Analysis of Downlink Scheduling for Network Coverage for Wireless Systems with Multiple Antenna

1Harish Kumar, 1Manish Kumar and 2Pushpneel Verma

1Department of ECE, 2Department of CSE,
Bhagwant Institute of Technology, Muzaffarnagar, India
E-mail: hc78@rediffmail.com, pushpneelverma@gmail.com

Abstract
In this paper, the focus on the downlink scheduling design optimized for network coverage for wireless systems with multiple antennas. & proposed a wireless system consisting of a base station (with \( n_T \) transmit antennas) and \( K \) client mobiles (each with single antenna). with multiple antennas, we proposed a systematic framework based on information the critical approach and formulate the scheduling design as a mixed convex and combinatorial optimization problem. network coverage is based on the maximum cell-radius such that the outage probability of a single user at the cell edge (at a target bit rate) is below a specified target \( P_{\text{out}} \). The scheduling algorithm in the MAC layer is responsible for the allocation of channel resource at every fading block. The system resource is partitioned into short frames. We assume time division duplexing (TDD) systems so that channel reciprocal holds. At the beginning of every frame, the base station estimates the channel matrix from the participating mobile users. The uplink channel estimation is used as the downlink channel information. was found that network coverage could also benefit from the multi-user selection diversity through wireless scheduling.

Index Terms: MIMO, Coverage-capacity tradeoff. Cross layer, Wireless

Introduction
It is well-known that cross-layer scheduling could achieve significant performance gain in network capacity due to multi-user selection diversity. Basically, system resource is allocated adaptively to user (s) with the best channel condition. There have been a lot of works that try to take advantage of the cross-layer optimization by cooperative scheduling which takes the link-level metrics into scheduling decisions which takes the link-level metrics into scheduling decisions [1]. In [2], [3], it is shown that maximizing the link diversity order (which maximizes the link capacity) does not always result in the optimal system capacity. Therefore, joint optimization to the system level performance of wireless systems and the scheduling design optimized for network capacity has been relatively well-studied.

While system capacity is an important measure to optimize. System coverage is also an important dimension to consider, especially during the initial deployment stage where network coverage is usually the bottleneck. The advantage of cross-layer scheduling on the network coverage is a relatively unexplored subject. Conventional concept of network coverage is based on the maximum cell-radius such that the outage probability of a single user at the cell edge (at a target bit rate) is below a specified target \( P_{\text{out}} \). In [4] the coverage performance of an uplink scheduling algorithm has been studied. It is found that network coverage could also benefit from the multi-user selection diversity through wireless scheduling. However, the scheduling algorithm considered is restricted to selecting one user at a time. Hence, the analysis does not generalize to the general case with multiple antennas where spatial multiplexing allows selecting multiple transmission at any scheduling slot. Moreover, the design of the scheduling algorithm is heuristic and it is not clear what is the optimal scheduling design with respect to network coverage. In this paper, we shall focus on the downlink scheduling design optimized for network coverage for wireless systems with multiple antennas.

In this paper we consider a wireless system consisting of a base station (with \( n_T \) transmit antennas) and \( K \) client mobiles (each with single antenna). With multiple antennas, we have additional degrees of freedom for spatial multiplexing and spatial diversity To include the spatial multiplexing into the framework, we first extend the conventional concept of network coverage to a more general utility-based network coverage to deal with the possibility of allocating resource to multiple users at the same time. We consider the network centric utility and the user centric utility as two examples of the utility based coverage concept. Raised on the generalized concept of network coverage, we propose a systematic framework based on information the critical approach and formulate the scheduling design as a mixed convex and combinational optimization problem. Due to the huge search space. The complexity of the optimal algorithm is enormous. We consider a genetic-based scheduling [5], which offers a reasonable complexity-performance tradeoff.

This paper is organized as follows. In section II. we shall outline the multi-user forward link physical layer model as well as the MAC layer model. In section III. we shall define the utility-based network coverage and formulate the downlink scheduling design optimized for network coverage. Optimal solution is outlined. In section IV. we shall introduce the genetic-based algorithm. In section V. we present numerical results to evaluate the performance of the optimal and genetic based schedulers. Finally, we conclude with a brief summary of results in section VI

[Page No. 15]
Overview of System Model with Multiple Transmit Antennas
We first elaborate the multi-user for ward link channel model with multiple antennas, the physical layer model and the MAC layer model below. To decouple the data source statistics from the system performance, we shall assume that source buffers are large in size so that they always contain source packets waiting to be transmitted. In other words, there will be no empty scheduling slots due to insufficient source packets at the buffer.

Downlink Channel Model
We consider a communication system with K mobile users having single receive antenna and a base station with nT transmit antennas. The microscopic channel fading between different users and different antennas is modeled as i.i.d. complex Gaussian distribution with unit variance. Furthermore, it is assumed that the encoding and decoding frame are short bursts which are much shorter than the coherence time of the fading channel.

Let Yk be the received signal of the k-th mobile. The K x 1 dimension vector of received signal Y at the K mobile stations is given by

\[
Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_K \end{bmatrix} = \begin{bmatrix} \sqrt{L_1} H_1 \\ \vdots \\ \sqrt{L_K} H_K \end{bmatrix} X + \begin{bmatrix} Z_1 \\ \vdots \\ Z_K \end{bmatrix}
\]

(1)

where X is the nT x 1 transmit symbol from the base station to the K mobiles, Z is the complex Gaussian noise with variance \(\sigma^2\), \(H_k\) is the 1 x nT dimension channel matrix between the nT transmit antennas (at the base station) and the k-th mobile and \(L_k\) is the path loss between the base station and the k-th mobile. The entries of \(H_k\) are i.i.d. zero mean complex Gaussian with unit variance. The path gain [6] is given by (2) at the top of the next page, where \(d_k\) denotes the distance between the base station an the k-th mobile, \(G_t, G_r\) are transmit and receive antenna gains and \(\lambda\) is the wavelength of the carrier.

Multi-user Physical Layer Model
Before we could discuss the scheduling optimization problem, it is very important to define the physical layer model because different physical layer implementations will definitely affect the system level performance. To isolate the physical layer performance from specific implementation details (such as channel coding and modulation, multiple access schemes like TDMA, FDMA, CDMA), and information theoretical approach is adopted Specifically, the maximum achievable rate at the physical layer (with arbitrarily low error probability) is given by the Shannon's capacity and is realized by random codebook and Gaussian modulation.

For simple processing, we shall adopt a simple downlink zero-forcing (ZF) scheme" at the base station [5], [7] as illustrated in Figure 1.

There are K streams of information data to K individual users at the base station transmitter. They are channel encoded independently. The vector \((K \times 1)\) of encoded symbols, \(U = [U_1, ..., U_K]\), are processed by the power control diagonal matrix \((K \times K)P = diag(p_1, ..., p_K)\) followed by the ZF matrix \((n_T \times K), W = [w_1, ..., w_K]\) where \(w_k\) is the \(n_T \times 1\) complex ZF weight of user k. Hence, the transmitted vector of symbol, X, is given by :

\[
X = w\sqrt{P}U = \sum_{k=1}^{K} \sqrt{p_k} U_k w_k
\]

(3)

where \(p_k \geq 0\) is the average transmit power during the current scheduling instance for user k and \(E[|U_k|^2] = 1\). Since encoding frame is short burst with quasi-static fading, no power adaptation within an encoding frame is needed. At any scheduling slot, Individual user(s) could be turned off by assigning \(p_k = 0\). an admissible set is defined as a set of user indices with non-zero transmit power \(A = \{k \in [1, K]: p_k > 0\}\). The total transmit power out of the base station at any scheduling slot is constrained by \(P_0\). That is:

\[
\sum_k P_k \leq P_0
\]

(4)

Calculation of the ZF Weights : Given an admissible set, the transmit power \(\{p_1, ..., p_K\}\), and a realization of the channel fading \(\{h_1, ..., h_k\}\), the received signal of user k is given by :

\[
Y_k = \sqrt{p_k} L_k h_k w_k u_k + \sum_{j=1, j \neq k}^{K} \sqrt{p_j} L_j h_j w_j \sum_{j=1}^{K} L_j h_j w_j U_j + Z_k
\]

(5)

Where the first term contains the desired signal and the middle term represents the multi-beam interference due to simultaneous transmission of independent information streams. The OTBF weight, \(w_k\) is selected to satisfy:

\[
w_k w_k^* = 1 \forall k
\]

(6)
And the orthogonal conditions

\[ h_j w_k = 0 \forall j \in A, j \neq k \]  

(7)

where \( A \) denotes admissible set of users with non-zero allocated power. The operator of * means complex conjugate transpose. Note that when \( p_k = 0 \), the information stream for user \( k \) is turned off. In other words, the number of simultaneous transmission is given by the cardinality of the admissible set \( A \). Observe that there are \( 2n_T \) degree of freedom in \( w_k \) and there are \( 2|A|-1 \) equation from the constraints (6) and (7).

Hence, we have:

\[ |A| \leq n_T \]  

(8)

This means that with \( n_T \) transmit antennas, the base station could support at most \( n_T \) spatial channels. The remaining degrees of freedom is utilized to maximize \( w_k^* (h_k^* h_k) w_k \). Please refer to [5] for the solution of the ZF weight.

Date rates supported by the Orthogonal Spatial Channels: With the ZF weights \( \{w_k\} \), the multi-beam interference becomes zero and there are \( |A| \) independent spatial channels. The received signal for mobile user \( k \) is given by:

\[ Y_k = \sqrt{p_k L_k} H_k w_k U_k + Z_k \]  

(9)

Hence, the maximum achievable date rate of the \( k \)-th spatial channel during the fading block is given by the maximum mutual information between \( U_k \) and \( Y_k \) and is given by:

\[ T_k = \log_2 \left( 1 + \frac{p_k L_k |h_k w_k|^2}{\sigma^2} \right) \]  

(10)

**MAC Layer Model**

The scheduling algorithm in the MAC layer is responsible for the allocation of channel resource at every fading block. The system resource is partitioned into short frames. We assume time division duplexing (TDD) systems so that channel reciprocal holds. At the beginning of every frame, the base station estimates the channel matrix from the participating mobile users. The uplink channel estimation is used as the downlink channel information. Due to short burst transmissions, the channel estimation is used as the downlink channel information. Due to short burst transmissions, the channel fading remains the same across the entire burst duration. The estimated CSI is passed to the scheduling algorithms in the MAC layer. The output of the scheduler consists of an admissible set. \( A = \{k \in [1, K]; p_k > 0\} \) (the set of user indices with non-zero power allocated at the current fading block), the corresponding power allocation \( \{p_k\} \) and the instantaneous rate allocation \( \{r_k\} \) of the selected users. The downlink payload is transmitted at the scheduled rate and the rate is also broadcast on the downlink common channels to mobile users.

**General Formulation of the Downlink Scheduler Design**

Before we proceed with the scheduler design, it is important to quantify what is meant by system performance. In multi user systems, the system performance can be defined by an instantaneous utility \( G(r_1, ..., r_K) \) where \( r_k \) is the instantaneous data rate of the \( k \)-th user. For meaningful optimization, we have:

\[ \frac{\partial G}{\partial r_k} > 0 \forall r_k \geq 0 \]  

(11)

**Utility-Based Network Coverage**

Consider a sequence of \( N \) realization of fading blocks, the instantaneous achievable data rate \( s \) of a scheduled mobile user \( s \) is random variables (functions of the specific fading realization) in that fading block. In conventional wireless systems where the scheduling is constrained to select one active user at any scheduling instance (fading block), outage is defined as the event that the instantaneous data rate of the scheduled user is below a target data rate \( R_0 \). Outage probability is defined as the likelihood of the outage event, 

\[ P_{\text{out}} (R_0) = \lim_{N \to \infty} N_{\text{outage}} (R_0) / N \]  

Traditional concept of network coverage is there fore defined as the maximum distance between a mobile and the base station such that the outage probability of the scheduled user (at a target bit rate, \( R_0 \)) is below a specified target \( P_{\text{out}} (R_0) \). That is, coverage is given by the maximum distance, \( d(P_{\text{out}}, R_0) \), such that

\[ \Pr [r(d) \leq R_0] \leq P_{\text{out}} \]  

(12)

where \( r(d) \) denotes the instantaneous data rate of any user at a distance \( d \) from the base station. This is a commonly employed definition in the cellular systems [6] as well as wireless systems with scheduler constrained to select one user at a time [4].

However, when the base station has multiple transmit antennas, there are multiple spatial channels as a result of the additional degrees of freedom. This implies that a general multiple antenna scheduler must allow more than one active transmission at any scheduling slot (fading block), Hence, before we could discuss the design of schedulers optimized for network coverage, we must extend the definition of coverage to include multiple active user transmissions at a time.
Coverage-Optimized Scheduling Problem Formulation

Observe that given a fixed target utility value $G_0$, the utility-based coverage is maximized if the worst case out-age probability is minimized. Hence, the coverage-optimized scheduler design is formulated below and is a mixed concave optimization and combinatorial search problem.

Examples of Utility Functions

The utility-based coverage is a general concept and depending on the specific forms of the utility function, the coverage can have very different physical interpretation. In this section we shall illustrate the concept of utility-based coverage using two examples, namely the network-centric coverage and the user-centric coverage.

Network Centric Coverage: The generalized coverage is called network centric if the utility function is given by:

$$
G_{network}(r_1, ..., r_K) = \sum_{k \in A} r_k
$$

(15)

The corresponding outage event will be interpreted as the situation when the total instantaneous sum throughput $\sum_k r_k$ corresponding coverage (defined based on the outage event) is not measuring what a single user gets but is measuring how the network delivers as a whole but on how much the worst user gets. Hence, the physical meaning of the outage and coverage is network centric.

User Centric Coverage: The generalized coverage is called user centric if the utility function is given by:

$$
G_{user}(r_1, ..., r_k) = \min_{k \in A} r_k
$$

(16)

In this case, the outage event is determined by the worst case user instead of the contribution from all selected users. The outage is measured not based on how much the network delivers as a whole but how much the worst user gets. Hence, the physical meaning of the outage and coverage is user centric.

Optimal Solution-Mixed Combinatorial and Convex Programming

Since the optimization problem involves minimizing $\text{Pr}[G(r_1, ..., r_K) \leq G_0]$, which is not very easy to deal with, we have the following lemma which establishes the equivalence between minimizing the worst-case outage and maximizing the utility function $G(r_1, ..., r_K)$.

Step 1: For every possible admissible set $A$, we compute the optimal power allocation $(p_1, ..., p_K)$ for those selected users.

Since $r_k$ is a function of weights $w_1, ..., w_K$ and $(p_1, ..., p_K)$ and since the weights are also functions of the admissible set $A$, we could express the utility function $G(r_1, ..., r_K) = G(A; p_1, ..., p_K)$ as:

$$
G(r_1, ..., r_K) = G(A; p_1, ..., p_K)
$$

(18)

The optimal power allocation with respect to the network utility, $G_{network}(r_1, r_2, ..., r_K)$, as well as the user-centric utility, $G_{user}(r_1, r_2, ..., r_K)$, are both given by:

$$
p^*_k = \left( \frac{1}{\lambda} - \frac{1}{L_k | h_k w_k |^2} \right)^+
$$

(19)

for all $k \in A$ and $\lambda$ is the Lagrange multiplier given by the solution of the equation $\sum_{k \in A} p^*_k = P_0$.

Step 2: Repeat step 1 to compute the utility functions for all other possible admissible set $A$.

For the network-centric utility, we have the following lemma about the admissible set $A$.

$$
G_{network}(r_1, ..., r_K) \text{ satisfies } |A| = n_T
$$

On the other hand, we have the following lemma about the admissible set $A$ with respect to the user-centric utility.

Genetic-Based Coverage-Optimized Scheduler Design

The computational complexity of the optimal algorithm in general exceeds the implementation limitation in most designs for moderate $K$ and $n_T$. In this section, we shall introduce a real-time genetic algorithm [5]. The main template of genetic algorithm is illustrated below.

Algorithm 1. Step 1-Initialization: Initialize a population with $N_P$ chromosomes (A chromosome is a sample of the optimizing variable $(\alpha_1, ..., \alpha_K)$ where $\alpha_k \in \{0,1\}$). These chromosomes are randomly picked, satisfying the constraint:

$$
\sum_{k=0}^{K} \alpha_k \leq n_T.
$$

$$
P_m = \frac{1}{\beta_1 + \beta_2 \sigma_G / G}
$$

(20)
Where $\sigma_G$ is the standard derivation of the fitness of the current population (before selection). $\beta_1$ and $\beta_2$ are two constants.

These two processes introduce randomness into the intermediate generation so that the new population will be a combination of the best chromosome in the current population as well as some new random elements.

**Step 4:** Termination: Replace the original population with the new population and repeat Step 2 and Step 3 until the number of interactions reaches $N_g$. When forming new population, it is ensured that the fitness chromosome in the current population is saved and inserted into the next population. And all members of the next population is checked against the constraint $\sum_\alpha k \leq n_T$. If any chromosome violates this constraint, ‘O’ is inserted into a randomly selected bit position in the violating chromosome until the constraint is satisfied.

The computation complexity of the genetic algorithm is bounded by $N_p \times N_p$ function evolutions. As will be illustrated in the next section, this represents enormous computational saving compared with the optimal algorithm.

$\sigma_G$ is an indication of the population convergence because a population converging onto a local optimal solution will have small $\sigma_G$ while a population before convergence will have large $\sigma_G$. Hence, we would like to reduce the randomness introduced through mutation $p_m$ to speed up convergence when $\sigma_G$ is large but increase the mutation probability $P_m$ to avoid getting stuck at local optimal points when $\sigma_G$ is small. In this paper we use $\beta_1 = 1.2, \beta_2 = 10$.

**Numerical Results and Discussions**

In this section, we shall compare the performance of the downlink schedulers optimized for utility-based coverage, we shall investigate the contributions of multi-user selection diversity and spatial multiplexing to the network coverage. To highlight the contribution of multi-user selection diversity, we compare the coverage with respect to the random scheduler, where $n_T$ users are randomly selected irrespective of their channel matrices at every fading block. To highlight the contribution of spatial multiplexing. We compare the coverage with respect to various $n_T$. Furthermore, we shall compare the effectiveness and performance-complexity tradeoff of the genetic algorithm.

In the simulation, each data point consists of 5000 realizations of channel fading. Channel fading of the $K$ users are generated based on independent complex Gaussian distribution (with unit variance). We assume 0dB antenna gain in the transmit and receive antenna. The carrier frequency is assumed to be 2GHz. Data rate is expressed in terms of bits per second per Hz.

**Contribution of Spatial Multiplexing/Spatial Diversity to Network Coverage**

Figure 3 illustrates the network centric coverage versus the network capacity (network loading) of the coverage-optimized scheduler at $NT = 1.2, 4$. We observe that a significant gain in network coverage is achieved by increasing the number of transmit antenna $n_T$ at high SNR and this illustrates the contribution of spatial multiplexing to network coverage. For example, in Figure 3, there are 50% and 90% area coverage gain at target network capacity of 5b/s/Hz comparing relative to $NT=1$ for $n_T=2, 4$ respectively.

On the other hand, Figure 4 illustrates the user-centric utility $G_{use}(r_1, ..., r_K) = \min_{x \in A} r_k$ threshold versus cell radius at various $Q=1, 2, ..., 5$ and $n_T=5$ where $Q=|A|$. The user centric utility threshold $G_0$ we considered refers to the threshold at 1% outage probability where outage is defined as the event $G_{use}(r_1, ..., r_K) < G_0$. Hence a point $(x, y)$ in the graph means that over 99% of the time, scheduled user at $x$ m from the base station will be able to transmit at least $y$ b/s/Hz. We observe that to achieve high coverage gain (w.r.t. user centric utility), the space time scheduler should exploit the spatial diversity ($Q=1$) instead of spatial multiplexing ($Q=5$). For example, there is a 13 times are a coverage gain at target network capacity of 5b/s/Hz relative to $Q=1$ and $Q=5$ respectively.
Contribution of Multi-user Selection Diversity to Network Coverage

Figure 6(a-b) illustrates the capacity versus coverage between optimal scheduling and random scheduling at \( n_T = 1.4 \). For example, there is 3.2 times are coverage gain between the optimal scheduler and the random scheduler at \( n_T=4 \) and target network capacity of 2 b/s/Hz. This illustrates that multi-user selection diversity contributes significantly to coverage area gains.

Figure 5 illustrates the worst-case outage probability vs number of users \( K \) at various \( n_T \). Observe that as \( K \) increases, the efficiency of multi-user selection diversity increases because at any scheduling instance, it is more likely to select user with good channel conditions. Hence, the outage probability is reduced at the same target network loading (network capacity). Yet, supporting a large \( K \) would induce a large signaling overhead for channel estimation at the base station. In practice, \( K=20 \), could deliver a majority of the multi-user selection diversity gain for \( n_T = 4 \).

Performance Comparisons of the Genetic Algorithms

Figures 6 illustrates the performance of the genetic algorithm. The genetic algorithm has relatively small performance loss compared with the optimal scheduler.

Complexity Comparisons

Table I compares the number of function evaluations of the optimal algorithm and the genetic algorithm at various \( n_T \) and \( K \). Observe that there is a 8 times and 36 times saving in computation of genetic algorithm (compared with the optimal algorithm) when \( (K,n_T) = (10,4) \) and \( (20,4) \) respectively. Furthermore, the MFPS (Million Function evaluations Per Second) requirement of the genetic algorithm is still within implementation limit. Hence the genetic algorithm may be used as real time scheduling algorithm.

Conclusion

In this paper, we focus on the design of coverage-optimized downlink scheduler for systems with multiple antennas. A systematic framework is proposed for the scheduler design. Problem based on information theoretical approach. There are \( K \) mobiles with single receive antenna and one base station with \( n_T \) transmit antennas. We proposed a generalized definition of network coverage, namely the utility-based coverage. The contributions of multi-user selection diversity and spatial multiplexing to the network coverage are analyzed. While the optimal scheduler delivers the best coverage, the computation complexity is huge. A real-time genetic algorithm is proposed, which offers enormous computational savings compared with the optimal algorithm and is an attractive candidate given it’s performance complexity tradeoff.

Table 1 : Comparison Of Computational Complexity (In Terms Of Number Of Function Evaluations) Of Genetic And Optimal Algorithms. Is The Genetic Column. The Format Of The Results Is \( N_p \times N_g \). The Brackets Indicates The Meps Assuming 2ms Packet Duration. [Meps= Million Function Evaluation Per Second].

<table>
<thead>
<tr>
<th>((k,n_T))</th>
<th>Genetic Algorithm</th>
<th>Optimal Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10.2)</td>
<td>10x2=20 [0.01]</td>
<td>55[0.027]</td>
</tr>
<tr>
<td>(10.4)</td>
<td>10x5=50 [0.025]</td>
<td>385[0.194]</td>
</tr>
<tr>
<td>(20.2)</td>
<td>10x5=50[0.025]</td>
<td>210[0.150]</td>
</tr>
<tr>
<td>(20.4)</td>
<td>20x5=100 [0.05]</td>
<td>3645[1.32]</td>
</tr>
</tbody>
</table>

References