Quality-optimized Downlink Scheduling for Video Streaming Applications in LTE Networks

Xiaolin Cheng, Prasant Mohapatra
Department of Computer Science, University of California at Davis, CA 95616
{xlcheng, pmohapatra}@ucdavis.edu

***Abstract*—As the next generation of all-IP mobile communication system, LTE offers unprecedented data transmission speed and low latency for a variety of applications and services. However efficient QoS provisioning for wireless networks is challenging due to unreliable and resource-constrained radio interface. In this paper, we investigate the important downlink scheduling problem in LTE networks with a focus on video streaming applications. Unlike the conventional scheduling rules which exploit the network layer metrics, our scheme is directly targeted on optimizing the application-layer video quality within the required end-to-end delay bound. The video quality optimized scheduling is formulated as a complex combinatorial optimization problem with an exponentially-growing search space. To solve this complex optimization problem, we exploit the GA (genetic algorithm) based metaheuristic approach. The intrinsic strength of population based solution of GA offers a superior advantage for this type of scheduling problems. Performance of the proposed crosslayer design and the GA solution is evaluated against the well-known M-LWDF scheduling rule and a trajectory method. The simulation results demonstrate the effectiveness of the GA based quality-optimized approach. It can enhance the video quality significantly and satisfy the delay bound.**

I. Introduction

4G wireless systems such as 3GPP LTE (Long Term Evolution) [1] features high data rate and low end-to-end latency which are the key requirements of multimedia applications, especially video. According to Cisco's forecast [2], two-thirds of the world's mobile data traffic will be video by 2015. LTE will boost the proliferation of video applications. However, it is challenging to provision QoS for video and maintain designed system performance given limited radio resources, unreliable radio propagation channel and high user demands.

Scheduling has been an important aspect of QoS support in wireless networks. Wireless scheduling has two particular characteristics which distinguish it from conventional wireline scheduling [3]: (1) The radio channel is unreliable and error-prone. Errors are bursty in nature during which packets cannot be successfully transmitted on radio link. Good scheduling algorithms need to be channel state adaptive. (2) Channel state varies randomly in time on both slow and fast time scale. An efficient scheduling algorithm should take advantage of this by giving preference to a user with good channel.

In OFDMA based LTE networks, data is transmitted on a large number of parallel, narrow-band sub-carriers. During each time slot, multiple users can be allocated a set of sub-carriers to have concurrent transmissions. The efficient scheduling of radio resources (termed resource blocks in LTE) is crucial to achieve high network performance.

In a resource allocation period, each resource block is associated with different channel quality (characterized by Signal-to-Noise-Ratio or SNR) which can be sent back on a feedback channel known as CQI (channel quality indicator) from mobile terminals. Based on the CQI, appropriate modulation and coding schemes can be applied to be channel adaptive and improve the transmission reliability and rate (known as AMC or adaptive modulation and coding).

BER (bit error rate) in a packet can be approximated based on SNR and AMC scheme. We use the rate-distortion model based method to estimate video quality based on bit rate and BER. The target of scheduling resource blocks is to optimize video quality within delay bounds. It can be formulated as a combinatorial optimization problem. To solve this optimization problem with a complex objective function, multiple constraints, and an intractable search space, we exploit the genetic algorithm (GA) based approach and obtain satisfactory user-end video quality and delay performance. Using genetic operators, GA has an inherent advantage of handling multiple solutions during each iteration to explore diverse landscapes and avoid local sub-optima in search space.

The rest of the paper is organized as follows. In Section II, we describe the detailed system model. In Section III, we formulate the combinatorial optimization problem for video quality optimized scheduling in LTE. Section IV proposes our GA-based solution and Section V presents the numerical results. Related work is discussed in Section VI. Section VII concludes the paper.

II. System Model

A. LTE Downlink Model

The LTE downlink transmission scheme [4] provides scalable bandwidth from 1.4MHz to 20MHz with a sub-carrier spacing $\triangle f$=15kHz. A radio frame is $10ms$ in duration, and divided into 10 equally sized sub-frames (each being $1ms$ long). Each sub-frame is called an *Transmission Time Interval* (TTI), and further divided into 2 slots ($0.5ms$ each). The transmitted downlink signal consists of N_{BW} sub-carriers for a duration of T_{slot}. It can be represented by a resource grid as depicted in Figure 1. A radio *resource block* (RB) is defined as one slot in the time domain ($0.5ms$) and 12 consecutive sub-carriers (180KHz) in the frequency domain. A resource block is the smallest element of resource allocation assigned by the base station scheduler.

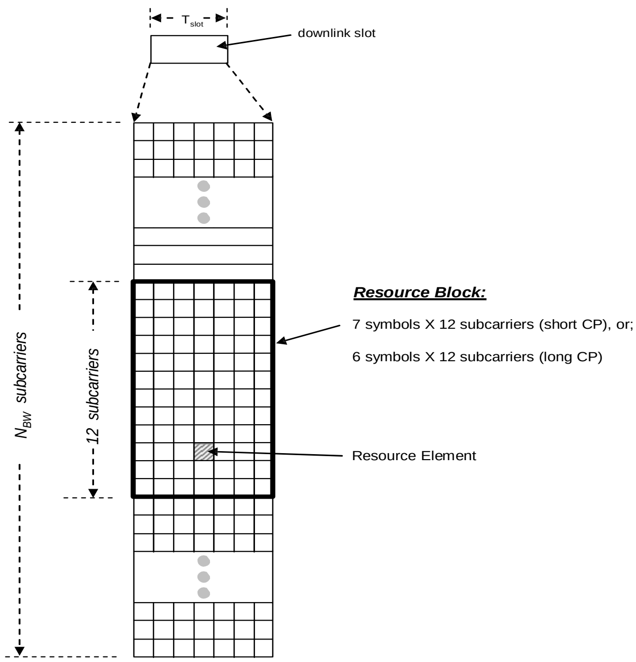


Fig. 1. LTE Downlink Resource Grid

B. Scheduler in LTE

LTE is based on OFDM, so it is possible to distribute available transmission radio resources in frequency domain to different mobile terminals. Resource allocations can be changed dynamically. MAC scheduler in eNodeB (base station) assigns both uplink and downlink radio resources.

3GPP does not specify the MAC scheduler, but leaves its design and implementation to vendors. This flexibility would impact the system performance significantly with different scheduling algorithms. Depending on the implementation, the scheduler can base its scheduling decision on the QoS class and the queuing delay of data, on the instantaneous channel conditions, or on fairness indicators. The channel conditions in a wide-band system vary in both time domain and frequency domain. If the mobile terminal can provide sufficiently detailed channel-quality information to the eNodeB, the scheduler can perform channel-dependent scheduling in time and frequency domains to improve system capacity and performance.

C. Channel Model

We consider a slowly-varying flat-fading channel. Its quality can be captured by a single parameter, namely received SNR (Signal-to-Noise Ratio) γ. The general Nakagami-k model is adopted to describe γ statistically [5]:

$$p_\gamma = \frac{k^k \gamma^{k-1}}{\overline{\gamma}^k \Gamma(k)} \exp(-\frac{k\gamma}{\overline{\gamma}}) \tag{1}$$

where $\overline{\gamma}$ is the average received SNR, Γ is the Gamma function, and k is the Nakagami-k fading parameter ($k \geq 0.5$). The Nakagami channel model applies to a large class of fading channels. It includes the Rayleigh channel as a special case when $k = 1$. In this paper, we use the Rayleigh model to characterize the downlink channel.

D. Adaptive Modulation and Coding

Adaptive modulation and coding (AMC) has been adopted in LTE to enhance the system throughput. It is one of the most important techniques of link adaptation [6]. Its objective is to maximize the data rate by adjusting transmission parameters to channel variations. When channel quality is good, AMC schemes with larger constellation sizes and higher channel coding rate can be applied to effectively achieve high transmission rate. When the quality of channel conditions is poor, transmission rate is reduced to ensure transmission quality.

TABLE I
AMC MODES AT THE PHYSICAL LAYER

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
Modulation	QPSK	QPSK	16-QAM	16-QAM	64-QAM
Coding Rate	1/2	3/4	9/16	3/4	3/4
Rate(bits/sym.)	1.00	1.50	2.25	3.00	4.50
a_m	0.3351	0.2197	0.2081	0.1936	0.1887
b_m	3.2543	1.5244	0.6250	0.3484	0.0871
SNR Threshold (dB)	1.7	4.5	7.2	10.7	16.1

Table I lists all the available AMC modes [7] adopted in this paper, with each mode consisting of a pair of modulation scheme and FEC coding scheme. Based on the parameters in the Table, the following equation is employed to approximate *Bit Error Rate* (BER):

$$BER_m = \frac{a_m}{e^{\gamma \times b_m}} \tag{2}$$

E. Rate-Distortion Based Video Quality Estimation

For video coding and communications, rate-distortion theory describes the relationship between the bit rate R and the achieved distortion D. The rate distortion region for a memoryless *i.i.d.* Gaussian source with the square error distortion measure was first introduced in [8]:

$$R(D) = \begin{cases} \frac{1}{2}\log_2(\sigma_x^2/D), & \text{if } D \leq \sigma_x^2 \\ \\ 0, & \text{if } D > \sigma_x^2 \end{cases} \tag{3}$$

where σ_x^2 is the variance of the source.

In our problem, during each scheduling period, a resource block has a bit rate R and BER P_{BER}. We can approximately estimate the video distortion achieved by this block:

$$D = 2^{-2R} \times (1 - P_{BER})^R + (1 - (1 - P_{BER})^R) \tag{4}$$

It is a complex function of both bit rate and BER.

III. PROBLEM FORMULATION

We consider a single-cell (one base station or eNodeB) scenario where the downlink bandwidth is divided into M resource blocks (RBs). The base station serves N active users. We denote the set of all RBs by $\mathcal{M} = \{m | m = 1, 2, \cdots, M\}$, and the set of all users by $\mathcal{N} = \{n | n = 1, 2, \cdots, N\}$. During each scheduling slot, the base station could allocate m RBs (RBs are not necessarily contiguous) to user n, but each RB is assigned to at most one user. We denote the power set of $\mathcal{M}$ as $\mathcal{P}$. For $\forall p \in \mathcal{P}$, it is a set of RBs. We have $x_i^p = 1$ if and only if p is allocated to user i. The scheduling of resource blocks for multiple users can be formulated as a combinatorial optimization problem as follows. We use the similar notations to ones used in [9].

Given a set of resource blocks $\mathcal{M}$, a set of active users $\mathcal{N}$ and a set of available AMC schemes $\mathcal{A}$ during a scheduling period in a cell:

$$\textbf{Minimize}: \ \textbf{Max}_{i\in\mathcal{N}} \ Distortion(i) \tag{5}$$

subject to:

$$\forall \text{ RB } m \in \mathcal{M} : \sum_{m\in p, i\in\mathcal{N}} x_i^p \leq 1 \tag{6}$$

$$\forall \text{ user } i \in \mathcal{N} : \sum_{p\in\mathcal{P}} x_i^p \leq 1 \tag{7}$$

$$\forall \text{ user } i \in \mathcal{N}, p \in \mathcal{P} : x_i^p \in \{0,1\} \tag{8}$$

$$\forall \text{ RB } m \in \mathcal{M} : \text{AMC}_{RB_m} \in \mathcal{A} \tag{9}$$

$$\forall \text{ user } i \in \mathcal{N} : \text{delay}_i \leq \text{DELAYBOUND}_i \tag{10}$$

Constraint (6) requires each RB be assigned to at most one user. Constraint (7) requires each user get no more than one set of RBs. If we assume N is the number of users in the cell, M is the number of available RBs and do not consider delay bounds, then we can estimate the number of feasible schedules during each scheduling period. For each RB, we could assign any one of N users, so this number would be N^M. Clearly the search space of this problem is huge (it grows exponentially with respect to the number of RBs). For example, if we have $M = 15$ and $N = 10$, then we may have 10^{15} feasible schedules. If evaluating each feasible schedule takes 10^{-9} second, the exhaustive search would need 11.57 days which is completely intractable.

It has been proved in [9] that optimal scheduling in LTE networks is NP-hard, so it is infeasible to construct a polynomial-time algorithm with less complexity to solve the complex optimization problem (5). A more practical approach would be designing an efficient heuristic.

IV. Genetic Algorithm Based Approach

Since the problem (5) has an exponentially-growing search space, it is not practical to pursue an optimal solution in an efficient way (with polynomial time complexity). The best strategy to address the problem is to view it as a "black-box" optimization problem and explore an effective metaheuristic approach [10]. Genetic Algorithm (GA) is particularly suitable for addressing this type of complex combinatorial scheduling problems [11]. GA is a population based metaheuristic inspired by the "survival-of-the-fittest principle". It has the particular strength of dealing with a set of solutions (i.e., a population) at each step, rather than working with a single and current solution, providing a natural and powerful way for exploring the search space. At each iteration, a number of genetic operators are applied to the individuals of the current population in order to generate individuals for the next generation. The "survival-of-the-fittest" principle ensures that the overall quality of the population improves as the algorithm progresses from one generation to the next. Figure 2 shows the flowchart of the GA based approach.

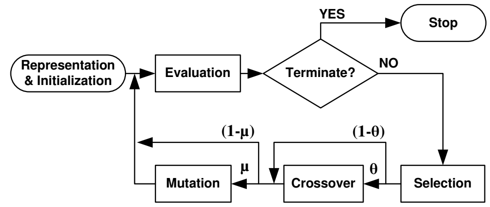


Fig. 2. The flowchart of the GA based approach

A. Solution Representation and Initialization

In order to encode a feasible solution in the genetic format, we need to define a gene first and then map a solution to a sequence of genes (i.e., a chromosome). Such encoding should be suitable for fitness computation (which is determined by the objective function) and genetic operations.

In our case, we have N users and M resource blocks. A natural encoding scheme would be to define a resource block as a gene. Then a particular sequence of resource blocks where each block may be allocated to different users can be represented as a chromosome. Figure 3 gives an example of an individual which is a feasible schedule of 7 RBs for 3 users.

Before entering the loop in Figure 2, we need to generate an initial population of individuals, i.e., a set of initial solutions. In our case, the initial population is a set of randomized sequences of M RBs for N users.

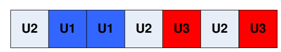


Fig. 3. An individual (a feasible scheduling solution)

B. Evaluation

The fitness function $f(x)$ is directly related to the objective function of the problem (5), the higher the fitness value, the better the individual. Since the objective is to minimize the video distortion, we adopt a fitness function as the inverse of the distortion value, i.e., $f(x) = 1/D(x)$. To be more application aware, we add more weight when transmitting a I-frame packet. An I frame is the key frame in each group of pictures. It is self encoded and does not need prediction information from other frames, so it is very important to the reconstructed video quality.

C. Selection

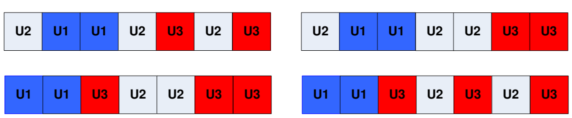
During this operation, we select individuals that have a better chance or potential to produce "good" offsprings in terms of their fitness values. With a proper selection operation, "good" genes among the population are more likely to be passed to the future generations.

Several selection schemes can be employed during this operation. A simple but efficient scheme known as *Tournament Selection* [11] randomly chooses m (tournament size) individuals from the population each time, and then selects the best of these m individuals in terms of their fitness values. By repeating this procedure multiple times, a new population can be selected. In this paper, we use tournament selection and tournament size is set to 2. To further weaken the selection pressure or avoid premature convergence where the search process falls into the local sub-optimum and cannot move toward the global optimal solution, probabilistic tournament selection can be used by specifying a probability for selecting the better fitting solution from the 2 competing solutions. If this probability is 1, the selection reduces to the conventional tournament selection. *Elitist selection* can be added to ensure that the very best solutions are retained over generations.

D. Crossover

Crossover mimics the genetic mechanism of reproduction in the natural world, in which genes from parents are combined and passed to offsprings. Crossover may create new individuals, thus exposing the search process to a new area of the fitness landscape or search space.

In our problem, during each iteration, two individuals (schedules) can be picked up randomly to be parents. We choose a RB randomly. This RB would be a "switching-point" where two individuals exchange part of their "chromosome". Figure 4 is an example of crossover operation. In Figure 4(a) two individuals are chosen to be applied crossover operation. The 4th RB is the "switching-point" in this case. In Figure 4(b), the 5th, 6th and 7th RBs of two individuals are exchanged, resulting in two new individuals (schedules).



(a) Parents (b) Children

Fig. 4. Genetic Crossover Operation

E. Mutation

The objective of the mutation operation (applied with a rate μ) is to diversify the genes of the current population, which helps prevent the solution from being trapped in a local optimum. This is a significant advantage over traditional heuristic methods. To apply mutation operation to our individual, we randomly pick a RB and allocate it to a different user than it was in the last scheduling period. Figure 5 is an example: the 4th RB is changed from user2 to user3.

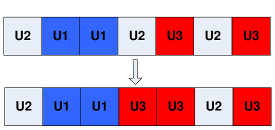


Fig. 5. Genetic Mutation Operation

F. Termination

As shown in Figure 2, a termination criterion needs to be set for GA to stop its iterating. In practice, the termination condition could be based on the total number of iterations (generations), maximum computing time, a threshold of desired fitness values, or the fitness values converge and do not change over a certain number of iterations.

V. Simulation Results

A. Comparison with M-LWDF Scheduling Rule

Resource allocation and scheduling has been an important problem in wireless HDR (High Data Rate) systems. Numerous schemes have been proposed to address this problem. In this section we compare our GA based quality optimized scheduling approach with the well-known Modified Largest Weighted Delay First (M-LWDF) scheduling rule [12].

M-LWDF is based on the Proportional Fairness (PF) scheduling rule [13]. Let $\mu_i(t)$ be the state of the channel of user i at time t, i.e. the actual rate supported by the channel, and $\overline{\mu}_i(t)$ denote the average historical rate of user i at time t. Then the PF rule can be defined as:

$$j = \arg\max_i \frac{\mu_i(t)}{\overline{\mu}_i(t)} \tag{11}$$

Compared to the Max-Rate rule [14] where a user with the highest instantaneous rate is scheduled, the PF rule balances the user requests by considering their historical rates. M-LWDF was proposed to further accommodate the delay requirement. If we denote $W_i(t)$ to be the waiting time of user i at time t, then the M-LWDF rule can be defined as:

$$j = \arg\max_i a_i W_i(t) \frac{\mu_i(t)}{\overline{\mu}_i(t)} \tag{12}$$

We use MATLAB to simulate the LTE downlink system model and implement our quality optimized scheduling scheme. We use the standard 400-frame *foreman* video clip which is encoded into a 1000Kbps MPEG4 video stream using FFMPEG [15]. The frame rate is 25 frame/sec, and every 10 frames have an I frame. The compressed video clip is streamed to 40 users in a single cell. The channel average SNR $\overline{\gamma}$ is 20dB. The bandwidth is 1.25MHz with 6 radio resource blocks. The delay requirement is 25ms. For GA parameters, the crossover rate is 0.8; the mutation rate is 0.1.

In Figure 6, we compare the PSNR values of received video stream obtained by GA with M-LWDF for a user. Clearly for most frames, GA has higher PSNR values than M-LWDF. Figure 7 presents the visual comparison of the image quality offered by GA with M-LWDF. GA provides a much more clear image of Frame 41 than M-LWDF. In addition, GA adds more weight on I frames which are crucial to the reconstructed video quality. Preserving I frames improves the video quality from the content-aware perspective. M-LWDF only considers the physical layer channel rate and waiting time which may not directly optimize the application layer video quality.

In Table II we compare the average PSNR, PLR (Packet Loss Ratio) and delay of two schemes. Our GA based approach produces lower PLR and delay which is consistent with the application layer quality results (PSNR values and reconstructed image quality).

In Figure 8, we plot the average delay of all users. They are all within the required delay bound. The fairness among all users is properly achieved.

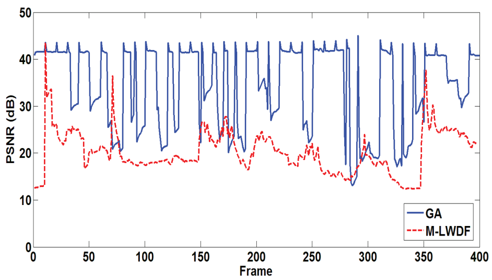


Fig. 6. PSNR Comparison Between GA and M-LWDF

(a) Frame 41 by GA (b) Frame 41 by M-LWDF

Fig. 7. Comparison of Image Quality Between GA and M-LWDF

B. Comparison with Simulated Annealing

In contrast to the population based GA approach where a set of solutions are manipulated in each iteration, trajectory methods are another broad class of metaheuristics which deal with a single solution in each step. To compare the GA-based approach with the trajectory methods, we implement simulated annealing (SA) which has been used for solving certain networking problems. SA was initially motivated by an analogy between the way a piece of metal cools and freezes into a minimum energy crystalline structure (annealing process), and the search for a minimum in a more general system [16]. When SA explores the solution space, it accepts a non-improving solution with a probability, which decreases with iterations. We use a probabilistic acceptance function:

$$Pr\{\bar{x} \rightarrow \hat{x}\} = \begin{cases} 1, & \text{if } D(\hat{x}) < D(\bar{x}) \\ \exp\{-\frac{|D(\hat{x}) - D(\bar{x})|}{T_k}\}, & \text{otherwise,} \end{cases} \tag{13}$$

where T_k is a control parameter analogous to temperature in a physical system, $\bar{x}$ is the current solution (allocation of resource blocks), $\hat{x}$ is a perturbation of $\bar{x}$, and $D(x)$ is the video distortion given a resource block allocation solution x. The fashion in which T_k is changed is called the *cooling schedule*. The following geometric cooling schedule is used in our simulations [16]:

$$T_{k+1} = \omega \cdot T_k \tag{14}$$

Nearly all transitions will be accepted at the initial stage of the search process. The control parameter is decremented every time when a non-improving solution is accepted, and remains at each value for a sufficient time for the system to "return to an equilibrium." ω is the decay coefficient. We set $\omega = 0.99$ for all simulations in this paper.

TABLE II
COMPARISON BETWEEN GA AND M-LWDF

	Average PSNR (dB)	Average PLR	Average Delay (ms)
GA	34.1	4.31%	20.4
M-LWDF	20.2	10.75%	35.8

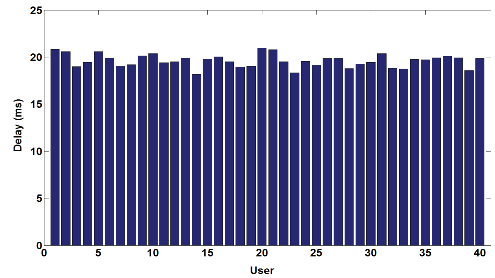


Fig. 8. Average Delay of All Users

We compare our GA based approach with the SA method using the same simulation settings of the comparison with the M-LWDF scheduling rule in the previous section. Figure 9 shows the PSNR values for a user generated by two schemes. Our solution has significantly higher PSNR values for the majority of the frames, therefore offering better video quality. In Table III the average statistics of three schemes are compared. One interesting observation is that SA method has a slightly higher packet loss ratio than M-LWDF but offers better video quality (higher average PSNR value). The results can be explained that the application layer metric (video quality) can be impacted by the network layer performance, but not directly related to the network layer metrics. From Figure 9 we can see although the video quality by SA method is worse than GA, it still successfully keeps and reconstructs more I frames than M-LWDF (see Figure 6) which helps improve the video quality. Application layer metric optimized scheduling is more efficient and effective in enhancing the video quality.

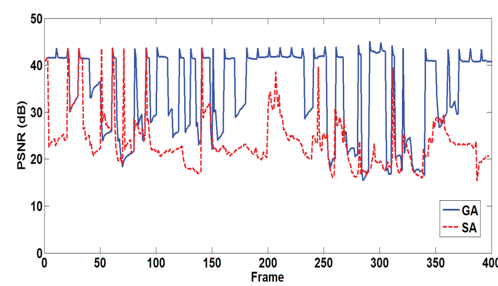


Fig. 9. PSNR Comparison Between GA and SA

TABLE III
COMPARISON OF PSNR AND NETWORK STATISTICS

	Average PSNR (dB)	Average PLR	Average Delay (ms)
GA	34.3	3.92%	19.7
SA	23.7	11.44%	22.6
M-LWDF	21.9	10.44%	35.5

In addition to providing better solutions, another strength of GA based approach over trajectory methods is that multiple "good" solutions can be found in each scheduling slot. Such extra good solutions can be used as alternative (or backup) schedules if needed.

C. Performance Consideration

A practical consideration for the GA-based scheduling scheme is its computational complexity. The population based approach of GA may have higher complexity. We presume it should not be an issue in its implementation. In LTE the scheduler runs on eNodeB (base station) which is computationally resourceful and powerful. The nature of GA provides itself a unique advantage of distributed and parallel computing applicability. Genetic operations are simple and can be easily implemented in parallel or distributively which are readily available in server cluster environments. In addition, as a scheduling framework, GA can be plugged into other specific scheduling rules for different QoS requirements. So it is scalable and extensible.

VI. Related Work

Scheduling and resource management in wireless networks has been an active research area over the years. Various schemes have been proposed to address this problem. In addition to the Max Rate, Proportional Fairness and M-LWDF rules discussed in Section V, other schemes have also been proposed with different focus. The Exponential rule [17] and the Log rule [18] aim at balancing delay and maximizing system throughput respectively.

For scheduling in LTE networks, in [19] the authors propose several approximation algorithms for downlink frequency domain packet scheduling. In [9], the authors present the hardness results on the LTE scheduling problem and prove that LTE uplink frequency domain scheduling problem is NP-hard. In [20], a quality-driven cross-layer approach was proposed for video delivery over LTE. In the first step, channel rate on each resource block, historical rates and delay requirement of multiple users are weighted and evaluated to determine the different priorities of allocating resources. In the second step, AMC and video coding parameters are selected to optimize the video quality. The scheduling step is essentially an implementation of M-LWDF and does not consider the video quality.

The potential of Genetic Algorithm in networking research has been recognized in recent years. GA has been explored to address various networking problems such as routing [21], scheduling [22] and buffer management [23] where it demonstrates effectiveness compared to the conventional solutions.

VII. Conclusion

In this paper, we investigate the downlink radio resource scheduling and allocation problem for video streaming applications in LTE networks from a crosslayer perspective. The application-layer video quality is considered as the basic scheduling criterion. The scheduling problem is formulated as a complex combinatorial optimization problem involving multiple constraints and an exponential search space which does not have an efficient polynomial-time solution. We use the rate-distortion theory based method to estimate the video quality and exploit genetic algorithm based metaheuristic approach to solve the optimization problem. Simulation results demonstrate that the proposed video quality optimized scheduling solution is superior to the existing well-known scheduling scheme and a trajectory method. It can effectively enhance the video quality at the receiver end while satisfying the delay requirements and achieving fairness among the users.

References

[1] 3GPP TS 36.300, "E-UTRA and E-UTRAN Overall Descrption Stage 2 (Release10)," Sep. 2011, v10.5.0.

[2] Cisco Visual Network Index: Global Mobile Data Traffic Forecast Update, 2011-2015, Online Whitepaper, Feb. 2010.

[3] S. Shakkottai and A. L. Stolyar, "Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR," in *in Proceedings of 17th International Teletraffic Congress (ITC-17*, Brazil, Dec. 2001, pp. 793–804.

[4] 3GPP TS 36.211, "E-UTRA Physical Channels and Modulation (Release9)," March 2010, v9.1.0.

[5] Q. Liu, S. Zhou, and G. B.Giannakis, "Cross-Layer Combining of Adaptive Modulation and Coding With Truncated ARQ Over Wireless Links," *IEEE Trans. on Wireless Communications*, vol. 3, no. 5, p. 1746, 1755 2004.

[6] M. Alouini and A. Goldsmith, "Adaptive Modulation over Nakagami Fading Channels," *Wireless Personal Communications*, pp. 119–143, May 2000.

[7] H. Luo, S. Ci, and D. Wu, "A Cross-layer Optimized Distributed Scheduling Algorithm for Peer-to-Peer Video Streaming over Multi-hop Wireless Mesh Networks," in *Proc. IEEE SECON*, Rome, Italy, June 2009, pp. 1–9.

[8] T. Berger, *Rate-distortion theory: A mathematical basis for data compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[9] H. Yang, F. Ren, C. Lin, and J. Zhang, "Frequency-Domain Packet Scheduling for 3GPP LTE Uplink," in *Proc. IEEE INFOCOM*, San Diego, CA, March 2010, pp. 1–9.

[10] C. Blum and A. Roli, "Metaheuristics in combinatorial optimization: overview and conceptual comparison," *ACM Computing Surveys*, vol. 35, no. 3, pp. 268–308, Sept. 2003.

[11] T. Back, D. Fogel, and Z. Michalewicz, Eds., *Handbook of Evolutionary Computation*. New York, NY: Oxford University Press, 1997.

[12] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "Scheduling in a Queueing System with Asynchronously Varying Service Rates," *Probability in the Engineering and Informational Sciences*, vol. 18, no. 2, pp. 191–217, 2004.

[13] A. Jalali, R. Padovani, and R. Pankaj, "Data Throughput of CDMA-HDR a High Efficiency-High Data Rate Personal Communication Wireress System," in *Proc. IEEE VTC*, Tokyo , Japan, May 2000.

[14] B. S. Tsybakov, "File Transmission over Wireless Fast Fading Downlink," *IEEE Transactions on Information Theory*, 2002.

[15] http://ffmpeg.org/.

[16] E. Aarts and J. Korst, *Simulated Annealing and Boltzman Machines*. New York, NY: John Wiley & Sons, 1989.

[17] S. Shakkottai and A. L. Stolyar, "Scheduling for Multiple Flows Sharing a Time-Varying Channel: The Exponential Rule," *American Mathematical Society Translations, Series 2*, vol. 207, 2002.

[18] B. Sadiq, S. J. Baek, and G. de Veciana, "Delay-optimal opportunistic scheduling and approximations: The log rule," in *IEEE INFOCOM*, Rio de Janeiro, Brazil, April 2009, pp. 1692–1700.

[19] S.-B. Lee, S. Choudhury, A. Khoshnevis, S. Xu, and S. Lu, "Downlink MIMO with Frequency-Domain Packet Scheduling for 3GPP LTE," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, April 2009.

[20] H. Luo, S. Ci, D. Wu, J. Wu, and H. Tang, "Quality-Driven Cross-Layer Optimized Video Delivery over LTE," *IEEE Communications Magazine*, pp. 102–109, Feb. 2010.

[21] S. Mao, Y. T. Hou, X. Cheng, H. D. Sherali, and S. F. Midkiff, "Multipath routing for multiple description video in wireless ad hoc networks," in *Proc. IEEE INFOCOM*, Miami, FL, 2005.

[22] C. Ngo and V. Li, "Centralized broadcast scheduling in packet radio networks via genetic-fix algorithms," *IEEE Trans. Commun.*, vol. 51, no. 9, pp. 1439–1441, Sept. 2003.

[23] G. Fatta, F. Hoffmann, G. Re, and A. Urso, "A genetic algorithm for the design of a fuzzy controller for active queue management," *IEEE Trans. Syst., Man, Cybern. C*, vol. 33, no. 3, pp. 313–324, Aug. 2003.