Journal of Southeast University (English Edition) Vol. 31, No. 2, pp. 163 - 169 June 2015 ISSN 1003—7985

Frame length optimization for multi-antenna downlink systems based on delay-bound violation probability constraints

Tan Yufeng Li Junchao Xia Weiwei Shen Lianfeng

(National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China)

Abstract: A frame length optimization scheme is proposed for multi-antenna downlink systems to guarantee diverse delay-bound violation probability constraints. Due to the difficulties of extracting the quality of service (QoS) metrics from the conventional physical-layer channel models, the link-layer models named effective bandwidth and effective capacity are applied to statistically characterize the source traffic patterns and the queuing service dynamics. With these link-layer models, the source traffic process and the channel service process are mapped to certain QoS parameters. The packet delay-bound violation probability constraints are converted into minimum data rate constraints and the optimization problem is thus formulated into simultaneous inequalities. With the assumption of ergodic block-fading channels, the optimal frame lengths of single-user and multiuser systems are calculated respectively by numerical iterative methods. Theoretical analyses and simulation results show that the given delay-bound violation probability constraints are well satisfied with the optimal frame length.

Key words: delay-bound violation probability; frame length optimization; effective bandwidth; effective capacity; multi-antenna systems; quality of service

doi: 10.3969/j. issn. 1003 - 7985. 2015. 02. 001

The last decade has seen an explosive growth in multimedia applications, which brings an unprecedented demand for wireless network performance enhancement to support high-data-rate services with a diverse quality of service (QoS) requirements. In multi-antenna systems, guaranteeing heterogeneous QoS to multiple users arouses great concern and remains challenging due to the time-varying fading channel, the co-channel interference (CCI) and resource competition among users. Note that the frame length of the downlink of multi-antenna systems has an obvious impact on the system throughput, where a trade-off exists between the multiuser diversity gain by choosing smaller frame lengths and the throughput improvement by making frames larger.

The frame length optimization problem has been studied separately at the physical (PHY) layer and the media access control (MAC) layer. The orthogonal frequency division multiplexing (OFDM) systems were studied in Refs. [1 - 3]. A optimized data transmission scheme was proposed by optimizing the auxiliary signaling interval to maximize the amount of payload data transmitted[1], while the optimal frame size is selected with concern regarding maximizing the spectrum efficiency[2]. Besides, Wu et al. [3] formulated the system efficiency as a function of the frame length, scheduling cost and channel conditions, and then found the optimal frame length by dynamic programming algorithms. On the other hand, this problem was investigated from the MAC-layer perspective in Refs. [4 - 6]. Iqbal et al. [5] studied the tradeoff between reliability and frame length with the redefined reliability function for the binary symmetric channels and the *K*-th order Markov channels. In Ref. [6], the slot allocation and frame length of ad hoc networks were jointly optimized through a decentralized TDMA-based MAC-layer scheduling protocol to maximum the slot reuse. However, instead of complex upper-layer QoS guarantees, the above schemes were evaluated based on the PHY-layer parameters including the bit error rate (BER) or the frame error rate (FER). Furthermore, users or terminals were described in terms of the signal-to-noise ratio (SNR) rather than statistical traffic characteristics. There remains a lack of cross-layer analysis on frame length optimization problems with diverse QoS constraints as well as appropriate traffic statistics and channel behavior characterization for multi-antenna systems.

In this paper, we seek to guarantee reliable transmission under the maximum tolerable delay-bound violation probabilities by optimizing the frame length of multi-antenna downlink systems. We study the frame length optimization problem with a joint consideration of the transmission rates at the PHY layer and diverse QoS requirements at the MAC layer. First, the source statistics and the queue dynamics are both incorporated, where the data sources are modeled as discrete-time, finite state, irreducible, stationary Markov chains, and the queue service

Received 2014-11-04.

Biographies: Tan Yufeng (1988—), female, graduate; Shen Lianfeng (corresponding author), male, professor, lfshen@ seu. edu. cn.

Foundation items: The National Science and Technology Major Project (No. 2012ZX03004005-003), the National Natural Science Foundation of China (No. 61171081, 61201175), the Research Fund of National Mobile Communications Research Laboratory of Southeast University (No. 2014A03).

Citation: Tan Yufeng, Li Junchao, Xia Weiwei, et al. Frame length optimization for multi-antenna downlink systems based on delay-bound violation probability constraints[J]. Journal of Southeast University (English Edition), 2015, 31 (2): 163 - 169. [doi: 10. 3969/j. issn. 1003 - 7985. 2015. 02. 001]

is characterized by the functions mapped to the link-level QoS guarantees. Different from the conventional PHY-level channel models which characterize the fluctuations in the amplitude of the radio signal, the concepts of effective bandwidth and effective capacity[7-9] are applied to illustrate the source traffic pattern and channel behavior, respectively. Secondly, the maximum tolerable delay-bound violation probabilities are converted into data rate constraints. Then, the cross-layer frame length optimization problem can be expressed as functions of source traffic parameters, QoS requirements and data rates. With certain source traffic models and QoS parameters, the optimal frame length can be calculated. As a result, reliable transmission under certain delay-bound violation probability constraints is guaranteed, while the quantitative relationship between the optimal frame length and QoS requirements is revealed.

1 System Model and QoS Guarantees

1.1 System model

Consider a multi-antenna downlink system with a base station of M_t transmit antennas and K users each holding one downlink traffic request, as shown in Fig. 1. The k-th user has N_k receiving antennas with $M_t \geqslant \sum_{k=1}^{K} N_k$, $k = 1, 2, \cdots, K$. Perfect channel state information (CSI) is known to the transmitter as well as the receiver, and the base station collects various QoS requirements from each user before transmission. We consider a spatially uncorrelated stationary and ergodic fading channel with a bandwidth B. The channel gain remains constant over a frame slot and changes independently from one slot to another based on the channel distribution. Let $\boldsymbol{H}_k \in \mathbf{C}^{N_k \times M_t}$ denote the downlink channel matrix of the k-th user, where the entries of $\boldsymbol{H}_k$ are assumed to be independent and identically distributed (i. i. d) zero-mean, equal variance, complex circularly symmetric Gaussian random variables. $\boldsymbol{x}_k \in \mathbf{C}^{N_k \times 1}$ demonstrates that the data vector of the k-th user

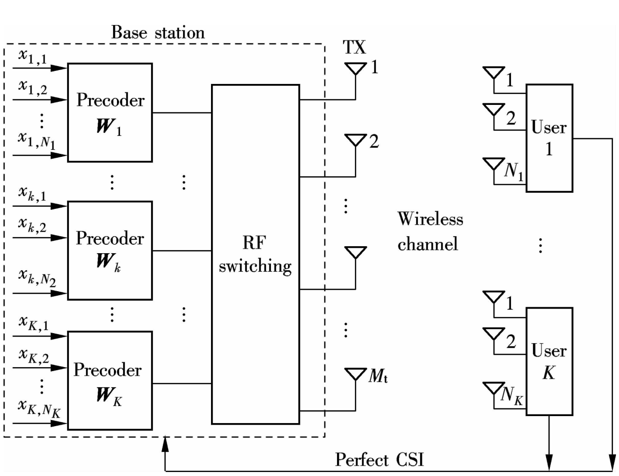

Fig. 1 Multi-antenna broadcast system with precoding and perfect CSI feedback

is preprocessed at the transmitter by $M_t \times N_k$ precoding matrix $\boldsymbol{W}_k$. Then, the N_k-dimensional received signal of the k-th user can be represented as

$$\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{W}_k \boldsymbol{x}_k + \boldsymbol{H}_k \sum_{j \neq k}^{K} \boldsymbol{W}_j \boldsymbol{x}_j + \boldsymbol{n}_k \tag{1}$$

where $\boldsymbol{n}_k \in \mathbf{C}^{N_k \times 1}$ denotes the additive white Gaussian noise (AWGN) with zero mean and the covariance matrix $\sigma^2 \boldsymbol{I}_{N_k}$.

With perfect knowledge of CSI, we can cancel inter-user interference, the second term in Eq. (1), by appropriately designing the precoding matrices. After that, the multiuser multiple-input multiple-output (MIMO) channel is decomposed into K parallel noninterfering single-user MIMO links. According to prior analysis of the MIMO channel capacity, the achievable rate of the k-th user is given by

$$R_k = \ln \left| \boldsymbol{I}_{M_{r,k}} + \boldsymbol{H}_k \boldsymbol{W}_k \boldsymbol{W}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \right| = \sum_{n=1}^{m_k} \ln(1 + p_{n,k} \sigma_{n,k}^2) \tag{2}$$

where $\{p_{n,k}, n = 1, 2, \cdots, m_k\}$ denote the SNR per receiving antenna under unity channel gain for the individual data substreams of the k-th user and $\{\sigma_{n,k}, n = 1, 2, \cdots, m_k\}$ are nonzero singular values of the equivalent channel matrix $\boldsymbol{H}_k \boldsymbol{W}_k$ with $m_k = \operatorname{rank}(\boldsymbol{H}_k \boldsymbol{W}_k)$. The equal power allocation to the transmit antennas is adopted instead of an adaptive power distribution strategy for the purpose of simplification, replacing $p_{n,k}$ with p.

1.2 Descriptions of QoS guarantees

Generally, QoS parameters at the MAC layer or upper layers are referred to delay bounds or delay jitter for a certain minimum rate, while those at the PHY layer are represented by SNR, BER at the receiver. In this paper, we assume a time-slotted packet communication model without the automatic repeat request (ARQ) protocol. For each user, packets generated by the source are sequentially stored in a queue in the transmit buffer and grouped into frames before being served in a first-come-first-serve manner. Let T be the frame or time slot duration mentioned earlier. To characterize the burstiness of the arrival traffic, we assume that the arrival stream $\{A_t\}$ is modulated by a discrete-time, finite state, irreducible, stationary Markov chain $\{X_t\}$ with the state space S and the transition matrix $\boldsymbol{Q} = (q_{ij})$, $i, j \in S$. Thus, the distribution of A_t at time t depends only on the source state X_t and the A_t's are independent while in one state of the chain. Consider a two-state ON-OFF Markov chain with exponentially distributed ON and OFF period as the traffic source in this paper. The number of packets generated per time slot during ON state follows an independent Poisson distribution of rate α_k and each packet has a fixed size N_b in bits.

Specifically, a packet will be dropped if its waiting time in the buffer exceeds its given delay bound. When the ARQ is not employed, a packet is considered successfully received at the receiver only if it has not been dropped due to delay violation and has been correctly received through the wireless channel. If D_k is the delay experienced by a source packet in the buffer, then the probability of exceeding a delay bound $D_k^{\max}$ is defined as the delay-bound violation probability, denoted by $\Pr(D_k > D_k^{\max})$. Given the maximum tolerable delay-bound violation probability as ε_k, the delay-QoS constraint of the k-th user can be expressed as

$$\Pr(D_k > D_k^{\max}) \leqslant \varepsilon_k \tag{3}$$

Hence, our objective is to find the optimal frame length satisfying constraint (3). Note that parameters $D_k^{\max}$ and ε_k show the stringency on delay requirements of different traffic, i. e., the more delay-sensitive traffic has a smaller $D_k^{\max}$ and ε_k.

Since we aim to guarantee the delay restriction via optimizing physical layer transmission, it is necessary to express the constraint (3) in terms of physical layer parameters. According to Ref. [7], the effective capacity channel is modeled as the dual of the effective bandwidth source model. For a source with certain $D_k^{\max}$ and ε_k, the effective bandwidth function determines the minimum constant service rate. On the other hand, the effective capacity of a channel limits the maximum date rate of a source with $D_k^{\max}$ and ε_k. In this paper, a statistical description for packet delay probability is adopted, which depends on the log moment generating functions of arrival and service processes, also called the Gartner-Ellis (GE) limit. For the k-th user, let $A_k(0,t)$ denote the amount of bits generated by the traffic source in the interval $[0, t)$. Assume that the GE limit of $A_k(0,t)$, defined as

$$\Lambda_k^A(z_k) = \lim_{t\to 0}\frac{1}{t}\log[E\{\exp(z_k A_k(0,t))\}] \tag{4}$$

exists and is finite, differentiable for all real z_k, here $E\{\cdot\}$ is the expectation operator.

Paralleling $\Lambda_k^A(z_k)$, the GE limit of the accumulative service process in $[0,t)$ denoted by $C_k(0,t)$ is given by

$$\Lambda_k^C(z_k) = \lim_{t\to 0}\frac{1}{t}\log[E_{H_k}\{\exp(z_k C_k(0,t))\}] \tag{5}$$

where the expectation $E_{H_k}\{\cdot\}$ is with respect to all channel states.

As proved in Ref. [10], if there is a unique $z_k^* > 0$ satisfying

$$\Lambda_k^A(z_k^*) + \Lambda_k^C(-z_k^*) = 0 \tag{6}$$

then the left-hand term of constraint (3) can be appropriately approximated as

$$\Pr(D_k > D_k^{\max}) \approx \Pr(D_k > 0)\exp(\Lambda_k^C(-z_k^*)D_k^{\max}) \tag{7}$$

Assume that the arrival rates of the traffic source are large enough so that the buffer is always non-empty, with $\Pr(D_k > 0) = 1$. Hence, a necessary and sufficient condition for constraint (3) is

$$\Lambda_k^C(-z_k^*) \leqslant \frac{\ln\varepsilon_k}{D_k^{\max}} \tag{8}$$

Consequently, the delay-QoS requirements of the user can be guaranteed.

2 Analysis of Frame Length Optimization

2.1 Single-user systems

We first focus on the case where the multi-antenna downlink system is single-user. According to the system model, we can derive the degenerate model in a point-to-point communication system of M_t transmit and N_k receiving antennas

$$\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{n}_k$$

for $\boldsymbol{H}_k \in \mathbf{C}^{N_k\times M_t}$, $\boldsymbol{x}_k \in \mathbf{C}^{M_t\times 1}$ and $\boldsymbol{y}_k$, $\boldsymbol{n}_k \in \mathbf{C}^{N_k\times 1}$. With a singular value decomposition (SVD), the achievable rate of Eq. (2) can be simplified as

$$R_k = \ln|\boldsymbol{I}_{N_k} + \boldsymbol{H}_k\boldsymbol{H}_k^{\mathrm{H}}| = \sum_{n=1}^{m_k}\ln(1 + p_{n,k}\sigma_{n,k}^2) \tag{9}$$

where N_k is the number of nonzero singular values of the full rank $\boldsymbol{H}_k$. Based on the aforementioned objective and constraints, the optimization problem under the single-user scenario is to discover the frame length region both satisfying Eq. (6) and condition (8). In the worst case scenario, we have $\Pr(D_k > D_k^{\max}) = \varepsilon_k$ and then the optimization is equivalent to finding the boundary satisfying the following binary simultaneous equations:

$$\left.\begin{aligned} &\Lambda_k^A(z_k^*) + \Lambda_k^C(-z_k^*) = 0 \\ &\Lambda_k^C(-z_k^*) = \frac{\ln\varepsilon_k}{D_k^{\max}} \end{aligned}\right\} \tag{10}$$

For any discrete-time Markov modulated process, let $G_i(z) = E\{\exp(zA_t) \mid X_t = i\}$ be the generating function of the conditional distribution of A_t with the source state $X_t = i$. With the constructed matrix $\boldsymbol{U}(z) = [q_{ij}G_i(z)]$, the GE limit for a Markov source is[11]

$$\Lambda^A(z) = \frac{1}{T}\log\rho(\boldsymbol{U}) \tag{11}$$

where $\rho(\boldsymbol{U})$ is the spectral radius of the square matrix $\boldsymbol{U}$. In the case of a two-state ON-OFF Markov source, $\Lambda_k^A(z_k)$ can be formulated by

$$\Lambda_k^A(z_k) = \frac{1}{T}\ln\left[\frac{1}{2}\left(a_k(z_k) + \sqrt{a_k^2(z_k) + 4b_k(z_k)}\right)\right] \tag{12}$$

where $a_k(z_k)=q_{k,1}+q_{k,2}G_k(z_k)$, $b_k(z_k)=(1-q_{k,1}-q_{k,2})G_k(z_k)$, and $G_k(z_k)=E\{\exp(z_k A_t)\mid X_t=\text{ON}\}=\exp(\alpha_k(\exp(z_k N_b)-1))$.

Let $c_{k,\tau}$ denote the data rate of the k-th user at time-slot τ. According to the assumptions on physical layer modules and channel fading, the queue dynamics can be modeled as a discrete-time stationary service process and, therefore, $C_k(0,t)$ can be expressed as $C_k(0,t)=\sum_{\tau=1}^{t/T}B(T-\delta)c_{k,\tau}$, where δ represents the overhead of each frame. As the rate sequence $\{c_{k,\tau},\tau=1,2,3,\cdots\}$ is an uncorrelated process, the limit of Eq. (5) becomes

$$\Lambda_k^C(z_k)=\frac{1}{T}\ln[E_{H_k}\{\exp(z_k B(T-\delta)r_k)\}] \quad (13)$$

where r_k is the data rate of the k-th user. Since the expectation is with respect to $\boldsymbol{H}_k$, $\Lambda_k^C(-z_k^*)$ can be computed if the distribution of singular values for $\boldsymbol{H}_k$ is determined. As mentioned in Section 2, the entries of $\boldsymbol{H}_k$ are assumed to be i. i. d zero-mean, equal variance, complex circularly symmetric Gaussian random variables. Then $\boldsymbol{H}_k\boldsymbol{H}_k^{\mathrm{H}}$ is an $N_k\times N_k$ random non-negative definite matrix and has real, non-negative eigenvalues $\{\lambda_n,n=1,2,\cdots,N_k\}$, where $\lambda_n=\sigma_{n,k}^2,n=1,2,\cdots,N_k$. Thus, the joint density of these eigenvalues is given by[12]

$$f_k(\boldsymbol{\lambda}_k)=\frac{\exp\left(-\sum_{i=1}^{N_k}\lambda_i\right)\prod_{i=1}^{N_k}\lambda_i^{M_t-N_k}\prod_{i<j}(\lambda_i-\lambda_j)^2}{N_k!\left(\prod_{i=1}^{N_k}(M_t-i)!(N_k-i)!\right)} \quad (14)$$

Let $g_k(\boldsymbol{\lambda}_k)=\exp(-z_k^*B(T-\delta)r_k)$ with $\boldsymbol{\lambda}_k=\{\lambda_n,n=1,2,\cdots,N_k\}$, and $\Lambda_k^C(-z_k^*)$ can be mathematically expressed as the N_k-fold integral,

$$\Lambda_k^C(-z_k^*)=\frac{1}{T}\ln\Big[\underbrace{\int_0^\infty\cdots\int_0^\infty}_{N_k\text{-fold}} g_k(\boldsymbol{\lambda}_k)f_k(\boldsymbol{\lambda}_k)\mathrm{d}\lambda_1\mathrm{d}\lambda_2\cdots\mathrm{d}\lambda_{N_k}\Big] \quad (15)$$

where $f_k(\boldsymbol{\lambda}_k)$ is the joint probability density function of $\boldsymbol{\lambda}_k$. As a result, the GE limit of the accumulative service process can be calculated by iterative methods.

2.2 Multiuser systems

Consider the multiuser systems with all K users. Multiple users share the system resources and the downlink channel. Assuming that $M_t=\sum_{k=1}^{K}N_k$, there are no extra transmit antennas for the eigenmode selection to optimize the symbol error rate (SER). CCI constitutes the major performance impairment since data is simultaneously transmitted to multiple users through the non-orthogonal spatial channels. Thus, we first design the multiuser downlink precoders for CCI cancellation by block diagonalization (BD) method implemented with standard QR decomposition. Letting $\overline{\boldsymbol{H}}_k=[\boldsymbol{H}_1^{\mathrm{H}}\cdots\boldsymbol{H}_{k-1}^{\mathrm{H}}\boldsymbol{H}_{k+1}^{\mathrm{H}}\cdots\boldsymbol{H}_K^{\mathrm{H}}]^{\mathrm{H}}$ denote the aggregate channel matrix of the k-th user, the perfect CCI cancellation is equivalent to constructing the null space of $\overline{\boldsymbol{H}}_k$. The zero-interference constraint is given by

$$\overline{\boldsymbol{H}}_k\boldsymbol{W}_k=0 \qquad \forall k=1,2,\cdots,K \quad (16)$$

Note that for an $m\times n$ matrix $\boldsymbol{\Phi}$ with $m\leqslant n$, we have $\boldsymbol{\Phi}(\boldsymbol{I}-\boldsymbol{\Phi}^\dagger\boldsymbol{\Phi})=0$. The precoding matrix $\boldsymbol{W}_k$ can be constructed as a linear combination of the column basis vectors of $(\boldsymbol{I}-\overline{\boldsymbol{H}}_k{}^\dagger\overline{\boldsymbol{H}}_k)$ obtained by the standard QR decomposition. With $M_t=\sum_{k=1}^{K}N_k$, the QR decomposition of $(\boldsymbol{I}-\overline{\boldsymbol{H}}_k{}^\dagger\overline{\boldsymbol{H}}_k)$ can be expressed as

$$\boldsymbol{I}-\overline{\boldsymbol{H}}_k{}^\dagger\overline{\boldsymbol{H}}_k=\boldsymbol{\Theta}_k\boldsymbol{V}_k=[\boldsymbol{\Theta}_k\quad\widetilde{\boldsymbol{\Theta}}_k]\begin{bmatrix}\boldsymbol{V}_k\\ \boldsymbol{0}\end{bmatrix} \quad (17)$$

where $\boldsymbol{V}_k\in\mathbf{C}^{N_k\times M_t}$ is an upper triangular matrix and $\boldsymbol{\Theta}_k\in\boldsymbol{U}(M_t,N_k)$. $\boldsymbol{U}(M_t,N_k)$ is the collection of $M_t\times N_k$ complex matrices with unit-norm orthogonal columns. Without eigenmode selection performed by extra transmit antennas, we can write the precoding matrix $\boldsymbol{W}_k$ as

$$\boldsymbol{W}_k=\boldsymbol{\Theta}_k \quad (18)$$

Given the precoders $\{\boldsymbol{W}_k,k=1,2,\cdots,K\}$, the constraint (16) is satisfied and the CCI at each user is perfectly eliminated. Consequently, the multiuser MIMO channel is decomposed into parallel noninterfering single-user MIMO links and $\widetilde{\boldsymbol{H}}_k=\boldsymbol{H}_k\boldsymbol{\Theta}_k$ is the equivalent channel of the k-th user.

Without eigenmode selection, $\boldsymbol{W}_k$ with unit-norm orthogonal columns is a function of $\{\boldsymbol{H}_j\mid j\neq k,j=1,2,\cdots,K\}$ and independent of $\boldsymbol{H}_k$. Also, each entry of the equivalent channel $\widetilde{\boldsymbol{H}}_k$ is a linear combination of the rows of $\boldsymbol{H}_k$ and the columns of $\boldsymbol{\Theta}_k$. When $\boldsymbol{H}_k$ is a complex Gaussian matrix with i. i. d. entries, the distribution of $\widetilde{\boldsymbol{H}}_k$ remains the same as the one of $\boldsymbol{H}_k$ because the linear operations of Gaussian random variance are still Gaussian. Hence, the distribution of the equivalent channel $\widetilde{\boldsymbol{H}}_k$ is given by Eq. (14), where $\{\lambda_n,n=1,2,\cdots,N_k\}$ are the eigenvalues of $\widetilde{\boldsymbol{H}}_k$.

In multiuser systems, there are K independent two-state ON-OFF Markov traffic sources respectively corresponding to the K users, modeled by different transition metrics and Poisson rates $\{\boldsymbol{Q}_k,\alpha_k\}_{k=1}^{K}$. The GE limit of the arrival process for each user is given by Eq.(12). After the downlink precoding, each user performs on the parallel independent channel and does not interfere with each other. Then the accumulative service process in $[0,t)$ is divided into K parallel subprocesses corresponding to K traffic sources, respectively. Therefore, the GE limit of the accumulative service process for each user can be expressed by Eq. (13) except that the expectation operation is with respect to the equivalent channel matrix $\widetilde{\boldsymbol{H}}_k$.

It is finally proven that the frame length optimization problem of multiuser systems is an extension of the single-user case. Thus, the optimal frame length region is defined as the intersection of all K frame length regions of each user,

$$T = \cap_{k=1}^{K} \{T_1, T_2, \cdots, T_K\} \tag{19}$$

where T_k is determined by Eqs. (6) and (8). The optimal frame length of multiuser systems is given by

$$T^* = \max\{T_1^*, T_2^*, \cdots, T_K^*\} \tag{20}$$

where $T_k^*, k = 1, 2, \cdots, K$ is calculated by Eq. (10).

3 Numerical Results and Simulation

In this section, we assess the delay-bound violation probability performance of the proposed frame length optimization scheme. According to Section 2.2, the frame length optimization of multiuser systems is an extension of the single-user case after the CCI cancellation precoding. A point-to-point MIMO system with block Rayleigh fading channels for performance evaluation is considered, where the channel coherence time $t_c = 100$ ms. According to the aforementioned analyses, the optimal frame length depends on the following parameters: 1) System-level parameters, including the number of transmit and receiving antennas M_t, N_k, channel bandwidth B, the overhead of frame δ and the SNR per receiving antenna under unity channel gain p; 2) Traffic parameters of two-state ON-OFF Markov sources, including the Poisson arrival rate α_k, the packet size N_b, and the transfer probabilities $q_{k,1}$, $q_{k,2}$; 3) Delay-QoS requirements, including the delay bound $D_k^{\max}$ and the maximum tolerable delay-bound violation probability ε_k. Tab. 1 summarizes values of all parameters used in simulations, and default values of the varying parameters are shown between the parentheses. We vary traffic parameters, delay-QoS requirements and SNR values respectively to draw corresponding curves of delay-bound violation probability vs. frame length. Each sample of the curve is an average of 1 000-run values.

Tab. 1 Simulation parameters

Parameter	Value (default)
p/dBm	20 to 40(27)
α_k/(packet · frame^{-1})	0 to 300(100)
$q_{k,1}$	0 to 1(0.8)
$q_{k,2}$	0 to 1(0.4)
$D_k^{\max}$/ms	50 to 300(100)
ε_k	10^{-6} to 10^{-2}(10^{-3})
M_t	4
N_k	2
B/MHz	20
δ/ms	0.08
N_b/B	1 024

3.1 Effect of traffic source parameters

We validate the effect of traffic source parameters on delay-bound violation probability as a function of frame length. The theoretical values of the optimal frame length are numerically calculated by Eq. (10), as shown in Tab. 2.

Tab. 2 Traffic parameters and theoretical optimal frame length values

α_k/(packet · frame^{-1})	$q_{k,1}$	$q_{k,2}$	T_k^*/ms
100	0.8	0.4	10.9
150	0.8	0.4	20.4
200	0.8	0.4	31.3
100	0.6	0.4	12.3
100	0.8	0.6	14.9

Fig. 2 shows the delay-bound violation probabilities vs. the frame lengths for five sets of traffic source parameters. According to Fig. 2 and Tab. 2, the differences between the simulation and theoretical results are negligible in all cases with $\varepsilon_k = 1.0^{-3}$. First, the required optimal frame length with fixed $q_{k,1}$ and $q_{k,2}$ becomes larger as α_k increases. Secondly, there is a positive correlation between the optimal frame length T_k^* and $q_{k,2}$, but a negative correlation between T_k^* and $q_{k,1}$.

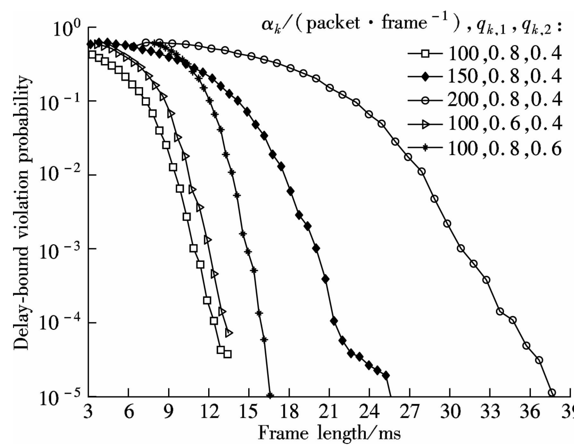

Fig. 2 Delay-bound violation probability vs. frame length of traffic source parameters

3.2 Effect of delay-QoS requirements

The curves of the delay-bound violation probability vs. the frame length with different delay-QoS requirements are shown in Fig. 3 and Fig. 4, while the theoretical results are given in Tab. 3. Clearly, the theoretical results and simulation agree reasonably well. As expected, the optimal frame length becomes larger along with a smaller ε_k and a fixed $D_k^{\max}$ since a smaller ε_k represents the more stringent delay requirement. Therefore, larger frames are chosen to transmit more packets and improve the system throughput. Furthermore, constraints (6) and (8) indicate that services with the same $\ln\varepsilon_k(D_k^{\max})^{-1}$ share the same optimal frame length in this scenario, regardless of different $D_k^{\max}$ and ε_k. For example, voice conference services with $D_k^{\max} = 100$ ms and $\varepsilon_k = 1.0^{-2}$ and cached streaming services with $D_k^{\max} = 300$ ms and $\varepsilon_k = 1.0^{-6}$, such as e-mail, both require the frame length of 8.9 ms to satisfy their delay-QoS constraints. As a result, the optimization needs to operate only once among these services.

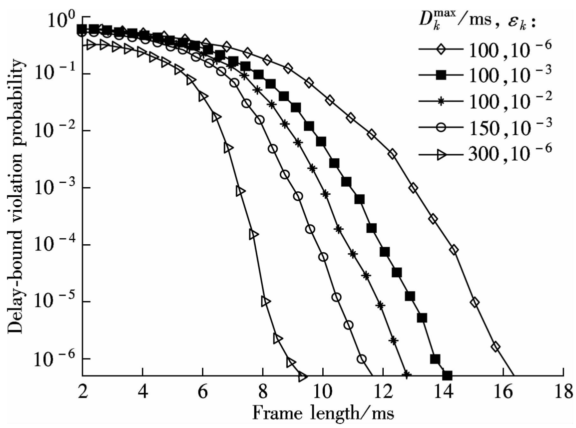

Fig. 3 Delay-bound violation probability vs. frame length of different $D_k^{\max}$ and ε_k

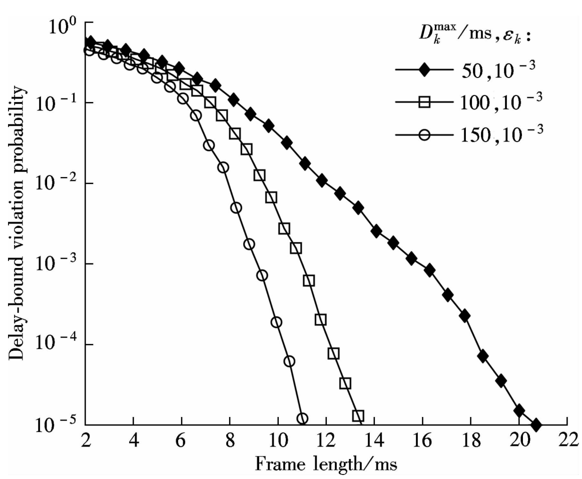

Fig. 4 Delay-bound violation probability vs. frame length of different $D_k^{\max}$

Tab. 3 Delay-QoS requirements and theoretical optimal frame length values

$D_k^{\max}$/ms	ε_k	T_k^*/ms
50	10^{-3}	15.8
100	10^{-6}	15.8
100	10^{-3}	10.9
100	10^{-2}	8.9
150	10^{-3}	8.9
300	10^{-6}	8.9

Fig. 4 draws the curves of the delay-bound violation probability for different $D_k^{\max}$ with $\varepsilon_k = 10^{-3}$. In accordance with the theoretical results in Tab. 3, different $D_k^{\max}$ results in different values of T_k^* and the frame length required to guarantee constraint (3) decreases with the increase in $D_k^{\max}$. Since the larger delay bound refers to the less stringent delay requirement, small frames can satisfy the less delay-sensitive traffic transmission. In Fig. 4, note that the delay-bound violation probability decreases almost linearly vs. the frame length above 6 ms with $D_k^{\max} = 100$ ms and $D_k^{\max} = 150$ ms.

3.3 Delay-bound violation probabilities of different SNR

In order to validate the effects of SNR on the proposed frame length optimization scheme, the curves of the delay-bound violation probability vs. frame length are plotted with different SNR values in Fig. 5. A highly acceptable agreement can be observed between theoretical results and the simulation. Similar to the results in Tab. 4, the lower SNR results in a larger frame, given certain delay-bound violation probability constraints. As the frame length changes, delay-bound violation probabilities of the higher SNR become more sensitive than those of the lower SNR due to larger frames can be more efficient than smaller frames in consideration of the overhead.

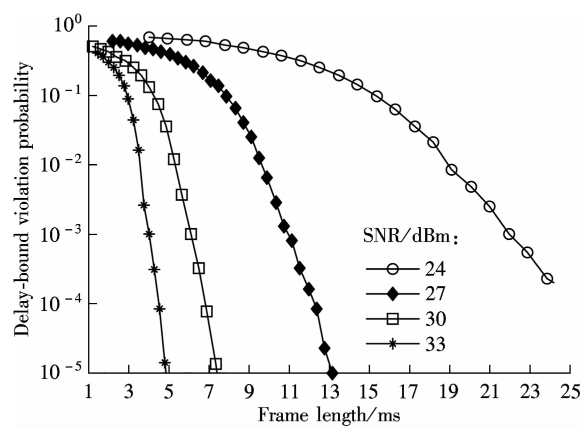

Fig. 5 Delay-bound violation probability vs. frame length of different SNR

Tab. 4 SNR values and theoretical optimal frame length results

p/dBm	T_k^*/ms
24	22.2
27	10.9
30	6.2
33	4.0

4 Conclusion

In this paper, a frame length optimization scheme is proposed for multi-antenna downlink systems to guarantee diverse delay-bound violation probability constraints. Concepts of effective bandwidth and effective capacity are applied to characterize the traffic source pattern and the queuing service dynamics. Different link-layer delay-QoS requirements are converted into PHY-layer data rate constraints. The cross-layer optimization of frame length is expressed as functions of traffic source parameters, delay requirements and data rates. The proposed scheme optimizes frame lengths of single-user and multiuser systems over flat Rayleigh fading channels. Theoretical analyses and simulation results show that delay-bound violation probability constraints can be well satisfied with the optimal frame lengths. Since the optimal frame length is designed without traffic scheduling, it is of interest to develop optimization schemes under overloaded scenarios.

References

[1] Trifonov P. Optimization of the downlink frame length in adaptive systems [C]//2005 *IEEE 7th CAS Symposium on Emerging Technologies: Circuits and Systems for 4G Mobile Wireless Communications*. St. Petersburg, Russia, 2005: 50 – 53.

[2] Keum Seung Won, Kim Jeong Gon. Decision of frame size and radio resource allocation for OFDMA/TDD-based next generation cellular systems [C]//2007 *International Conference on Convergence Information Technology*. Gyeongju, Korea, 2007:984-989.

[3] Wu Yi, Tang Zhiqiang, Larsson E G. Optimization of frame length in OFDMA systems taking into account the control signaling cost [C]//2011 *IEEE 73rd Vehicular Technology Conference (VTC Spring)*. Yokohama, Japan, 2011: 1 – 5.

[4] Lettieri P, Srivastava M B. Adaptive frame length control for improving wireless link throughput, range, and energy efficiency [C]//*IEEE 17th Annual Joint Conference of the IEEE Computer and Communications Societies*. San Francisco, CA, USA, 1998: 564 – 571.

[5] Iqbal A, Qureshi H K, Saleem M, et al. On the tradeoff between reliability and frame length in wireless networks [C]//*Third International Conference on Next Generation Mobile Applications, Services and Technologies*. Cardiff, Wales, 2009: 265 – 270.

[6] Chaudhary M H, Scheers B. Progressive decentralized TDMA based MAC: joint optimization of slot allocation and frame lengths [C]//2013 *IEEE Military Communications Conference*. San Diego, CA, USA, 2013: 181 – 187.

[7] Wu Dapeng, Negi R. Effective capacity: a wireless link model for support of quality of service [J]. *IEEE Transactions on Wireless Communications*, 2003, **2**(4): 630 – 640.

[8] Davy A, Meskill B, Domingo P J. An empirical study of effective capacity throughputs in 802. 11 wireless networks [C]//2012 *IEEE Global Communications Conference*. Anaheim, CA, USA, 2012: 1770 – 1775.

[9] Anwar A H, Seddik K G, ElBatt T, et al. Effective capacity of delay constrained cognitive radio links exploiting primary feedback [C]//*11th International Symposium on Modeling & Optimization in Mobile, Ad Hoc & Wireless Networks*. Tsukuba, Japan, 2013: 412 – 419.

[10] Chang C S. *Performance guarantees in communication networks* [M]. Berlin, Germany: Springer-Verlag, 2000.

[11] Harsini J S, Lahouti F. Adaptive transmission policy design for delay-sensitive and bursty packet traffic over wireless fading channels [J]. *IEEE Transactions on Wireless Communications*, 2009, **8**(2): 776 – 786.

[12] Kuo Ping Heng, Smith P J, Garth L M. Joint density for eigenvalues of two correlated complex Wishart matrices: characterization of MIMO systems [J]. *IEEE Transactions on Wireless Communications*, 2007, **6**(11): 3902 – 3906.

基于超时违约概率约束的多天线系统下行帧长优化

谭雨凤　李俊超　夏玮玮　沈连丰

(东南大学移动通信国家重点实验室, 南京 210096)

摘要:为了满足数据传输的时延约束,提出一种多天线系统的下行帧长优化方案. 由于传统的物理层信道模型在评估服务质量方面的局限性,该方案应用有效带宽与有效容量概念从链路层对源数据模式与队列服务动态进行统计特性建模. 该链路层模型分别描述了源数据产生过程、信道服务过程与服务质量参数之间的映射关系,将超时违约概率约束转化为数据传输速率限制,将帧长优化问题转换为求解多元联合不等式组. 在块衰落信道场景中,求解满足多元联合方程的帧长域下界,得到超时违约概率约束下的最佳系统帧长. 理论分析和仿真结果表明,在最佳帧长设置下,用户数据传输的超时违约概率满足要求.

关键词:超时违约概率;帧长优化;有效带宽;有效容量;多天线系统;服务质量

中图分类号:TN92