

# Alocação de Recursos Adaptativa para Maximização da Satisfação dos Usuários em Redes Celulares

Roberto P. Antonioli, Emanuel B. Rodrigues, Tarcisio F. Maciel, Diego A. Sousa, Fco. Rodrigo P. Cavalcanti

**Resumo**—Este artigo propõe um algoritmo de alocação de recursos de rádio adaptativo para o enlace direto de um sistema de telecomunicações que oferece múltiplos serviços. O algoritmo proposto é baseado em teoria da utilidade e objetiva maximizar simultaneamente a satisfação dos usuários de todos os serviços do sistema enquanto protege a satisfação de um determinado serviço com maior prioridade. O algoritmo é escalável para diversas classes de serviços e pode ser aplicado nos sistemas sem fio atuais e futuros. A avaliação de desempenho conduzida em cenários 4G LTE realistas mostrou que o algoritmo proposto proporciona ganhos na capacidade do sistema.

**Palavras-Chave**—Múltiplos Serviços, Maximização da Satisfação do Usuário, Teoria da Utilidade, Alocação de Recursos de Rádio.

**Abstract**—This article proposes an adaptive radio resource allocation algorithm for the downlink of wireless communication systems that offer multiple services. The proposed algorithm is based on utility theory and aims to maximize the user satisfaction of all services in the system simultaneously while protecting the satisfaction of a given service with higher priority. The algorithm is scalable to several service classes and can be applied in the current and future wireless systems. A performance evaluation conducted in realistic 4G LTE scenarios demonstrated that the proposed algorithm provides gains in the system capacity.

**Keywords**—Multiple Services, User Satisfaction Maximization, Utility Theory, Radio Resource Allocation.

## I. INTRODUÇÃO

As redes celulares de banda larga do século XXI são caracterizadas por um número cada vez maior de dispositivos móveis (do inglês *User Equipments* (UEs)) que acessam uma ampla variedade de serviços. Os serviços acessados por estes dispositivos variam desde aqueles com características de tempo real (do inglês *Real Time* (RT)) como vídeo, voz sobre IP (do inglês *Voice over IP* (VoIP)) e outras aplicações multimídia, até serviços de tempo não-real (do inglês *Non-Real Time* (NRT)), tais como a transferência de arquivos, navegação na Internet e aplicações com taxa de bits constante (do inglês *Constant Bit Rate* (CBR)).

Considerando a perspectiva das operadoras das redes celulares, o principal objetivo é utilizar o espectro eletromagnético e orçamento limitados da melhor maneira possível para maximizar a receita obtida fazendo com que os usuários fiquem satisfeitos. Satisfazer usuários significa atender aos seus requisitos de qualidade de serviço (do inglês *Quality of Service* (QoS)) [1], o que não é uma tarefa simples tendo em vista

que as redes celulares mais recentes fornecem uma ampla variedade de serviços com requisitos de QoS conflitantes.

Neste contexto, algoritmos de alocação de recursos de rádio (do inglês *Radio Resource Allocation* (RRA)) tornam-se um dos mecanismos mais importantes usados pelas operadoras para garantir a satisfação do usuário. Vários algoritmos de RRA são discutidos em [2]. O algoritmo *Modified Largest Weighted Delay First* (MLWDF) proposto em [3] considera a informação do estado do canal (do inglês *Channel State Information* (CSI)) do usuário e tamanho da fila de pacotes com o objetivo de manter o atraso do pacote do começo da fila (do inglês *Head Of Line* (HOL)) da maioria dos usuários abaixo de um dado requisito. O algoritmo MLWDF foi estendido em [4] e chamado de *Queue-HOL-MLWDF* (QHMLWDF) para que a RRA considerasse o tamanho da fila de pacotes e o atraso do pacote HOL. Em [5], o algoritmo *Max-Delay-Utility* (MDU) foi proposto empregando funções de utilidade do tipo degrau para definir a prioridade dos usuários na RRA. Um algoritmo de RRA baseado na CSI e funções de utilidade baseadas no atraso do pacote HOL foi proposto em [6], em que uma função com formato de Z é usada para escalonar os usuários de serviços RT e uma função exponencial é usada para escalonar os usuários de serviços NRT. Em [7], o algoritmo clássico *Exponential* (EXP) foi estendido e um algoritmo de RRA foi proposto, chamado de *EXP/Proportional Fair* (PF), que escalona os usuários RT e NRT segundo os algoritmos EXP e PF, respectivamente.

Diferentemente dos algoritmos encontrados na literatura, este trabalho propõe um algoritmo de RRA baseado na teoria da utilidade que visa maximizar a satisfação dos usuários no enlace direto de redes celulares que fornecem múltiplos serviços. A principal contribuição deste trabalho é um algoritmo de RRA com as seguintes características:

- **Normalizado:** as prioridades baseadas em utilidade para todos os serviços são equivalentes em magnitude e podem ser comparadas sem considerar nenhum ajuste;
- **Adaptativo:** uma função de utilidade que define a prioridade dos serviços é adaptada dinamicamente por meio de passos não-lineares;
- **Flexível:** a operadora pode escolher um serviço para ter maior prioridade e o algoritmo proposto garante um limiar mínimo de satisfação para este serviço. Esta característica ainda não foi explorada na literatura.

O restante deste trabalho é organizado da seguinte forma. Na Seção II apresentamos a modelagem do sistema. As Seções III e IV contêm o detalhamento da formulação do problema de otimização e o algoritmo de RRA derivado da

Os autores são vinculados ao Grupo de Pesquisa em Telecomunicações Sem Fio (GTEL), Universidade Federal do Ceará (UFC), Fortaleza-CE. Diego Aguiar Sousa é vinculado ao Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), Acaraú-CE.

mesma. Na Seção V mostramos as funções de utilidade empregadas no algoritmo proposto. A avaliação do desempenho é realizada na Seção VI e, finalmente, a Seção VII apresenta as conclusões do trabalho.

## II. MODELO DO SISTEMA

Neste artigo é considerado um cenário composto de uma única célula de um sistema baseado no padrão de evolução de longo prazo (do inglês LTE) operando no enlace direto. Neste cenário, uma estação rádio-base (do inglês *Base Station* (BS)) localizada no centro da célula serve o conjunto  $\mathcal{J} = \{1, 2, \dots, J\}$  dos UEs localizados dentro da região de cobertura da célula.

O sistema considerado emprega múltiplo acesso por divisão em frequências ortogonais. Devido a restrições de sinalização, a unidade mínima de alocação considerada aqui é um bloco de recursos de rádio (do inglês *Resource Block* (RB)) com duração de 1 ms (intervalo de tempo de transmissão, do inglês *Transmission Time Interval* (TTI)) no domínio do tempo e de 180 kHz (12 subportadoras de 15 kHz) no domínio da frequência. A BS realiza a alocação dos RBs representados por  $\mathcal{K} = \{1, 2, \dots, K\}$  considerando que cada RB pode ser alocado para apenas um UE durante um TTI.

Assume-se que a BS e os UEs possuem apenas uma antena cada para fazer transmissões e recepções. Os coeficientes de canal  $h_{j,k}$  entre a BS e o UE  $j \in \mathcal{J}$  no RB  $k \in \mathcal{K}$  englobam os principais efeitos de propagação do canal (perda de percurso, sombreamento e desvanecimento em pequena escala) e ganhos na transmissão ou recepção proporcionados pelas antenas.

Um termo de ruído e uma componente de atraso são inseridos na estimação da CSI, o que caracteriza uma imperfeição na estimação do canal. O modelo de imperfeição na estimação do canal é  $\hat{h}_{j,k}[n] = \sqrt{(1-\psi)}h_{j,k}[n] + \sqrt{\psi}\eta[n]$ , em que:  $h_{j,k}[n]$  seria o valor da estimação do estado do canal caso não houvesse erro na estimação;  $\hat{h}_{j,k}[n]$  é o valor da CSI após inserção do erro (ruído);  $\psi \in (0, 1)$  representa a degradação da estimação do canal;  $\eta[n] \in \mathbb{C}$  representa um erro que é inserido na estimação do canal [8]. Além disso, os valores da estimação do canal que é reportado pelos UEs para a BS e que é usado pelos algoritmos de RRA estão atrasados por uma certa quantidade  $\Delta n$  de TTIs.

Assumimos que a potência total  $P_t$  da BS passa por uma distribuição igualitária de potência (do inglês *Equal Power Allocation* (EPA)) entre todos os RBs. Logo, a potência  $p_k$  alocada ao RB  $k$  é igual a  $p_k = \frac{P_t}{K}$ . O valor da relação sinal ruído de um UE  $j$  no RB  $k$  é  $\gamma_{j,k} = p_k \frac{|h_{j,k}|^2}{\sigma^2}$ , em que  $\sigma^2$  representa a potência média do ruído aditivo gaussiano branco na banda de frequência de um RB.

A modelagem utilizada considera que a BS utiliza uma técnica de adaptação de enlace que permite transmissões de taxas de dados diferentes de acordo com o valor estimado de  $\gamma_{j,k}$  para um usuário  $j$  no RB  $k$ . Portanto, a taxa alocada pela BS para o usuário  $j$  no RB  $k$  é  $r_{j,k} = f(\gamma_{j,k})$ , em que  $f(\cdot)$  representa a função de adaptação de enlace. A taxa total do usuário  $j$  é então dada por  $R_j = \sum_{k \in \mathcal{K}_j} r_{j,k}$ , em que  $\mathcal{K}_j \subset \mathcal{K}$  é o subconjunto de RBs alocados para o usuário  $j$ .

## III. FORMULAÇÃO DO PROBLEMA

A teoria da utilidade, que foi originalmente concebida para aplicações no ramo da economia, nos permite capturar o nível de satisfação dos usuários para uma determinada alocação de recursos. Além disso, ela fornece a possibilidade de combinarmos métricas das camadas física e da (sub)camada de acesso ao meio para obter uma melhor otimização entre camadas na alocação dos recursos [5].

Com o embasamento fornecido pela teoria da utilidade, formulou-se um problema geral de otimização com o objetivo de maximizar a utilidade dos usuários derivada da rede (satisfação) para todas as distintas classes de serviço simultaneamente. Considerando as definições da Seção II, o problema estudado foi formulado como

$$\text{maximizar}_{\rho_{j,k}, p_k} \sum_{j \in \mathcal{J}} V[U(x_j)] \quad (1a)$$

$$\text{sujeito a } \rho_{j,k} \in \{0, 1\}, \forall j \in \mathcal{J} \text{ e } \forall k \in \mathcal{K}, \quad (1b)$$

$$\sum_{j \in \mathcal{J}} \rho_{j,k} = 1, \forall k \in \mathcal{K}, \quad (1c)$$

$$\sum_{k \in \mathcal{K}} p_k = P_t, \quad (1d)$$

$$p_k \geq 0, \forall k \in \mathcal{K}, \quad (1e)$$

em que:  $\rho_{j,k}$  é uma variável de assinalamento que assume o valor 1 se o RB  $k$  é alocado para o usuário  $j$ , e 0 caso contrário;  $U(x_j)$  é a função de utilidade do usuário baseada em uma variável genérica  $x_j$  que pode representar um determinado uso de um recurso ou uma métrica de QoS associada ao usuário  $j$ ; e  $V(\cdot)$  é uma função de utilidade de serviço que depende das funções de utilidade dos usuários  $U(\cdot)$  e diferencia as classes de serviço presentes no sistema. As restrições (1b) e (1c) garantem que um determinado RB não pode ser utilizado por dois ou mais usuários no mesmo TTI. As restrições (1d) e (1e) impõem que a soma das potências alocadas para todos os RBs não pode ser maior que a potência total disponível e que essas potências devem ser não-negativas.

Em geral, a solução ótima do problema de otimização proposto é muito difícil de ser encontrada. Portanto, utilizamos uma abordagem para achar uma solução sub-ótima que divide o problema original em duas etapas: (i) primeiro, uma alocação dinâmica dos recursos é feita considerando uma alocação de potência fixa, e (ii) depois fazemos uma EPA, com uma atribuição fixa dos recursos [9].

## IV. SOLUÇÃO SUBÓTIMA

O algoritmo proposto neste trabalho é uma extensão para múltiplos serviços dos algoritmos para cenários com serviço único RT ou NRT apresentados em [1]. Em [1] demonstrou-se que é possível derivar problemas de otimização equivalentes aos problemas originais de maximização da utilidade para os serviços NRT ou RT, separadamente, em que o problema de otimização simplificado tem função objetivo em termos da taxa instantânea do usuário.

Considerando o cenário com serviços RT e NRT, o problema de otimização é formulado para maximizar a utilidade total em termos das métricas de QoS dos usuários, ou seja, vazão de

dados (do inglês *throughput*) e atraso do pacote HOL para os serviços NRT e RT, respectivamente. Assumimos ainda que o conjunto total de usuários  $\mathcal{J}$  é separado em dois subconjuntos:  $\mathcal{J}_{\text{nrt}}$  e  $\mathcal{J}_{\text{rt}}$  para os usuários NRT e RT, respectivamente. Para este cenário com múltiplos serviços, a função objetivo do problema de otimização geral (1) é reformulada como:

$$\max_{\rho_{j,k}, P_k} \left\{ \sum_{j \in \mathcal{J}_{\text{nrt}}} V[U_{\text{nrt}}(T_j[n])] + \sum_{j \in \mathcal{J}_{\text{rt}}} V[U_{\text{rt}}(d_j^{\text{hol}}[n])] \right\}. \quad (2)$$

A função de utilidade de serviço  $V(\cdot)$  é utilizada nos cenários com múltiplos serviços para diferenciar a prioridade dos serviços dependendo do objetivo a ser atingido. Essa função pode ser projetada: (i) para fornecer prioridade igual para todos os serviços; (ii) para fornecer uma prioridade maior fixa para um determinado serviço; ou ainda (iii) para ser adaptada dinamicamente a fim de atender algum requisito de QoS de um serviço com maior prioridade.

Seguindo a mesma abordagem utilizada em [1], podemos simplificar o problema de otimização estudado de tal forma que a função objetivo simplificada torne-se linear em termos da taxa instantânea dos usuários (a demonstração matemática detalhada dessa simplificação está fora do escopo deste trabalho). A função objetivo simplificada é dada por

$$\begin{aligned} & \max_{\rho_{j,k}, P_k} \left\{ \sum_{j \in \mathcal{J}_{\text{nrt}}} V[U_{\text{nrt}}(T_j[n])] + \sum_{j \in \mathcal{J}_{\text{rt}}} V[U_{\text{rt}}(d_j^{\text{hol}}[n])] \right\} \approx \\ & \max_{\rho_{j,k}, P_k} \left\{ \sum_{j \in \mathcal{J}} w_j^s \cdot w_j \cdot R_j[n] \right\} = \\ & \max_{\rho_{j,k}, P_k} \left\{ \sum_{j \in \mathcal{J}_{\text{nrt}}} w_j^s \cdot w_j^{\text{nrt}} \cdot R_j[n] + \sum_{j \in \mathcal{J}_{\text{rt}}} w_j^s \cdot w_j^{\text{rt}} \cdot R_j[n] \right\}, \quad (3) \end{aligned}$$

em que:  $w_j^s$  é a prioridade de serviço do usuário  $j$  e serviço  $s$ ;  $w_j$  é a prioridade de usuário associada ao usuário  $j$ ; e  $R_j[n]$  é a taxa instantânea do usuário  $j$  no TTI  $n$ .

A prioridade do usuário é mais especificamente definida como  $w_j^{\text{nrt}} = U'_{\text{nrt}}(T_j[n-1])$  se  $j$  é um usuário NRT ou  $w_j^{\text{rt}} = |U'_{\text{rt}}(d_j^{\text{hol}}[n])|$  se  $j$  é um usuário RT [1]. Note que  $U_{\text{nrt}}(T_j[n])$  é uma função de utilidade crescente em função do *throughput* do usuário  $j$  no TTI  $n$ ;  $U'_{\text{nrt}}(T_j[n-1])$  é a utilidade marginal do usuário  $j$  dada pela derivada de  $U_{\text{nrt}}$  em relação ao *throughput* do usuário  $j$  no TTI  $n-1$ ;  $U_{\text{rt}}(d_j^{\text{hol}}[n])$  é uma função de utilidade decrescente em função do atraso do pacote HOL  $d_j^{\text{hol}}[n]$  do usuário  $j$  no TTI  $n$ ; e  $U'_{\text{rt}}(d_j^{\text{hol}}[n])$  é a utilidade marginal do usuário  $j$  dada pela derivada de  $U_{\text{rt}}$  em relação ao atraso do pacote HOL do usuário no TTI  $n$ . A prioridade de serviço é definida como  $w_j^s = V'(U_{\text{nrt}}(T_j[n-1]))$  se  $j$  é um usuário NRT ou  $w_j^s = V'(U_{\text{rt}}(d_j^{\text{hol}}[n]))$  se  $j$  é um usuário RT. A expressão para  $w_j^s$  é obtida utilizando um processo similar ao utilizado para obter  $w_j$  em [1] e a regra da cadeia das derivadas (a demonstração detalhada é omitida por limitação de espaço).

A função objetivo simplificada dada por (3) caracteriza um problema de maximização de taxa ponderada [10], em que os pesos são  $w_j^s$  e  $w_j$ , e é linear em termos de  $R_j[n]$ . Devido a essa linearidade, quando a função objetivo é dada por (3), o problema de alocação de recursos dinâmica descrito tem uma solução fechada [5]. Considerando o caso mais realista e

complexo que é o cenário com múltiplos serviços, a solução de (3) é que o usuário com índice  $j^*$  é escolhido para receber dados durante o TTI  $n$  no RB  $k$  se ele satisfaz a condição

$$j^* = \arg \max_j \{ w_j^s \cdot w_j \cdot r_{j,k}[n] \}, \quad (4)$$

em que  $r_{j,k}[n]$  denota a taxa máxima de transmissão que o usuário pode atingir no RB  $k$  no TTI  $n$ . Se dois ou mais usuários tiverem a mesma prioridade no RB  $k$ , o usuário com maior relação sinal-ruído é selecionado para transmissão.

Vale ressaltar a simplicidade do algoritmo proposto. No início, tínhamos um problema de otimização baseado em utilidade em que a solução ótima é muito difícil de ser encontrada. Após uma simplificação, uma equação para um algoritmo de RRA foi derivada, a qual é simples de ser implementada já que ela envolve apenas a multiplicação de dois pesos ( $w_j^s$  e  $w_j$ ) por  $r_{j,k}[n]$ .

## V. FUNÇÕES DE UTILIDADE

A política de alocação de recursos proposta aqui é chamada de *Joint Satisfaction Maximization* (JSM) e é uma extensão para múltiplos serviços das políticas para serviço único *Throughput-based Satisfaction Maximization* (TSM) e *Delay-based Satisfaction Maximization* (DSM) propostas em [1], onde os autores demonstraram que o uso de funções sigmoideais proporcionam níveis de satisfação do usuário melhores se comparado com algoritmos clássicos de RRA.

Neste trabalho, a função logística é utilizada como função de utilidade do usuário, a qual é descrita pela equação

$$U(x_j[n]) = \frac{1}{1 + e^{\mu(x_j[n] - x_j^{\text{req}})/\sigma}}. \quad (5)$$

Essa função foi escolhida devido à sua parametrização, em que: o parâmetro  $\sigma$  é um valor não negativo que regula a inclinação da função;  $\mu$  define se a função é crescente ( $\mu = -1$ ) ou decrescente ( $\mu = 1$ ); e  $x_j^{\text{req}}$  é o requisito de QoS do serviço e define o deslocamento horizontal da função.

A variável de entrada  $x_j[n]$  é a métrica de QoS de cada serviço. Como cada serviço tem sua métrica particular de QoS, com o intuito de normalizar o algoritmo proposto e deixá-lo independente da métrica de QoS, dividimos as variáveis  $x_j[n]$  e  $x_j^{\text{req}}$  pelo requisito de QoS ( $x_j^{\text{req}}$ ).

Como já mencionado, a função  $U_{\text{nrt}}(T_j[n])$ , utilizada pela política TSM, é uma função crescente ( $\mu = -1$ ), o que significa que um determinado usuário torna-se satisfeito rapidamente quando sua vazão de dados aumenta e fica maior que o valor de requisito de vazão de dados. A função  $U_{\text{rt}}(d_j^{\text{hol}}[n])$ , utilizada pela política DSM, é decrescente ( $\mu = 1$ ), o que implica que quanto menor o atraso do pacote HOL, maior será a satisfação do usuário. Resultados satisfatórios foram obtidos quando os parâmetros são  $x_j^{\text{req}} = 1$  e  $\sigma = 0,1088$ .

A principal inovação deste trabalho é a proposição da função de utilidade de serviços, que proporciona uma diferenciação adaptável na prioridade dos serviços. A função escolhida para exercer esse papel foi

$$V(z_j[n]) = \log(1 + e^{(z_j[n]-1)/\lambda}). \quad (6)$$

Essa função é uma versão escalada da tangente hiperbólica e nos permite definir regiões para a diferenciação dos serviços,

de tal forma que cada serviço tenha uma prioridade relativa à magnitude da função na região correspondente.

A função marginal do serviço  $V'(z_j[n])$ , ou prioridade de serviço  $w_j^s$ , é obtida pela derivada primeira de  $V(z_j[n])$ , o que resulta em uma função marginal logística. Os parâmetros de  $V'(z_j[n])$  são:  $\lambda$  é o parâmetro que controla a inclinação da função; e  $z_j[n]$  é  $U_{\text{nrt}}(T_j[n-1])$  se  $j$  é um usuário NRT e  $U_{\text{rt}}(d_j^{\text{hol}}[n])$  se  $j$  é um usuário RT.

Neste trabalho, nós consideramos que a prioridade de serviço  $w_j^s$  é adaptada dinamicamente para proteger um serviço com maior prioridade. Como normalmente feito na literatura, nós escolhemos que o serviço a ser priorizado é aquele com características RT. Na nossa abordagem, a satisfação dos usuários RT deve permanecer sempre acima de 90% por meio da adaptação do parâmetro  $\lambda$  da função de utilidade do peso do serviço. A adaptação deste parâmetro é feita por meio de uma tabela de pesquisa (do inglês *look-up table*), que foi calculada usando uma função de interpolação cúbica para que as curvas obtidas ficassem igualmente espaçadas entre os valores de  $\lambda = 0,1088$  e  $\lambda = -0,1088$ . Para  $\lambda = 0,1088$ , a função marginal do serviço assume um formato em que a prioridade do serviço RT é máxima, enquanto que para  $\lambda = -0,1088$  a prioridade do serviço NRT é máxima. Além disso, para que tivéssemos um formato da função com prioridade igual para ambos os serviços, o valor central da tabela de pesquisa é  $\lambda = 5.000$ . O resultado foi uma tabela de pesquisa com 41 valores de  $\lambda$  não-linearmente espaçados.

## VI. AVALIAÇÃO DE DESEMPENHO

Nesta seção, analisamos o desempenho do JSM e o comparamos com outros 4 algoritmos da literatura. O modelo de sistema descrito na Seção II foi utilizado para todas as simulações. A Tabela I apresenta os principais parâmetros de simulação, os quais estão alinhados com a arquitetura do *3rd Generation Partnership Project (3GPP)* [11], [12].

TABELA I  
PARÂMETROS DE SIMULAÇÃO.

Parâmetro	Valor	Ref.'s
Pot. de transmissão da eNB ( $P_t$ )	12 W	[11]
Padrão de radiação da antena	Tri-setorizada	[11]
Raio da célula	1 km	
Velocidade do UE	3 km/h	[11]
Frequência central da portadora	2 GHz	[11]
Largura de banda	3 MHz	[13]
Número de RBs ( $K$ )	15	[13]
Perda de percurso	$128.1 + 37.6 \log_{10}(d)^a$	[11]
Ganho de antena	$G_h(\theta_h) + G_v(\theta_v)^b$	[14]
Desv. padrão do somb. log-normal	8 dB	[11]
Desvanecimento de pequena escala	3GPP Typical Urban	[12]
Potência do ruído por subportadora	-123.24 dBm	
Figura de ruído	9 dB	
Adaptação de enlace	Curvas de enlace do LTE	[15]
Degradação da CSI ( $\psi$ )	0,05	
Atraso da CSI ( $\Delta n$ )	15 TTIs	
Modelo de tráfego RT/NRT	VoIP/CBR	
Duração do TTI	1 ms	
Duração de cada simulação	30 s (30.000 TTIs)	
Número de realizações	150	

<sup>a</sup>  $d$  é a distância entre a BS e o UE em km.

<sup>b</sup>  $\theta_h$  e  $\theta_v$  são os ângulos horizontal e vertical relativos à BS.

O cenário com múltiplos serviços utilizado foi composto por dois serviços: CBR, que é um serviço NRT e VoIP, que é um serviço RT. Os dois serviços foram modelados com cadeias de Markov de dois estados. O período de tempo que os usuários permanecem em cada estado é calculado por meio de uma distribuição de probabilidade exponencial com média unitária.

O serviço VoIP gera pacotes de 320 bits a cada 20 ms durante o período ativo e os pacotes que não forem transmitidos dentro de 20 ms são descartados. Um usuário VoIP é considerado satisfeito se sua taxa de perda de pacotes é menor que 2% [1]. O serviço CBR gera pacotes de 2048 bits a cada 4 ms durante o período ativo. Um usuário CBR é considerado satisfeito se o seu *throughput* é maior que 512 Kbps [1].

Como feito na literatura, o serviço com maior prioridade é aquele com características RT, ou seja, o VoIP. Portanto, a satisfação dos usuários VoIP deve permanecer sempre acima de 90% por meio da adaptação do parâmetro  $\lambda$  de  $w_j^s$ .

Os 4 algoritmos da literatura utilizados para comparação são: QHMLWDF [4], MDU [5], EXP/PF [7] e o algoritmo proposto em [6] (refere-se a esse algoritmo como *Lei*).

A Fig. 1 ilustra um gráfico radar para o cenário com 200 usuários VoIP e 5 usuários CBR. Cada eixo do gráfico representa uma métrica diferente de comparação: justiça (baseada no índice de justiça de Jain [1]) e satisfação de cada serviço, e o *throughput* total da célula normalizado. Percebe-se que considerando as métricas de justiça e *throughput* total, todos os algoritmos tem desempenho similar, exceto o MDU que apresenta o pior desempenho. Considerando esse cenário específico, podemos notar que o JSM apresenta o desempenho mais balanceado, já que ele consegue balancear a satisfação dos dois serviços sem penalizar um deles. Note que o EXPPF e MDU obtiveram melhor satisfação RT que o JSM. No entanto, a satisfação NRT fornecida por eles foi bastante penalizada, enquanto o JSM manteve a satisfação RT e NRT acima do limiar de 90% utilizando a adaptação do peso de serviço. Portanto, percebe-se que somente o JSM apresenta um balanço satisfatório entre as satisfações dos dois serviços.

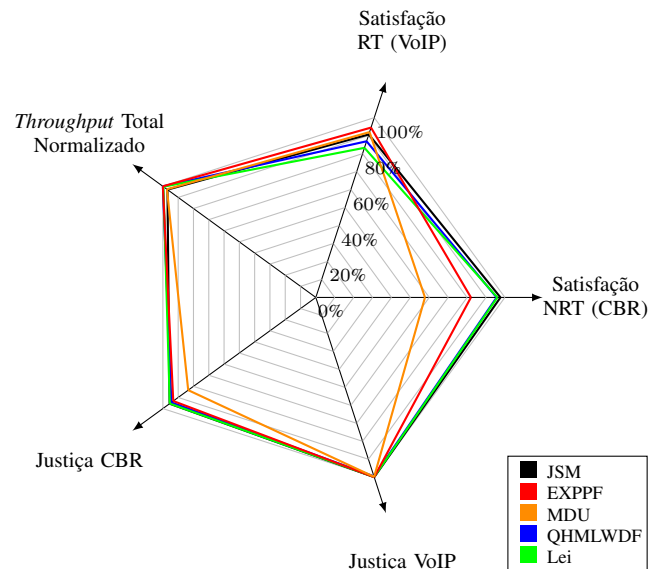


Fig. 1. Gráfico radar para cenário com 200 usuários VoIP e 5 usuários CBR.

Ampliando nossa análise, utilizamos o plano de capacidade conjunta, que é uma ferramenta poderosa para analisar o desempenho de algoritmos de RRA em cenários com múltiplos serviços. Este plano mostra as regiões de capacidade do sistema que são definidas como o número de usuários para o qual os níveis de QoS são mantidos acima de um determinado limiar simultaneamente para todas as classes de serviço.

Para cenários com um único serviço, a capacidade do sistema é a última carga de usuários tal que a satisfação é maior que 90%. Essas cargas compõem os eixos das abscissas e das ordenadas no plano de capacidade conjunta. Para cenários com múltiplos serviços, a capacidade do sistema é a carga máxima de usuários tal que os algoritmos conseguem fornecer níveis de satisfação maiores que 90% simultaneamente para todos os serviços. Essas cargas compõem os pontos internos do plano.

A Fig. 2 mostra o plano de capacidade conjunta obtido seguindo os critérios do parágrafo anterior. O desempenho de cada algoritmo é ilustrado como uma curva no plano de capacidade. Quanto maior for a área sob a curva, maior é a porcentagem de usuários satisfeitos respeitando um limite de satisfação mínimo de 90%. Portanto, nota-se claramente que o algoritmo JSM apresentou o melhor desempenho. O JSM foi capaz de aumentar significativamente a capacidade conjunta do sistema, atingindo ganhos totais (área sob a curva) que vão de 36% até 165%. Além disso, obteve-se ganhos que vão até 350% e 63% para usuários CBR e VoIP, respectivamente.

O principal motivo para o melhor desempenho do JSM é que ele consegue distribuir melhor os recursos principalmente devido à adaptação dinâmica da prioridade dos serviços. Além disso, o JSM aloca os recursos para os usuários CBR e VoIP com base no *throughput* e no atraso do pacote HOL, respectivamente. Com isso, o JSM explora as métricas de QoS que definem a satisfação dos usuários de cada serviço.

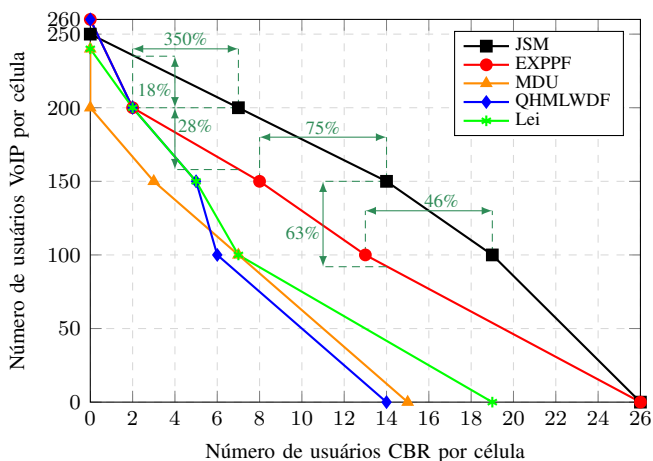


Fig. 2. Plano de capacidade conjunta do sistema para cenários com serviços CBR e VoIP.

## VII. CONCLUSÕES

Neste artigo, foi proposto e avaliado um algoritmo de RRA, fundamentado pela teoria da utilidade, que visa maximizar a satisfação dos usuários em sistemas celulares que fornecem múltiplos serviços. O JSM mantém um determinado nível de satisfação de um serviço com maior prioridade adaptando uma função de utilidade que diferencia a prioridade dos serviços.

Os resultados apresentados considerando um cenário com serviços RT (VoIP) e NRT (CBR) demonstraram que o JSM proporciona ganhos de 36% até 165% na capacidade conjunta do sistema. Além disso, um melhor compromisso entre a satisfação dos serviços no sistema é obtido pelo JSM.

Vale ressaltar ainda que aqui escolhemos o serviço RT para ter sua satisfação protegida. No entanto, a operadora tem a flexibilidade de escolher qualquer serviço, NRT ou RT, para ter a satisfação protegida pelo JSM.

## AGRADECIMENTOS

Agradecemos o apoio técnico-financeiro da Ericsson Research, Wireless Access Network Department, Suécia e do Ericsson Innovation Center, Brasil, com relação ao Contrato de Coperação Técnica EDB/UFC.40. Roberto P. Antonioli agradece à FUNCAP pela bolsa de estudos. Os autores agradecem ao Dr. Carlos F. M. e Silva pelo paciente trabalho de revisão deste trabalho.

## REFERÊNCIAS

- [1] E. B. Rodrigues, F. R. M. Lima, T. F. Maciel, and F. R. P. Cavalcanti, "Maximization of user satisfaction in OFDMA systems using utility-based resource allocation," *Wireless Commun. and Mob. Computing*, pp. 1–17, 2014.
- [2] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in LTE cellular networks: Key design issues and a survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 678–700, May 2013.
- [3] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.*, vol. 39, no. 2, pp. 150–154, Feb 2001.
- [4] M. M. Nasralla and M. G. Martini, "A downlink scheduling approach for balancing QoS in LTE wireless networks," in *Proc. of the IEEE Personal, Indoor and Mob. Radio Commun. (PIMRC)*, Sept 2013, pp. 1571–1575.
- [5] G. Song, "Cross-layer resource allocation and scheduling in wireless multicarrier networks," Ph.D. dissertation, Georgia Institute of Technology, Georgia, USA, 2005.
- [6] H. Lei, L. Zhang, X. Zhang, and D. Yang, "A packet scheduling algorithm using utility function for mixed services in the downlink of OFDMA systems," in *Proc. of the IEEE Vehic. Tech. Conf. (VTC)*, Sept 2007, pp. 1664–1668.
- [7] R. Basukala, H. A. M. Ramli, and K. Sandrasegaran, "Performance analysis of EXP/PF and M-LWDF in downlink 3GPP LTE system," in *First Asian Himalayas Internat. Conf. on Internet*, Nov 2009, pp. 1–5.
- [8] T. F. Maciel and A. Klein, "On the performance, complexity, and fairness of suboptimal resource allocation for multiuser MIMO-OFDMA systems," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 406–419, Feb. 2010.
- [9] J. Gross and M. Bohge, "Dynamic mechanisms in OFDM wireless systems: A survey on mathematical and system engineering contributions," Telecommunication Networks Group (TKN), Technical University Berlin, Tech. Rep. TKN-06-001, May 2006.
- [10] L. M. C. Hoo, B. Halder, J. Tellado, and J. M. Cioffi, "Multiuser transmit optimization for multicarrier broadcast channels: asymptotic FDMA capacity region and algorithms," *IEEE Trans. Commun.*, vol. 52, no. 6, pp. 922–930, June 2004.
- [11] 3GPP, "Evolved universal terrestrial radio access (E-UTRA): Further advancements for E-UTRA physical layer aspects," 3rd Generation Partnership Project (3GPP), TR 36.814, Mar. 2010.
- [12] —, "Deployment aspects," 3rd Generation Partnership Project (3GPP), TR 25.943, Dec. 2009.
- [13] M. Pelcat, S. Aridhi, and J. Piat, *Physical Layer Multi-Core Prototyping - A Dataflow-Based Approach for LTE eNodeB*, 1st ed. Springer Science & Business Media, 2013.
- [14] F. Gunnarsson, M. Johansson, A. Furuskar, M. Lundevall, A. Simonsson, C. Tildestav, and M. Blomgren, "Downtilted base station antennas - a simulation model proposal and impact on HSPA and LTE performance," in *Proc. of the IEEE Vehic. Tech. Conf. (VTC)*, Sep. 2008, pp. 1–5.
- [15] C. Mehlführer, M. Wrulich, J. C. Ikuno, D. Bosanska, and M. Rupp, "Simulating the Long Term Evolution physical layer," in *17th European Signal Processing Conf.*, Glasgow, Scotland, Aug. 2009, pp. 1471–1478.