

Solução Ótima para Alocação de Recursos de Rádio com Restrições de QoS em Sistemas Multicelulares com Múltiplos Serviços

Lászlón R. Costa^{1,2}, F. Rafael M. Lima^{1,2}, Yuri C. B. Silva¹,
F. Rodrigo P. Cavalcanti¹

¹Grupo de Pesquisa em Telecomunicações Sem Fio (GTTEL),
Universidade Federal do Ceará, Fortaleza, Brasil

²Programa de Pós-Graduação em Engenharia Elétrica e de Computação (PPGEEC),
Universidade Federal do Ceará, Sobral, Brasil

{laszlon, rafaelm, yuri, rodrigo}@gttel.ufc.br

Resumo. *Este trabalho propõe uma solução ótima para distribuição de RB (do inglês, Resource Block) considerando os efeitos da interferência, com o objetivo de maximizar a soma da taxa de dados total de um sistema multi celular com múltiplos serviços e restrições que assegurem um percentual mínimo de usuários satisfeitos para cada serviço. De forma a atingir uma experiência uniforme na área de cobertura, consideraremos critérios de satisfação de usuários localizados no centro e borda da célula. Resultados obtidos através de simulações computacionais mostram a importância de aplicar estratégias de gerenciamento de interferência em sistemas modernos e demonstram as vantagens da solução do problema formulado quando comparado às soluções encontradas na literatura.*

Abstract. *This work proposes an optimal solution to the Resource Block (RB) distribution in order to maximize the sum rate in a multicell and multiservice system with constraints on the percentage of users that should be satisfied per service. In order to achieve a more uniform experience in the coverage we propose takes into account the satisfaction of users located at the center and edge of cells. Results of computer simulations show the importance of interference management in modern systems and the advantages of proposed solution when compared with others solutions found in the literature.*

1. Introdução

Com a rápida popularização e adesão ao uso de dispositivos móveis inteligentes como *smartphones*, *tablets* e *laptops* portáteis, têm crescido também a demanda por altas taxas de dados e a necessidade do uso de diversos serviços em redes móveis. Para atender esta crescente demanda, as operadoras devem garantir que suas redes suportem o aumento do tráfego de dados, o que tem se mostrado cada vez mais desafiador. O desenvolvimento de novas estratégias para suprir o aumento da demanda de tráfego de dados estão em consonância com o desenvolvimento da Quinta Geração (5G) de redes móveis.

O aumento da capacidade de uma rede celular pode ser alcançado basicamente por três métodos: alocação de novas bandas de frequência, otimizar o uso da camada física do sistema ou novos paradigmas para a arquitetura da rede celular. O espectro de frequência que atualmente é utilizado em sistemas móveis é escasso, além do fato de que o aumento dessa banda pode ser custoso para as operadoras que devem enfrentar entraves burocráticos para concessão de mais bandas de frequência. Melhorias na camada

física, como novos esquemas de codificação e modulação, transmissão e recepção ou mesmo técnicas de múltiplo acesso estão chegando cada vez mais perto do limite teórico da eficiência espectral. Sendo assim, melhorias na arquitetura da rede se mostram como um dos principais meios para melhorar o desempenho das tecnologias de comunicações para a próxima geração [Bhushan et al. 2014].

Um método para aumentar a capacidade e atender grandes volumes de tráfego tal qual ocorre em regiões muito populosas, consiste em diminuir a distância entre as Estações Rádio Base (ERBs). Considerando uma densidade de usuários fixa, essa técnica garantirá que as ERBs servirão menos terminais e necessitarão de menos recursos de rádio. Os recursos mais relevantes em um sistema de comunicação são a potência de transmissão, banda de frequência e *time slot*, que são gerenciados de forma a atingir um melhor objetivo como maximização de taxa de transmissão, redução do atraso ou mesmo diminuição dos erros de bits. Na densificação teremos mais recursos de rádio disponíveis por área, ou seja, é possível maior reuso de recursos aumentando a capacidade de usuários. A densificação é uma das principais linhas de pesquisa para o 5G [Bhushan et al. 2014].

Entretanto, a maior densidade de canais por área obtida através da densificação da rede apresenta desafios relacionados ao aumento de interferência e carga de sinalização. Uma rede com distâncias entre ERBs reduzidas também implica em efeitos de perda de percurso menores, tornando comum casos que o sinal de interferência tenha a mesma ordem de grandeza dos sinais de interesse. Usuários localizados nas bordas das células tendem a receber maior interferência e potência de interesse mais fraca, requisitando assim, uma atenção especial. Sendo assim, a distribuição de recursos deve ser cuidadosamente planejada e executada de forma a combater os efeitos colaterais de densificação de ERBs. Uma forma de garantir uma boa gestão global de recursos na rede é através de alguma entidade central que, com total conhecimento dos requisitos e recursos do sistema, gerencie a rede baseado nas necessidades dos usuários e recursos disponíveis. As considerações deste trabalho podem ser implementadas na arquitetura de rede C-RAN (do inglês, *Cloud Radio-Access Network*) que utiliza processamento centralizado para múltiplas ERBs [Mohsen and Hassan 2015].

2. Estado da Arte e Contribuições

Problemas de RRA (do inglês, *Radio Resource Allocation*) são formulados como problemas de otimização que são compostos de uma função objetivo e restrições [Sadr et al. 2009, Bohge 2010].

Há na literatura um grande volume de trabalhos para alocação de recursos com diversos objetivos, porém sua grande maioria aplicada em cenário com uma única célula desprezando, portanto, os efeitos da interferência intercelular. Alocação de RB (do inglês, *Resource Block*) em sistemas OFDMA (do inglês, *Orthogonal Frequency Division Multiple Access*) em um cenário com multicelulas foi estudado no trabalho [Rahman and Yanikomeroglu 2007] em que é proposto um problema de alocação de RBs com o objetivo de maximizar a soma da taxa de dados para um sistema que oferece um único serviço com apenas a taxa de dados como requisito de QoS (do inglês, *Quality of Service*) em termos de taxa de dados. Foi utilizado um conceito de grupos interferentes, que é composto por células que utilizam os mesmos recursos e possuem antenas com padrão de irradiação dominante na direção da área de cobertura. No cenário abordado não há um limite de tolerância mínimo para o número de usuários que devem ser satisfeitos.

Uma análise entre taxa de dados e QoS em termos de atrasos de pacotes para um sistema com multicelulas é proposto em [Akbari and Vakili 2012]. O problema formulado é não convexo e combinatorial e nenhuma análise em relação a solução ótima para o problema é apresentada. O trabalho [Rahman and Yanikomeroğlu 2008] propõe uma solução subótima semi distribuída em uma arquitetura hexagonal setorizada para maximização de taxa de dados em que é levada em conta a interferência de apenas dois anéis interferentes. A proposta utiliza o algoritmo Húngaro [Kuhn 2010] em cada ERB e após este passo uma resolução de conflitos é feita por uma entidade central.

Abordagens que também consideram a associação entre usuários e ERBs em conjunto com a alocação de recursos para maximização da mínima taxa dos usuários no sistema são analisadas em [Sadr and Adve 2014] e [Gotsis and Alexiou 2013]. Especificamente em [Sadr and Adve 2014], os autores propõem a utilização de quatro passos para este objetivo, sendo eles: (1) a associação de usuários com ERBs, (2) estimação da carga de acordo com a taxa exigida pelos usuários, (3) distribuição de recursos entre as ERBs através de coloração de grafos e (4) alocação dos recursos para os usuários com informações locais. Múltiplos serviços e um número mínimo de usuários a serem satisfeitos não são considerados, além de que o escalonamento de recursos realizado pelas ERBs, no quarto passo, não consideram a interferência do sistema. A eficiência da solução baseia-se na distribuição dos recursos entre as ERBs no terceiro passo, evitando que células vizinhas usem o mesmo recurso. Uma abordagem diferente é estudada em [Gotsis and Alexiou 2013], que propôs uma formulação ótima para um problema de maximização da SINR (do inglês, *Signal to Interference plus Noise Ratio*) mínima do sistema através da associação de usuários, alocação de RBs e alocação de potência. O algoritmo tenta solucionar de forma ótima um ILP (do inglês, *Integer Linear Program*) dado uma SINR mínima alvo. O maior valor possível da SINR alvo é encontrada através do método da bisseção, sendo necessário resolver o problema ILP para cada iteração. Toda a formulação é baseada em SINR alvo mínima e não considera restrições de taxa. O método torna possível que o sistema encontre a melhor SINR mínima, porém não há garantias que os recursos alocados garantam taxas de dados necessária para uso de serviços. Ambos os trabalhos [Sadr and Adve 2014] e [Gotsis and Alexiou 2013] concluem que a associação ótima entre usuários e ERBs é próxima de uma associação baseada apenas em qualidade de canal.

A redução da interferência em cenário com multicelulas pode ser alcançada através da minimização da potência de transmissão das células do sistema. O trabalho [Lopez-Perez et al. 2014] propõe um problema de otimização ILP com objetivo de minimizar a potência total transmitida no sistema com restrições de QoS dos usuários em termos de taxa de dados. O problema formulado é de alta complexidade e os autores propõem utilizar uma abordagem distribuída, em que o problema é resolvido localmente e de forma independente por cada ERB. Este método espera que com a redução contínua de potência irradiada no sistema, a interferência seja mitigada levando a uma convergência em que a rede se auto organize. O método não traz uma solução ótima e o problema formulado só apresentará a solução ótima se houver apenas uma célula no sistema.

De acordo com nossa pesquisa bibliográfica, não encontramos na literatura uma proposta de problema de otimização que considere a heterogeneidade de requisitos de QoS de diferentes serviços de tráfego e que permita a satisfação de um número mínimo de usuários por serviço. Uma solução com essas características para um cenário com uma

única célula foi proposto em [Lima et al. 2012] em que é formulado um novo problema de RRA para maximização da taxa total do sistema restrito a requisitos de satisfação mínimas por cada serviço oferecido no sistema.

O objetivo deste trabalho é propor uma generalização do problema formulado em [Lima et al. 2012] para um ambiente com multicelulas limitado por interferência. O novo problema de otimização também explorará os requisitos de QoS em termos de taxa de dados criando justiça para usuários de borda em cenários com células hexagonais. Além da nova formulação este trabalho propõe uma solução ótima. Simulações computacional foram realizadas, evidenciando a vantagem em considerar interferência ao realizar alocação de recursos.

Este artigo segue a seguinte organização. Na seção 3 apresentamos o cenário e as considerações adotadas neste trabalho. A formulação do problema e a caracterização da solução ótima é apresentada na seção 4. As simulações computacionais e análise dos resultados estão presentes na seção 5. Por fim, as conclusões deste trabalho estão apresentadas na seção 6.

3. Modelagem do Sistema

Neste trabalho consideramos um sistema TDD (do inglês, *Time Division Duplexing*) composto por M células, utilizando antenas omnidirecionais, que servem um total de J usuários. Cada usuário pode utilizar um dos S serviços de tráfego oferecidos pela operadora. Como exemplo de serviços temos VoIP (do inglês, *Voice over IP*) e *web browsing*. O trabalho é baseado no enlace direto e o acesso dos usuários é realizado por meio de uma combinação de OFDMA e TDMA (do inglês, *Time Division Multiple Access*). O recurso mínimo a ser alocado durante um TTI (do inglês, *Transmission Time Interval*) é o RB que é composto por uma combinação de subportadoras e símbolos OFDMA adjacentes. A qualidade de um RB é medida através de sua SINR, que mede a relação entre a potência do sinal recebido pela soma da potência do sinal interferente adicionado ao ruído térmico do sistema. Para uma modelagem coerente será considerado que todas as subportadoras de um mesmo RB tenham ganhos de canais correlacionados.

Seja $\mathcal{J}_m^{\text{ERB}}$ o conjunto de usuários que é servido por uma ERB m e $\mathcal{J}_n^{\text{RB}}$ o conjunto de usuários que utilizam o RB n , podemos modelar a SINR observada pelo usuário j servido pela ERB m no RB n como:

$$\gamma_{j,n} = \frac{p^{\text{RB}} h_{j,m} |\alpha_{j,m,n}|^2}{\sigma^2 + \sum_{m' \neq m}^M p^{\text{RB}} h_{j,m'} |\alpha_{j,m',n}|^2}, \quad (1)$$

em que $h_{j,m}$ é a perda de percurso com o desvanecimento de longo prazo entre um usuário j e uma ERB m enquanto que o desvanecimento rápido desse *link* em um dado RB n é representado por $\alpha_{j,m,n}$. A potência do ruído e de transmissão em um RB são respectivamente σ^2 e p^{RB} .

A taxa de dados de transmissão em uma comunicação sem fio depende do estado do canal e potências de transmissão, ruído e sinais interferentes. Considere que o mapeamento entre a qualidade de canal e taxa de dados obtida na camada física é dada por uma função de *link* adaptativo $f(\cdot)$. Em sistemas reais a adaptação de enlace leva em conta o MCS (do inglês, *Modulation and Coding Scheme*) que mapeia a taxa de dados de acordo com a SINR e a taxa de erro de bloco, ou BLER (do inglês, *Block Error Rate*). Quanto maior o nível de MCS maior a capacidade de transmitir bits, porém também serão maiores

Tab. 1. Mapeamento geral entre SINR e taxa de dados transmitida por RB.

| Região de SINR | Taxa de dados transmitida por RB |
|---|----------------------------------|
| $\gamma_{j,n} < \gamma^1$ | 0 |
| $\gamma^1 \leq \gamma_{j,n} < \gamma^2$ | r^1 |
| $\gamma^2 \leq \gamma_{j,n} < \gamma^3$ | r^2 |
| \vdots | \vdots |
| $\gamma^{V-1} \leq \gamma_{j,n} < \gamma^V$ | r^{V-1} |
| $\gamma_{j,n} \geq \gamma^V$ | r^V |

as probabilidades de erro de bloco em função da SINR. Sendo assim quanto maior o nível de MCS mais alta será a SINR necessária para garantir uma taxa de erro mínima aceitável que possibilite uma transmissão satisfatória. Desta forma, a taxa de dados total de um usuário j , usando um recurso n em um nível de MCS v é dada por:

$$r_{j,n,v} = BR_v \cdot (1 - \text{BLER}(v, \gamma_{j,n})), \quad (2)$$

em que BR_v é a taxa de bits bruta em um nível de MCS v e BLER é a função que mapeia a probabilidade de erro de blocos de bit levando em conta o nível de MCS e a SINR. Sem perder a generalidade consideraremos que para um usuário usar um nível de MCS v será necessário atingir um nível de SINR em que a BLER seja tolerável. A taxa de dados de um usuário j em um RB n será mapeada de forma discreta por uma função adaptativa escada $f(\gamma_{j,n})$, como ilustrado na Tabela 1 em que γ^v e r^v são respectivamente a SINR mínima tolerável para atingir uma MCS v e a taxa de dados referente a MCS v . Para este trabalho será considerado o esquema de MCS do LTE (do inglês, *Long Term Evolution*) [3GPP 2009], composta por 15 níveis.

O sistema considerado neste trabalho oferece múltiplos serviços, tornando possível o usuário utilizar a rede para diferentes fins, como por exemplo: serviços de dados, voz, mensagens de texto e serviços de emergências. Cada serviço possui diferentes necessidades, e como consequência há diferentes critérios de qualidade. Os critérios de QoS mais comuns em redes móveis são taxa de dados, tempo de atraso de pacotes, confiabilidade de dados, etc. Neste contexto, um sistema que oferece diversos serviços deve alocar seus recursos de acordo com as necessidades dos serviços utilizados pelos usuários, evitando desperdícios de recursos e procurando atender critérios da satisfação. Neste trabalho consideraremos apenas requisitos de taxa para múltiplos serviços, em que cada serviço oferecido será associado a uma taxa de dados.

Buscando manter justiça no acesso a rede entre usuários de centro e de borda das células hexagonais, ou seja, garantir para ambos os tipos de usuários os mesmos requisitos de taxa de dados, iremos considerar uma duplicação de cada serviço oferecido. Embora um usuário de borda e de centro possam utilizar o mesmo tipo de serviço faremos uma distinção entre estes. Por exemplo: considere uma operadora que oferece dois serviços s_1 e s_2 com exigências de taxa t^{s_1} e t^{s_2} respectivamente. Na prática modelaremos para o problema quatro serviços virtuais s_1, s_2, s_3 e s_4 com requisitos de taxa $t^{s_1}, t^{s_2}, t^{s_1}$ e t^{s_2} , respectivamente. A diferença nessa formulação é que os usuários dos serviços s_3 e s_4 serão apenas usuários de borda enquanto s_1 e s_2 são exclusivos para usuários do centro. Diversos critérios podem ser utilizados para classificar usuários de centro e borda da célula. Como exemplo podemos citar geolocalização ou potência média recebida. Por simplicidade, neste trabalho iremos considerar que a localização geográfica do terminal móvel é

conhecida. Será mostrado nas próximas seções que esta criação de serviços virtuais resultará em ganhos de justiça em termos de número de usuários satisfeitos na borda da célula. A Figura 1 faz uma representação gráfica da rede mostrando o limitante entre centro e borda da célula. Para a aplicação proposta de solução ótima é necessário que a entidade

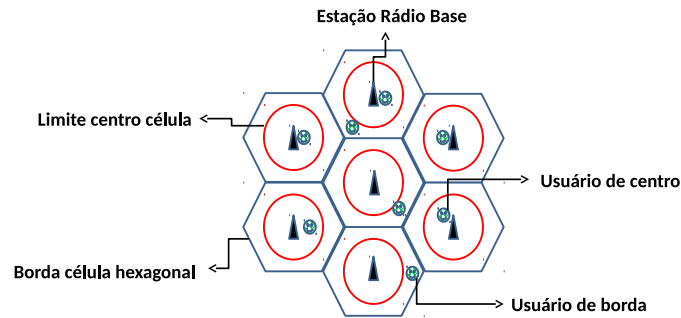


Figure 1. Representação gráfica da rede com divisão entre centro e borda das células.

central que determinará a alocação de recursos tenha conhecimento, além dos ganhos de canais entre usuários e ERB servidora, o estado de canal entre terminais e ERBs interferentes. Para isto, consideraremos que o sistema é capaz de estimar os canais interferentes. Essa estimativa é possível utilizando SRS (do inglês, *Sounding Reference Signal*), que foi definido no padrão LTE-A (do inglês, *Long Term Evolution - Advanced*) [3GPP 2012]. Este método foi projetado para se ter uma estimativa de ganhos de canal em qualquer recurso de frequência, mesmo nos que não estejam sendo utilizados para transmissão. Periodicamente os terminais enviam sinais de referência baseadas na sequência de Zadoff-Chu que quando deslocadas no tempo têm a propriedade da ortogonalidade entre cada cópia deslocada. Para mais detalhes convidamos a leitura de [Heimiller 1961] e [Chu 1972].

Considerando área densa de baixa mobilidade e sujeito a uma entidade central administradora da rede, é possível alocar diferentes *shifts* de uma mesma sequência de Zadoff-Chu por vez, dentro de um *cluster* cooperativo de Estação Rádio Bases (ERBs), respeitando o tempo de coerência dos canais no ambiente. Assim, quando um usuário na transmissão do canal reverso enviar um sinal de referência ortogonal para a ERB servidora, todas as ERBs interferentes poderão isolar este sinal e assim estimar o ganho da interferência que causarão nestes usuários. Para os resultados apresentados nas próximas seções, estaremos interessados na solução ótima e consideramos o perfeito conhecimento dos estados de canais. Análise da robustez dos algoritmos envolvidos na presença de incerteza em relação ao conhecimento da qualidade de canal são tópicos a serem explorados em trabalhos futuros.

4. Formulação do Problema e Solução Ótima

Consideraremos um sistema com potência de transmissão fixa para cada RB. Esta é uma consideração coerente, pois em geral na literatura é preferido um assinalamento

de recursos com potência fixa que soluções com alocações conjuntas de potência devido a alta complexidade dos problemas, tratamento matemático e dificuldades de implementação [Lopez-Perez et al. 2014]. Análises prévias feitas em [Song and Li 2005] e [Jang and Lee 2003] mostram que a melhora causada pela alocação de canais com potência diferentes é negligenciável quando aplicado em cenários com grande número de usuários e diversos requisitos de QoS. Baseado nestes resultados procuraremos obter ganhos na taxa de transmissão através apenas da gestão na distribuição de RBs, levando em conta a interferência e o ganho de canal.

A alocação de RBs para os usuários servidos por uma dada ERB é realizado de forma exclusiva, ou seja, em uma célula cada RB poderá ser alocado a no máximo um usuário, eliminando assim interferência intracelular. Sendo assim apenas a interferência intercelular poderá estar presente. Para contabilizar toda a interferência no sistema usaremos o conceito de grupos interferentes. Seja \mathcal{G} o conjunto de todos os grupos interferentes que é formado por todas as possíveis combinações entre usuários, incluindo grupos com um único usuário. Por exemplo, considere que um sistema tenha três ERBs servindo quatro usuários dispostos na forma de: $\mathcal{J}_1^{\text{ERB}} = \{1\}$, $\mathcal{J}_2^{\text{ERB}} = \{2, 3\}$ e $\mathcal{J}_3^{\text{ERB}} = \{4\}$. Note que não pode haver alocações de um mesmo RB para os usuários 2 e 3, pois ambos são servidos pela mesma ERB. Sendo assim devemos excluir grupos que contenham estes dois usuários. As combinações válidas de interferentes serão $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$, $\{2, 4\}$, $\{3, 4\}$, $\{1, 2, 4\}$ e $\{1, 3, 4\}$. O número mínimo de grupos interferentes válidos para um sistema é equivalente a quantidade de usuários, se considerarmos que há apenas uma ERB para todos os usuários. Portanto, quanto mais densa a rede menos grupos interferentes. O caso de uma única ERB corresponde ao cenário adotado em [Lima et al. 2012].

Para formulação do problema de otimização devemos definir variáveis que representem associações entre as entidades no sistema. A associação entre um usuário j e uma ERB m será representada pela variável binária $z_{j,m}$ que assumirá valor 1 caso j seja servido pela ERB m e 0 caso contrário. Neste trabalho a associação entre usuários e ERB será realizada através da qualidade média do canal, ou seja, o usuário será servido pela ERB que apresente o melhor estado de canal médio, tornando $z_{j,m}$ uma constante que pode ser pré-calculada antes resolver o problema de otimização. De forma semelhante a participação de um usuário j em um grupo interferente g será representado pela variável binária $o_{g,j}$ e será calculada antes da aplicação da solução do problema. Assim $o_{g,j}$ assumirá valor 1 quando o usuário j estiver presente no grupo g e 0 caso contrário. A taxa de dados em um único bloco de recurso n de um usuário j que participa de um grupo g será representada por $r_{g,j,n}$. Note que como sabemos quais os usuários que fazem parte de um grupo interferente e conhecemos também os ganhos de canais de interesse e interferentes podemos utilizar a Equação (1) para determinar a SINR dos participantes do grupo e mapear suas taxas de dados. Essa arquitetura só é possível se considerarmos que a solução é realizada por uma unidade central que recebe os dados de todas as ERBs e define a alocação em todo o sistema. Definimos também a variável $c_{j,s}$ que assumirá valor 1 caso o usuário j esteja utilizando o serviço s e 0 caso contrário. Neste trabalho consideraremos que cada usuário usará apenas um serviço por vez. A escolha de qual recurso será utilizado por um grupo é a variável binária de otimização $x_{g,n}$ que assume valor 1 para associação entre o grupo g e RB n , e 0 caso contrário. O problema é formulado como:

$$\max_{x_{g,n}} \sum_j \sum_n \sum_g x_{g,n} r_{g,j,n} O_{g,j} \quad (3a)$$

$$\text{sujeito a } \sum_g x_{g,n} = 1 \quad \forall n, \quad (3b)$$

$$\sum_j u \left(\sum_g \sum_n x_{g,n} O_{g,j} r_{g,j,n} C_{j,s} z_{j,m}, r_{\text{obj}}^j \right) \geq k_s \quad \forall s \text{ e } \forall m, \quad (3c)$$

$$\sum_j \sum_g x_{g,n} O_{g,j} z_{j,m} \leq 1 \quad \forall m \text{ e } \forall n, \quad (3d)$$

$$x_{g,n} \in \{0, 1\} \quad \forall n \text{ e } \forall g. \quad (3e)$$

A função objetivo (3a) é a soma da taxa de dados de todos os usuários e todos os RBs no sistema. A restrição (3b) garante que um recurso apenas será alocado para um único grupo de interferência, desse modo essa restrição garante que a escolha da alocação sempre contabilizará todos os efeitos de interferência possíveis. A segunda restrição (3c) é o requisito de QoS em que $u(b, d)$ é a função degrau que assumirá valor 1 caso $b \geq d$ e 0 caso contrário. Nesta restrição r_{obj}^j representa a taxa alvo para o usuário j atingir a satisfação. Esta taxa depende do serviço utilizado em que cada serviço tem sua própria taxa mínima necessária. O número mínimo de usuários a serem satisfeitos que usam o serviço s é dado por k_s . A última restrição (3d) garante a exclusividade no uso dos RBs dentro da área de cobertura de cada ERB. O problema formulado anteriormente tem a função objetivo e todas as restrições lineares, com exceção da restrição (3c) pois a função degrau não é linear. Por conta da não linearidade do problema, a solução ótima só poderá ser encontrada por busca exaustiva, que possui alta complexidade computacional. Porém, podemos linearizar o problema adicionando uma nova variável ρ_j que assume o valor 1 caso o usuário j esteja satisfeito em relação a seu QoS, ou seja, $\sum_g \sum_n x_{g,n} O_{g,j} r_{g,j,n} \geq r_{\text{obj}}^j$ e 0 caso contrário. A reformulação do problema será:

$$\max_{x_{g,n}, \rho_j} \sum_j \sum_n \sum_g x_{g,n} r_{g,j,n} O_{g,j} \quad (4a)$$

$$\text{sujeito a } \sum_g x_{g,n} = 1 \quad \forall n, \quad (4b)$$

$$\sum_g \sum_n x_{g,n} O_{g,j} r_{g,j,n} \geq \rho_j r_{\text{obj}}^j \quad \forall j, \quad (4c)$$

$$\sum_j \rho_j C_{j,s} z_{j,m} \geq k_s, \quad \forall s \text{ e } \forall m, \quad (4d)$$

$$\sum_j \sum_g x_{g,n} O_{g,j} z_{j,m} \leq 1 \quad \forall m \text{ e } \forall n, \quad (4e)$$

$$x_{g,n} \in \{0, 1\} \quad \forall n \text{ e } \forall g, \quad (4f)$$

$$\rho_j \in \{0, 1\} \quad \forall j. \quad (4g)$$

Note que o novo problema segue a mesma estrutura formulada anteriormente, porém, a restrição (3c) foi substituída pelas restrições (4c) e (4d), em que respectivamente, é garantido a taxa mínima para satisfação dos usuários selecionados e o número

mínimo de terminais satisfeitos por serviço. O problema é combinatorial da classe ILP e a solução ótima agora pode ser encontrada utilizando métodos como o algoritmo BB (do inglês, *Branch-and-Bound*) [Nemhauser and Wolsey 1988]. A complexidade deste algoritmo cresce exponencialmente com número de restrições e variáveis. Caso um problema tenha l variáveis e p restrições o número de subproblemas lineares a serem resolvidos será de $(\sqrt{2})^l$. Cada subproblema necessitará de $2(p + l)$ iterações em que serão executados $(pl - p)$ multiplicações, $(pl - p)$ somas e $(l - p)$ comparações [Zhang and Letaief 2004, Sierksma 2001]. Na formulação proposta temos $NG + J$ variáveis e $N + J + M(S + N)$ restrições, em que G e N são o número de grupos e RBs no sistema. Sendo assim o número de operações para solução no pior caso é dado por

$$\sqrt{2}^{NG+J} 2(NG+2J+N+M(S+N))(2(NG+J)(N+J+M(S+N))-3(N+J+M(S+N))+NG+J), \quad (5)$$

considerando o termo de maior ordem a complexidade computacional, no pior caso, será $O(2^{NG})$.

Para utilização de *softwares* para solução de problemas de otimização, como é o caso do CPLEX [IBM 2009], é necessário transcrever o problema para uma forma mais compacta e baseado em operações matriciais. Sendo assim, demonstraremos abaixo uma construção de matrizes para o problema (4) em forma compacta com apenas uma restrição matricial de igualdade e uma de desigualdade.

Sendo $(\cdot)^T$ a transposição de uma matriz e agrupando alguns valores do problema em vetores poderemos agrupar as variáveis de otimização na forma de $\mathbf{x}_g = [x_{g,1}, \dots, x_{g,N}]$, $\mathbf{x} = [\mathbf{x}_1^T, \dots, \mathbf{x}_g^T]^T$. Usaremos uma nova variável de otimização definida como $\mathbf{y} = [\mathbf{x}^T | \boldsymbol{\rho}^T]$. Para isolar as variáveis desejadas precisamos aplicar filtros que torna possível as operações: $\mathbf{x} = \mathbf{A}_1 \mathbf{y}$ e $\boldsymbol{\rho} = \mathbf{A}_2 \mathbf{y}$. Os filtros que mantêm essa propriedade são $\mathbf{A}_1 = [\mathbf{I}_{GN} | \mathbf{0}_{GN \times J}]$ e $\mathbf{A}_2 = [\mathbf{0}_{J \times GN} | \mathbf{I}_J]$ em que \mathbf{I}_J é uma matriz identidade de dimensão $J \times J$ e $\mathbf{0}_{n_1 \times n_2}$ uma matriz de zeros de dimensão $n_1 \times n_2$. Considere também os vetores $\mathbf{o}_g = [o_{g,1}, o_{g,2}, \dots, o_{g,J}]^T$, $\mathbf{r}_{g,n} = [r_{g,1,n}, \dots, r_{g,J,n}]^T$ e $\mathbf{a} = [\mathbf{o}_1^T \mathbf{r}_{1,1}, \dots, \mathbf{o}_1^T \mathbf{r}_{1,N} | \dots | \mathbf{o}_G^T \mathbf{r}_{G,1}, \dots, \mathbf{o}_G^T \mathbf{r}_{G,N}]$. Sendo assim a nova função objetivo será $\mathbf{A}_1 \mathbf{y}$.

Para as restrições também precisaremos definir novos vetores. A associação usuário-ERB é representada pelo vetor binário $\mathbf{z}'_m = [z_{m,1}, \dots, z_{m,J}]$. Também será necessário definir um vetor binário com a associação entre usuário-ERB-serviço que será composta pelo vetor $\mathbf{z}_{m,s} = [z_{m,1,s}, \dots, z_{m,J,s}]^T$ em que $z_{m,j,s}$ assume valor um caso o usuário j esteja sendo servido pela ERB m e usando o serviço s . O número de usuários mínimos a serem satisfeitos em cada serviço e em cada ERB será dado pelo

$$\text{vetor } \mathbf{k} = \left[\underbrace{k_1, \dots, k_S}_{M \text{ vezes}} \mid \dots \mid k_1, \dots, k_S \right]^T.$$

Como não poderemos alocar recursos para grupos com mais de um usuário da mesma ERB precisaremos definir uma função auxiliar que determina se um grupo possui mais de um usuário servido pela mesma ERB. Matematicamente essa função será:

$$w_g := \begin{cases} 0, & \text{se } \sum_m u(\mathbf{o}_g^T \mathbf{z}'_m, 2) > 0, \\ 1, & \text{caso contrário.} \end{cases} \quad (6)$$

Com isso o vetor que usaremos para determinar a validade de um grupo interferente na nossa formulação será $\mathbf{w} = \left[\underbrace{w_1, \dots, w_1}_{N\text{vezes}} \mid \dots \mid \underbrace{w_G, \dots, w_G}_{N\text{vezes}} \right]^T$. Com todos os vetores formulados acima devemos definir a partir delas matrizes que serão: $\mathbf{B} = \left[\underbrace{\mathbf{I}_N, \dots, \mathbf{I}_N}_{G \text{ vezes}} \right]$, $\mathbf{C} = \begin{bmatrix} o_{1,1} r_{1,1,1} \cdots o_{1,1} r_{1,1,N} \mid \cdots \mid o_{G,1} r_{G,1,1} \cdots o_{G,1} r_{G,1,N} \\ \vdots & & \ddots & & \vdots \\ o_{1,J} r_{1,J,1} \cdots o_{1,J} r_{1,J,N} \mid \cdots \mid o_{G,J} r_{G,J,1} \cdots o_{G,J} r_{G,J,N} \end{bmatrix}$, $\mathbf{Z} = \left[\mathbf{z}_{1,1} \cdots \mathbf{z}_{1,S} \mid \cdots \mid \mathbf{z}_{M,1} \cdots \mathbf{z}_{M,S} \right]^T$, e $\mathbf{E} = \text{diag}(r_{\text{obj}}^1, \dots, r_{\text{obj}}^J)$, em que $\text{diag}(\cdot)$ é a função que representa uma matriz diagonal em que a diagonal principal são os valores de entrada e $\mathbf{1}_{J_s}$ é um vetor coluna composto de 1's com o tamanho J_s que é o número de usuários utilizando o serviço s .

Baseado nas definições matriciais acima podemos redefinir a restrição (4b) como $\mathbf{B} \mathbf{A}_1 \mathbf{y} = \mathbf{1}_N$, em quem $\mathbf{1}_N$ é um vetor coluna composto por 1's de dimensão N . A restrição (4c) será reformulada como $\mathbf{C} \mathbf{A}_1 \mathbf{y} \geq \mathbf{E} \mathbf{A}_2 \mathbf{y} \Rightarrow (\mathbf{C} \mathbf{A}_1 \mathbf{y} - \mathbf{E} \mathbf{A}_2 \mathbf{y}) \geq \mathbf{0}_J$. As restrições (4d) e (4e) serão respectivamente $\mathbf{Z} \mathbf{A}_2 \mathbf{y} = \mathbf{k}$ e $\mathbf{A}_1 \mathbf{y} \leq \mathbf{w}$.

Note que temos duas restrições de igualdade e duas de desigualdade. Para compactarmos o problemas iremos concatená-las em apenas duas restrições, uma de cada classe. Considere as concatenações $\mathbf{U} = [(\mathbf{B} \mathbf{A}_1)^T \mid (\mathbf{Z} \mathbf{A}_2)^T]^T$, $\mathbf{P} = [-(\mathbf{C} \mathbf{A}_1 - \mathbf{E} \mathbf{A}_2)^T \mid \mathbf{A}_1^T]^T$, $\mathbf{v} = [\mathbf{1}_N^T \mid \mathbf{k}^T]^T$ e $\mathbf{q} = [\mathbf{0}_J^T \mid \mathbf{w}^T]^T$, então podemos reformular todo o problema como:

$$\max_{\mathbf{y}} \quad \mathbf{a} \mathbf{A}_1 \mathbf{y} \quad (7a)$$

sujeito a

$$\mathbf{U} \mathbf{y} = \mathbf{v}, \quad (7b)$$

$$\mathbf{P} \mathbf{y} \leq \mathbf{q}. \quad (7c)$$

5. Resultados

Para avaliar o desempenho das soluções ótima proposta realizamos simulações em diversos cenários utilizando o *software* matemático MATLAB. Para encontrar as soluções ótimas dos problemas ILP utilizamos o CPLEX como método de resolução. As soluções utilizadas para comparação foram:

- **Apenas SNR:** Usaremos este termo para referir-se a solução ótima que foi proposta em [Lima et al. 2012]. Neste problema a alocação de recursos é realizada por uma única célula considerando apenas a SNR (do inglês, *Signal-to-Noise Ratio*). Replicamos a formulação do problema ILP proposto e resolvemos o problema em cada célula sem considerar o conhecimento sobre a interferência.
- **Único Serviço:** Método que não distingue serviços para usuários de borda e centro da célula. Para viabilizar a comparação com nossa proposta, consideraremos que o número mínimo de usuários satisfeitos para um dado serviço de tráfego neste caso é igual a soma do número de usuários que devem ser satisfeitos para o mesmo serviço no centro e borda da célula.

Os principais parâmetros de simulação são apresentados na Tabela 2. Para validação estatística dos métodos utilizamos TTIs totalmente descorrelacionados, de forma que a cada TTI novas posições de usuários e ganhos de canais são gerados. Neste

caso o problema deve resolver a alocação em um *snapshot*, que é o instante em que todos os parâmetros são considerados fixos, sem variação temporal dos elementos do sistema.

Tab. 2. Parâmetros da Simulação Células Hexagonais

| Parâmetro | Valor | Unidade |
|---|--|---------|
| Raio da célula | 200 | m |
| Raio do centro da célula | 150 | m |
| Potência de transmissão por RB | 8.2607 | dBm |
| N. de subportadoras por RB | 12 | - |
| Desvio padrão do sombreamento | 10 | dB |
| Perda de Percurso | $30.6 + 36.7 \cdot \log_{10}(d)$ | dB |
| Densidade espectral do ruído | $3.16 \cdot 10^{-20}$ | W/Hz |
| Número de TTIs (para cada requisito de QoS) | 1000 | - |
| N. de serviços | 2 (1 centro e 1 borda) | - |
| N. de terminais por serviço | 3 (6 para Único Serviço) | - |
| N. de Células | 3 | - |
| N. de RBs | 15 | - |
| N. min. de usuários satisfeitos por serviço | 2 em cada (4 para Único serviço) | - |

Note que consideramos dois serviços para a simulação. Os mesmo requisitos de do inglês, *Quality of Service* (QoS) são aplicados em ambos, porém o serviço 1 agrupa usuários do centro da célula enquanto o serviço 2 agrupa os usuários de borda. O cenário simulado possui apenas um serviço real, porém múltiplos serviços podem ser considerados na formulação proposta na seção 4. O mesmo cenário também será aplicado para o **Único Serviço**, com a distinção que a restrição de usuários satisfeito será a soma do número mínimo de usuários de centro e de borda aplicados para a solução proposta.

Uma métrica avaliada foi a capacidade do método de resolver o problema no cenário adotado. Para isto analisaremos o *outage* ocorrido em cada TTI. Neste trabalho consideraremos que houve *outage* em um TTI caso o método não atinja o mínimo de satisfação de QoS para usuários de centro e de borda no problema proposto. Para medir probabilidade do método encontrar uma solução que respeite todas as restrições analisaremos a taxa de *outage*, que é a porcentagem de *outage* verificado nos TTIs simulados. A Figura 2 mostra o gráfico da taxa requisitada versus a taxa de *outage*.

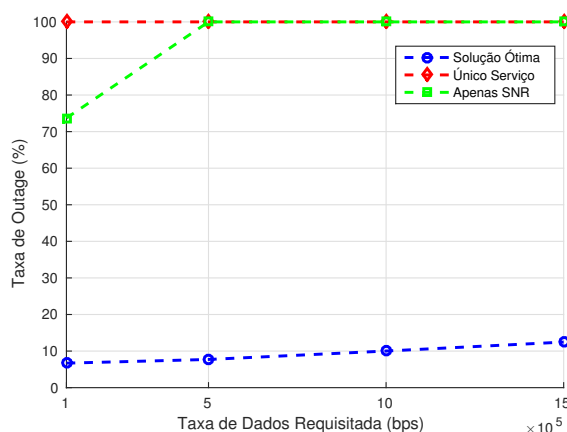


Figure 2. Taxa de *outage* versus taxa de dados requisitada para os métodos simulados.

Os piores resultados em termos de *outage* foram os métodos que consideram **Apenas SNR** [Lima et al. 2012] e **Único Serviço**, mostrando um resultado próximo de 100%

na maioria os requisitos de taxa. O desempenho do método **Apenas SNR** é facilmente explicado pelo fato de que a interferência é muito degradante em sistemas com pequenas células. Este resultado comprova a inviabilidade em desconsiderar a interferência em cenários mais densos e reforça a importância do estudo do gerenciamento de interferência.

O *outage* do método com apenas um serviço pode ser explicado pelo fato do mesmo não fazer distinções entre borda e centro da célula, sendo assim, existe uma tendência em satisfazer os usuários mais próximos do transmissor. Mesmo que o sistema tenha recursos suficientes para satisfazer todos os usuários, o objetivo de maximização da soma da taxa total fará que após a solução conseguir satisfazer as restrições mínima QoS, o sistema aloque recursos livres para melhores grupos que possuem as melhores SINR, pois estes grupos levarão a um maior ganho de taxa na soma total. Neste caso é esperado soluções com o menor número possível de usuários de borda satisfeitos, prejudicando os terminais móveis de borda. Para validar esta hipótese analisamos a média de usuários de centro de célula e borda satisfeitos em todo sistema considerando as soluções ótimas e **Único Serviço** em todos os TTIs. A Figura 3 contém a média de usuários satisfeitos para centro e borda em todo sistema para o maior requisito de taxa simulado, 100 kbps.

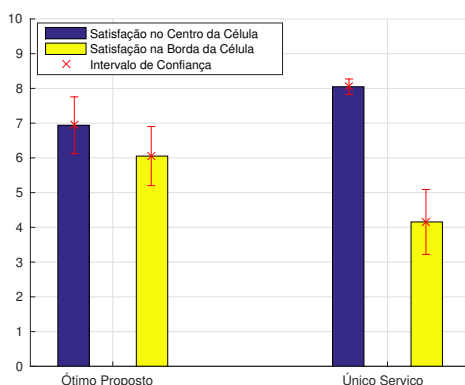


Figure 3. Média de usuários satisfeitos para requisito de taxa de 100 kbps.

Para este resultado consideramos os TTIs em que os métodos encontraram alguma solução. No caso do método **Único Serviço**, consideramos os TTIs em que há 4 usuários satisfeitos em cada célula, mesmo que a solução não tenha atingido o número mínimo de usuários satisfeitos na borda e no centro da célula, o que contribui para a taxa de *outage* mostrado anteriormente. É importante destacar que a solução ótima proposta satisfaz em média 7 usuários de centro e 6 de borda de célula. No total o sistema simulado possui 9 usuários em cada caso, sendo necessário satisfazer os 6 móveis obtidos na solução ótima. Em contra partida um desbalanceamento entre a satisfação dos usuários é verificada na solução com **Único Serviço**. Em média as soluções apresentaram satisfação em 8 usuários do centro contra 4 da borda, ferindo o requisito de justiça adotado. Este resultado evidencia a falta de justiça da solução sem requisitos de taxa próprios para usuários de borda, o que prejudica a experiência uniforme dos usuários na cobertura do sistema.

6. Conclusões

A densificação é um dos principais métodos para ganhos de capacidade em sistemas de telecomunicações e é considerada uma das principais linhas de pesquisa para a quinta geração de comunicações móveis. O aumento da densidade de Estação Rádio Base (ERB)

além de permitir um maior reuso de recursos de rádios no sistema também implica no aumento da interferência que se torna um limitante. Técnicas de gerenciamento de interferência vêm sendo estudados na literatura para diversos objetivos como, aumento da soma de taxa total do sistema, maximização da mínima SINR (do inglês, *Signal to Interference plus Noise Ratio*), entre outros.

Uma forma de mitigar a interferência é administrar o reuso de recursos em cada célula de forma a explorar os ganhos dos canais de interesse e interferente eliminando (ou diminuindo) assim a interferência. Neste contexto este trabalho propõe uma solução ótima para alocação de recursos com o objetivo de maximizar a taxa de dados em um sistema multisserviços, considerando critérios de QoS (do inglês, *Quality of Service*) em termos de taxa mínima por serviço e número mínimo de usuários atendidos pelo sistema mantendo justiça entre usuários de centro de borda das células. O problema proposto foi linearizado na forma de um ILP (do inglês, *Integer Linear Program*) e transformado em uma versão compacta que possibilita o uso de *softwares* comerciais para a solução. Não foi encontrada na literatura até o momento alguma proposta com estas características para cenário com multicelulas, tornando a proposta um novo critério de comparação para futuros trabalhos relacionados.

Apesar da solução ótima do problema ser possível utilizando BB (do inglês, *Branch-and-Bound*), a complexidade computacional da solução é exponencial, o que dificulta sua aplicação em cenários com número maior de usuários. Uma perspectiva de trabalho futuro é a proposta de uma solução heurística de baixa complexidade para o problema.

Agradecimento

Este trabalho foi apoiado pelo *Innovation Center Ericsson Telecomunicações S.A.*, Brasil sob contrato de cooperação técnica EDB/UFC.42. O estudante László R. Costa gostaria de agradecer a Fundação Cearense de Apoio a Pesquisa (FUNCAP) pelo suporte financeiro.

References

- 3GPP (2009). Evolved Universal Terrestrial Radio Access (E-UTRA). Technical Report TR 36.213, Third Generation Partnership Project.
- 3GPP (2012). Feasibility study for further advancements for e-utra (lte-advanced). Technical Report TR 36.912 V11.0.0, Third Generation Partnership Project.
- Akbari, M. and Vakili, V. (2012). Resource allocation for ofdma systems with trade-off between throughput and quality of service. In *2012 Sixth International Symposium on Telecommunications (IST)*, pages 420–425.
- Bhushan, N., Li, J., Malladi, D., Gilmore, R., Brenner, D., Damnjanovic, A., Sukhvasi, R., Patel, C., and Geirhofer, S. (2014). Network densification: the dominant theme for wireless evolution into 5g. *IEEE Communications Magazine*, 52(2):82–89.
- Bohge, M. (2010). *Dynamic Resource Allocation in Packet-Oriented Multi-Cell OFDMA Systems*. PhD thesis, Berlin Technology University.
- Chu, D. (1972). Polyphase codes with good periodic correlation properties (corresp.). *IEEE Transactions on Information Theory*, 18(4):531–532.
- Gotsis, A. G. and Alexiou, A. (2013). On coordinating ultra-dense wireless access networks: Optimization modeling, algorithms and insights. *CoRR*, abs/1312.1577.

- Heimiller, R. (1961). Phase shift pulse codes with good periodic correlation properties. *IRE Transactions on Information Theory*, 7(4):254–257.
- IBM (2009). IBM ILOG CPLEX Optimizer. <http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/>. Acessado em 09 de dezembro de 2015.
- Jang, J. and Lee, K. B. (2003). Transmit power adaptation for multiuser ofdm systems. *IEEE Journal on Selected Areas in Communications*, 21(2):171–178.
- Kuhn, H. W. (2010). The Hungarian Method for the Assignment Problem. In Jünger, M., Liebling, T. M., Naddef, D., Nemhauser, G. L., Pulleyblank, W. R., Reinelt, G., Rinaldi, G., and Wolsey, L. A., editors, *50 Years of Integer Programming 1958-2008*, pages 29–47, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Lima, F. R. M., Maciel, T. F., Freitas, W. C., and Cavalcanti, F. R. P. (2012). Resource assignment for rate maximization with qos guarantees in multiservice wireless systems. *IEEE Transactions on Vehicular Technology*, 61(3):1318 –1332.
- Lopez-Perez, D., Chu, X., Vasilakos, A., and Clausen, H. (2014). Power minimization based resource allocation for interference mitigation in ofdma femtocell networks. *IEEE Journal on Selected Areas in Communications*, 32(2):333–344.
- Mohsen, N. and Hassan, K. S. (2015). C-ran simulator: A tool for evaluating 5g cloud-based networks system-level performance. In *IEEE International Conference: Wireless and Mobile Computing, Networking and Communications (WiMob)*,, pages 302–309.
- Nemhauser, G. L. and Wolsey, L. A. (1988). *Integer and Combinatorial Optimization*. Wiley-Interscience, New York, NY, USA.
- Rahman, M. and Yanikomeroglu, H. (2007). Multicell downlink ofdm subchannel allocations using dynamic intercell coordination. In *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, pages 5220–5225.
- Rahman, M. and Yanikomeroglu, H. (2008). Interference avoidance through dynamic downlink ofdma subchannel allocation using intercell coordination. In *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, pages 1630–1635.
- Sadr, S. and Adve, R. (2014). Partially-distributed resource allocation in small-cell networks. *IEEE Transactions on Wireless Communications*, 13(12):6851–6862.
- Sadr, S., Anpalagan, A., and Raahemifar, K. (2009). Radio Resource Allocation Algorithms for the Downlink of Multiuser OFDM Communication Systems. *IEEE Communications Surveys Tutorials*, 11(3):92 –106.
- Sierksma, G. (2001). *Linear and integer programming: theory and practice*. CRC Press, 2nd edition.
- Song, G. and Li, Y. (2005). Cross-layer optimization for ofdm wireless networks-part i: theoretical framework. *IEEE Transactions on Wireless Communications*, 4(2):614–624.
- Zhang, Y. J. and Letaief, K. B. (2004). Multiuser Adaptive Subcarrier-and-Bit Allocation with Adaptive Cell Selection for OFDM Systems. *IEEE Transactions on Wireless Communications*, 3(5):1566–1575.