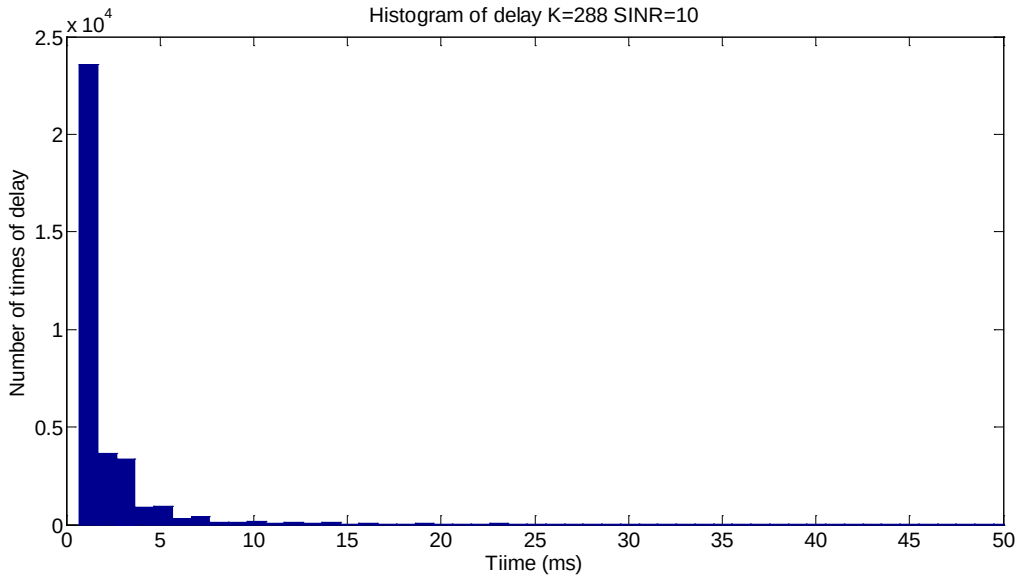


Figure 12. Histogram of delay with K=288 and SINR = 10 dB (50 users)

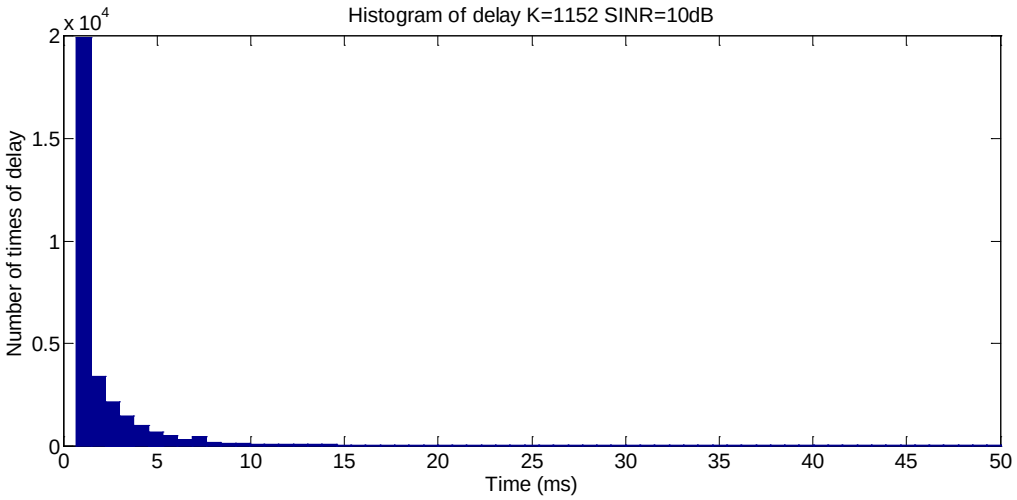


Figure 13. Histogram of delay with K=1152 and SINR = 10 dB (50 users)

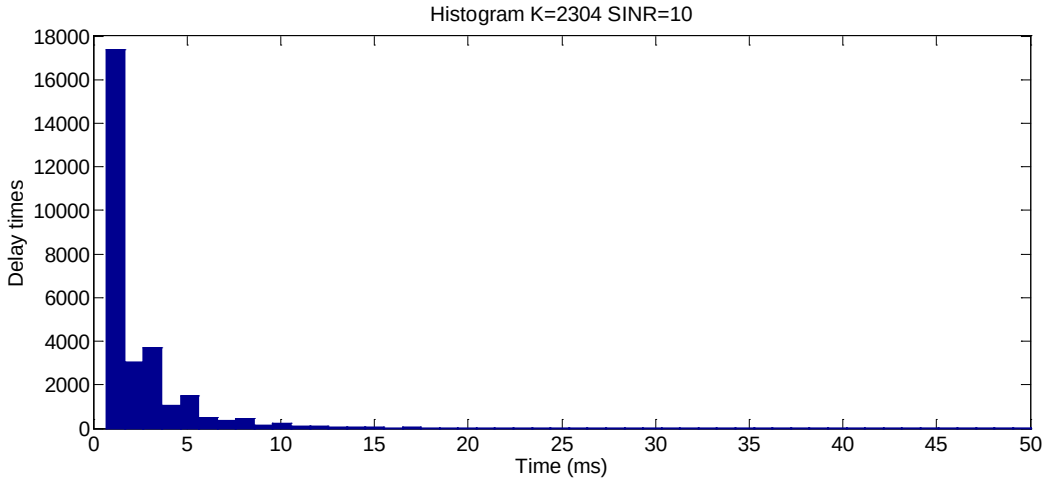


Figure 14. Histogram of delay with K=2304 and SINR = 10 dB (50 users)

From the Figures of Histogram of delay time (Figure 12~Figure 14) for 50 users, most of the delay time locations are within 15 ms which is far less than tolerable delay time for video streaming. For SINR=10, the shorter the codeword length is, much more shorter delay time appears which can be concluded that shorter codeword length get better performance when the SINR values are relatively low and this is also proved in Figure 11-6.

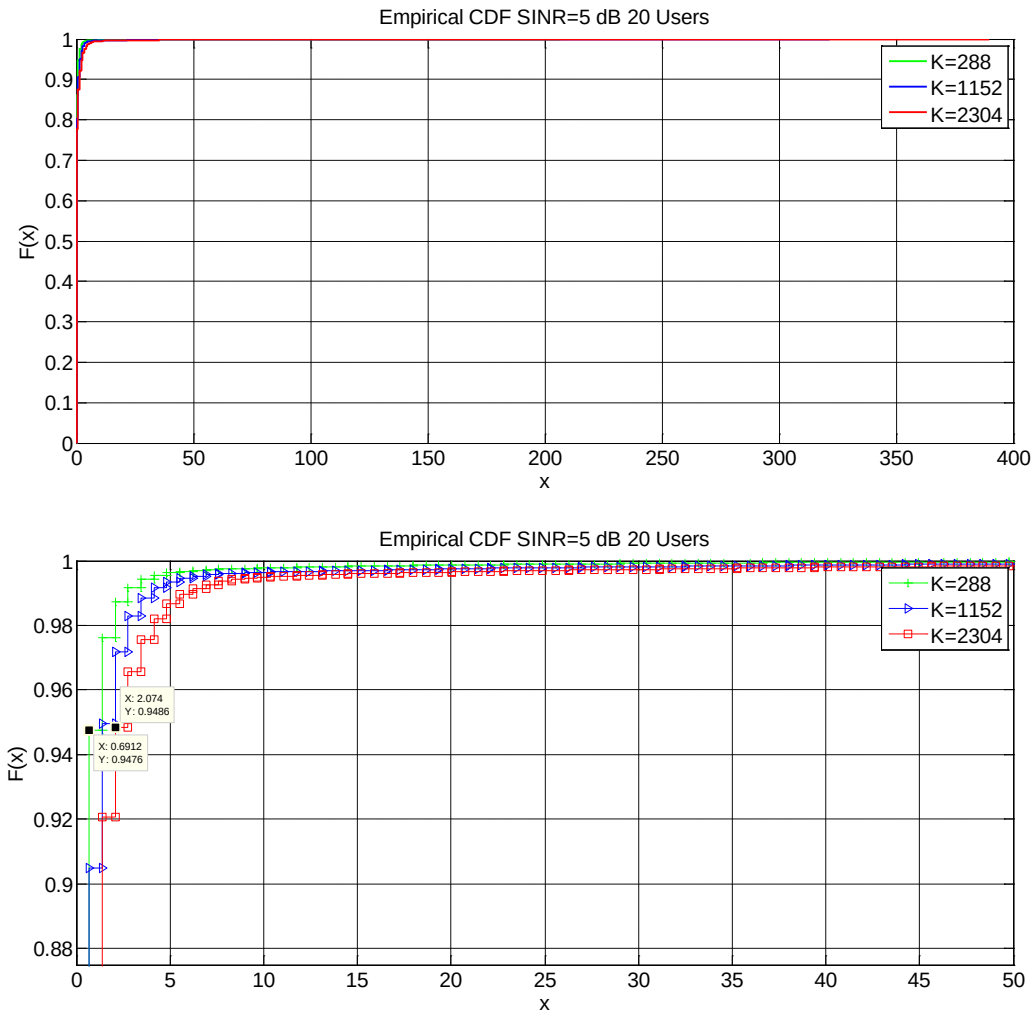


Figure 15. CDF of delay time of 20 users with SINR = 5 and the enlarged segment

In the figure of CDF of delay time for 20 users (Figure 15), it can be seen that the performance for each codeword are almost the same with even less difference. And in the enlarged segment, the conclusion obtained for 50 users in Figure 11-6 still holds true that shorter codeword length performs better with low SINR values even for less users.

7 Conclusions

I present the performance of the MI-ACM algorithm with 3 different codeword lengths in the physical layer, namely K=288, K=1152 and K=2304 using realistic VoIP and Video source models. Simulation results show that, in VoIP application with the number of users or the rate increases, it is possible that some users would not be satisfied, but it needs simulation on system-level for more investigation. Since our simulation was processed on link-level model, 50 users are almost the maximum number of users that the system could handle based on the computational capability we have.

From the results of video application, it can be concluded that shorter codeword length could obtain better performance with the respect of delay time for relatively low SINR values. And for the respect of rate, the system could support video source with even higher rate than 64 Kbps obviously. Video streaming application is a kind of elastic service and its buffer size could be determined by the up layer services. Considering the worst case scenario in the bad channel conditions, the maximum delay could be more than 2s (as seen in Figure 11-1～11-3). Therefore, it can be suggested that the upper layer could set up the buffer size as approximate corresponds to 1 to 2 seconds or even more for acceptable user experience.

Furthermore, to attain a generally more specific performance of MI-ACM algorithm with VoIP traffic model for more (than 50) users, it would require system-level evaluation which can be the future work on this subject. In this thesis, the VoIP and Video Streaming traffic models are assumed for each UE. The evaluation of the scheme on mixed traffic model scenario could provide more valuable opinions on scheduler in order to obtain better system throughput.

Reference

[1] Y. Sui, D. Aronsson, T. Svensson, “Evaluation of Link Adaptation Methods in Multi-User OFDM Systems with Imperfect Channel State Information,” Signals and Systems, Chalmers University of Technology.

[2] IST-4-027756 WINNER II D2.2.3, “Modulation and Coding Schemes for the WINNER II System,” Nov. 2007.

[3] K. Safjan, J. Oszmiański, M. Döttling, A. Bohdanowicz, “Frequency-Domain Link Adaptation for Wideband OFDMA Systems,” Nokia Siemens Networks.

[4] J. Oszmiański, K. Safjan, M. Döttling, A. Bohdanowicz, “Impact of Traffic Modeling and Scheduling on Delay and Spectral Efficiency of the WINNER System,” Nokia Siemens Networks.

[5] S.Stiglmayr, M. Bossert, “Adaptive Coding and Modulation in OFDM System using BICM and Rate-Compatible Punctured Codes,” in European Wireless, France, April 2007.

[6] M. Sternad, S. Falahati, T. Svensson, D. Aronsson, “Adaptive TDMA/OFDMA for Wide-area Coverage and Vehicular Velocities,” Proc. IST Mobile and Vehicular, 2005.

[7] S. Pfletschinger, A. Piatyszek, S. Stiglmayr, “Frequency-Selective Link Adaptation using Duo-Binary Turbo Codes in OFDM Systems,” Proc. 16th IST Mobile & Wireless Communication Summit 2007, Budapest, Hungary, July 2007.

[8] J. Li, Y. Xia, S. Xingang, W. Zhiliang, “Conversational Model Based VoIP Traffic Generation,” Department of Computer Science and Technology & Network Research Center, Tsinghua University, Beijing, China.

[9] IST-4-027756 WINNER II, D1.1.2 V1.2 “WINNER II Channel Models,” Sep. 2007.

[10] 3GPP TSG-RAN1#48, “LTE Physical Layer Framework for Performance Verification,” Feb. 2007.

[11] C.H. NG, L. Bai, B.H. Soong, “Modeling Multimedia Traffic over ATM using MMBP,” *IEE Pro.-Commun.*, Vol. 144, No. 5, Oct. 1997.

[12] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, J.D. Robbins, “Performance Models of Statistical Multiplexing in Packet Video Communication,” *IEEE Trans. Comm.*, vol. 36, pp. 834-43, 1988.

[13] P. Sen, B. Maglaris, N. Rikli, D. Anastassiou, “Models for Packet Switching of Variable-Bit-Rate Video Sources,” *IEEE J. on Select. Areas in Comm.*, vol. 7, no. 5, June 1989.

[14] F. Yegenoglu, B. Jabbari, Ya-Qin Zhang, “Motion-Classified Autoregressive Modeling of Variable Bit Rate Video,” *IEEE Trans. CSVT.* vol 3, no. 1, Feb. 1993.

[15] G. Bai, “Traffic Modeling for Multimedia Communication Systems,” GMD-German National Research Center for Information Technology.

[16] D. Turaga, T. Chen, “Activity-Adaptive Modeling of Dynamic Multimedia Traffic,” Electrical and Computer Engineering Carnegie Mellon University.