

Explorando a afinidade de usuários para descarregamento de dados mais eficiente em redes celulares de pequeno porte

Adriana Viriato Ribeiro¹, Leobino N. Sampaio¹, Artur Ziviani²

¹Programa de Pós-Graduação em Ciência da Computação (PGCOMP)
Instituto de Matemática – Universidade Federal da Bahia (UFBA)
Salvador – BA – Brasil

²Laboratório Nacional de Computação Científica (LNCC)
Petrópolis, RJ - Brasil

adrianaavr@dcc.ufba.br, leobino@ufba.br, ziviani@lncc.br

Abstract. *Network-densification by Small Cell Networks has become the main alternative for mobile network operators when dealing with exponential traffic-growth. A Content-Centric Network can be an alternative to improve data offloading in this scenario. In this paper, we propose a user clustering strategy to take advantage of the similarity function that considers frequency and common-content requisitions to correlate users. Simulation results show that the proposed strategy can be adaptive to different scenarios and is able to increase hit ratio and minimize data offloading by 20% and 30%, respectively.*

Resumo. *A densificação de redes celulares através do uso de estações-base de pequeno porte é uma recente alternativa, adotada por operadoras de telefonia, para lidar com o crescente volume de tráfego de dados. No intuito de alcançar melhor taxa de descarregamento de dados, esse tipo de técnica tem sido associada às Redes Orientadas a Conteúdo. Neste artigo é proposta uma estratégia de agrupamento de usuários que explora a combinação da similaridade entre eles a partir dos conteúdos solicitados em comum e da frequência de solicitação a esses conteúdos. Os resultados de simulação demonstram a capacidade de adaptação da estratégia proposta a diferentes cenários e evidenciam melhoras de 20% em relação à taxa de acerto e de 30% ao descarregamento de dados.*

1. Introdução

A densificação de redes tem sido vislumbrada como uma das alternativas para as operadoras de telefonia móvel lidar com o crescente volume de tráfego de dados decorrente, principalmente, de serviços *streaming* [Aijaz et al. 2013, Bhushan et al. 2014]. Trata-se de uma maior oferta do serviço de telefonia a partir do aumento do número de antenas de transmissão para expandir a cobertura e prover descarregamento de dados (do inglês, *Data offloading*). Essas antenas geralmente são instaladas em pequenas estações-base (do inglês, *Small-Cell Base Station* – SCBS). As SCBS são constituídas por meio da implantação de equipamentos e tecnologias de transmissão de curto alcance, oferecendo, desta forma, redes celulares de pequeno porte (tradução adaptada do termo em inglês, *Small Cell Networks* – SCNs).¹

¹Na literatura, a expressão em inglês *Small cell* refere-se às chamadas “células pequenas” que englobam femto-células, pico-células e microcélulas. Essas redes aumentam a cobertura do serviço prestado pelas operadoras e são utilizadas pelos usuários na redução do congestionamento das macro-células.

As SCNs oferecem serviços complementares aos disponibilizados pelas macro estações-base (do inglês, *Macro Base Station* – MBS) com o propósito de aumentar a cobertura do serviço e atender às demandas de tráfego locais, melhorando os tempos de resposta da rede. Tais objetivos tem sido potencializados através das Redes Orientadas à Conteúdo (ROC) [Jacobson et al. 2009]. Propostas descritas na literatura sugerem o uso de ROCs em cenários de SCN para manter os conteúdos mais populares na SCBS e, conseqüentemente, mais próximos dos consumidores finais [ElBamby et al. 2014, Pantisano et al. 2015, Chen and Kountouris 2015, Hajri and Assaad 2016]. Dado que as políticas clássicas de gerenciamento de *cache*, tais como *Least Frequently Used* (LFU) e *Least Recently Used* (LRU), não levam em consideração as preferências individuais dos usuários, trabalhos mais recentes têm sugerido o uso de técnicas cientes de contexto [Chang et al. 2016] e de agrupamento com o objetivo de reunir usuários com base em características similares e associá-los às SCBS mais apropriadas [Ioannou and Weber 2016]. Dessa forma, espera-se que, através de agrupamentos, os *caches* consigam alcançar melhores taxas de acerto e descarregar a MBS [ElBamby et al. 2014, Pantisano et al. 2015].

Apesar do agrupamento apresentar ganhos em relação ao descarregamento de dados, a maioria dos trabalhos relatados na literatura [ElBamby et al. 2014, Pantisano et al. 2015] que recorrem a esta estratégia leva em consideração apenas a frequência dos conteúdos na definição dos grupos que serão associados às SCBS. O problema, contudo, é que a frequência dos conteúdos não relata de forma precisa o nível de afinidade entre os usuários, visto que, dois consumidores podem ter interesse apenas por um único conteúdo em comum e, ainda assim, serem considerados similares em decorrência da frequência de solicitações e da função de similaridade utilizada.

Este trabalho, portanto, apresenta uma estratégia que explora a formação de grupos de usuários similares em cenários de uso de *caches* em SCN. A técnica de agrupamento proposta se baseia nos interesses dos consumidores de conteúdos e visa obter melhores taxas de descarregamento de dados na MBS através do aumento da taxa de acerto nas SCBS. A estratégia proposta busca identificar grupos que retratem melhor o nível de afinidade entre os usuários considerando, não somente a frequência de solicitações, mas também a quantidade de conteúdos solicitados em comum. Os resultados obtidos demonstraram que quando a similaridade leva em consideração tanto a quantidade de conteúdos em comum quanto a frequência, as taxas de acerto e descarregamento são superiores a quando apenas um fator é utilizado. Além disso, a função de similaridade proposta pode ser adaptada a diferentes cenários, uma vez que a influência da frequência e do conteúdo pode ser ajustada a partir da atribuição de um peso às diferentes partes da equação.

O restante deste artigo está organizado da seguinte forma. A Seção 2 discute brevemente as principais propostas que fazem uso de técnicas de agrupamento em cenários de SCN. A Seção 3 apresenta a estratégia adotada. A Seção 4 descreve o ambiente de avaliação adotado e a Seção 5 analisa os principais resultados experimentais obtidos. Por fim, a Seção 6 conclui o artigo e discute possíveis trabalhos futuros.

2. Gerenciamento de *Cache* em Redes Orientadas à Conteúdo

As Redes Orientadas à Conteúdo [Jacobson et al. 2009] tem como característica a desvinculação entre um conteúdo e seu localizador, com o objetivo de focar em *qual* conteúdo o usuário quer, em vez de *onde* o conteúdo está localizado. Essa abordagem

apresenta novas estruturas que dão suporte ao tráfego de conteúdo nomeado: a *Pending Interest Table* (PIT) armazena as requisições que ainda não foram atendidas, a *Forwarding Information Base* (FIB) mantém as informações de encaminhamento e a *Content Store* é uma estrutura de armazenamento similar a um *cache*. A utilização do *cache* nos próprios equipamentos de rede é um dos diferenciais dessa abordagem e essa característica tem sido utilizada em diferentes contextos.

Existem diversas políticas de gerenciamento de *cache* [Ioannou and Weber 2016] com diversas adaptações para se adequar aos diferentes cenários, no entanto, essas estratégias geralmente não levam em consideração as preferências individuais dos usuários. Desta forma, começaram a ser criadas estratégias que fizessem agrupamentos entre usuários com características similares para que os *caches* conseguissem alcançar melhores taxas de acerto e descarregamento.

Em [ElBamby et al. 2014] é descrita uma solução que faz uma associação entre usuários e *caches* de SCBS. A solução visa melhorar o descarregamento de dados na MBS e diminuir o tempo de resposta ao usuário. A associação entre usuário e SCBS é feita de acordo com a similaridade entre eles, que é calculada utilizando a lei do cosseno em relação à frequência dos conteúdos requisitados. Após a associação, os *caches* das SCBS são atualizados de acordo com um método de aprendizado por regressão também proposto no artigo. A estratégia conseguiu minimizar o atraso em 42% e 27% e obteve um ganho de 280% e 90% quando comparado com o agrupamento feito de forma aleatória e com esquemas sem agrupamentos.

[Pantisano et al. 2015] propõe um algoritmo de associação entre usuários e *caches* com a intenção de fazer uma melhor alocação da largura de banda, aumentar QoS do usuário e reduzir retransmissões. O cálculo de similaridade se baseia em técnicas de Filtro Colaborativo e é realizado levando em consideração a frequência da média de requisições recebidas por cada SCBS, da média de requisições para cada conteúdo e da média de requisições de cada usuário para todos os conteúdos. A intenção é explorar as relações entre as requisições de conteúdo e os usuários, levando em consideração a probabilidade de um conteúdo ser requisitado por outro usuário. Os resultados demonstraram ganhos em relação à alocação de largura de banda e taxa de acerto com melhoras de até 25%.

3. Estratégia de Agrupamento por Interesses Similares

A solução desenvolvida visa aumentar a taxa de acerto nas SCBS para diminuir a sobrecarga na MBS. Para tanto, a solução conta com o uso de agrupamentos baseados na similaridade de usuários em relação aos seus interesses. Nesta seção serão descritos a estrutura da rede e os tipos de pacotes trafegados, a função de similaridade, o algoritmo de agrupamento e as formas de comunicação existentes.

3.1. Estrutura da Rede e Tipos de Pacotes

O cenário é composto por três entidades distintas:

- a MBS é a estação principal, capaz de prover todos os conteúdos que estão disponíveis na rede;
- a SCBS tem capacidade de armazenamento limitada e é utilizada para armazenar réplicas de conteúdo mais próximas dos usuários;

- os dispositivos móveis (DM) são os clientes/assinantes de conteúdo que fazem requisições através do envio de pacotes de interesse.

A comunicação entre os equipamentos é feita através do envio de pacotes. Os pacotes de interesse tem a finalidade de requisitar conteúdos nomeados, os pacotes de dados também são nomeados e comportam o conteúdo propriamente dito, e os pacotes de controle são pacotes de dados cujo conteúdo é utilizado na manutenção da rede.

Cada entidade é responsável por manipular esses pacotes de acordo com a sua funcionalidade. Os dispositivos móveis são responsáveis pelas requisições de dados através do envio de pacotes de interesse. Cada dispositivo contém uma tabela de requisições que é composta pelo nome do conteúdo e a frequência em que foi solicitado. Desta forma, toda vez que um pacote de interesse é enviado, a tabela de requisições é atualizada. Essa tabela é enviada à MBS através de um pacote de controle e é utilizada para realização do cálculo de similaridade e agrupamento entre os dispositivos.

As principais funções da MBS são as de processar pacotes de interesse e controle. Ao receber um pacote de interesse, a MBS faz uma busca no *cache* para encontrar o pacote de dados referente àquele interesse e enviá-lo ao nó solicitante. A MBS é capaz de prover todos os conteúdos disponíveis na rede. Além disso, é responsável por realizar o cálculo de similaridade e a formação dos agrupamentos, conforme a Seção 3.2.

A SCBS pode receber e solicitar pacotes de interesse e dados. Ao receber um pacote de interesse, verifica no *cache* se existe um pacote de dados relacionado àquele conteúdo, se existir, encaminha o pacote de dados para o nó solicitante. Caso contrário, o nome do conteúdo e o nó solicitante são adicionados à PIT. Em seguida, a SCBS solicita o conteúdo à MBS e, ao recebê-lo, adiciona no seu *cache* e envia para o nó requisitante. Vale ressaltar que a ROC faz agregação de requisições, portanto, se mais de um nó solicitar o mesmo conteúdo, apenas uma entrada na PIT é adicionada com o nome do conteúdo e a lista de endereços dos nós solicitantes. Desta forma, o pacote de interesse é enviado uma única vez através da rede e o pacote de dados é replicado para todos os nós assinantes.

De acordo com as diferentes funcionalidades de cada entidade, é possível observar que o envio de pacotes de dados pode ocorrer de três maneiras distintas, como pode ser observado na Figura 1:

- a MBS envia o conteúdo diretamente para o DM;
- a SCBS tem o conteúdo armazenado em *cache* e envia para o DM;
- a SCBS não tem o conteúdo armazenado e a MBS envia o conteúdo para ela.

Cada uma dessas situações ocorre de acordo com a configuração da rede em determinado instante. Antes da formação dos grupos, todos os pacotes de dados são enviados pela MBS. Após a formação dos agrupamentos, os usuários são servidos pelas SCBS. A MBS só enviará conteúdos para a SCBS quando a SCBS não tiver o conteúdo solicitado armazenado em *cache*. Em uma ROC, um pacote de dados só é enviado em resposta a um pacote de interesse, logo, toda vez que a SCBS necessitar de um conteúdo que não existe no seu *cache*, deverá enviar uma requisição através do envio de um pacote de interesse.

Como foi citado anteriormente, além dos pacotes de interesse e de dados, existem os pacotes de controle. Os pacotes de controle são utilizados para realizar a formação dos agrupamentos. Os primeiros pacotes de controle enviados contém as tabelas de

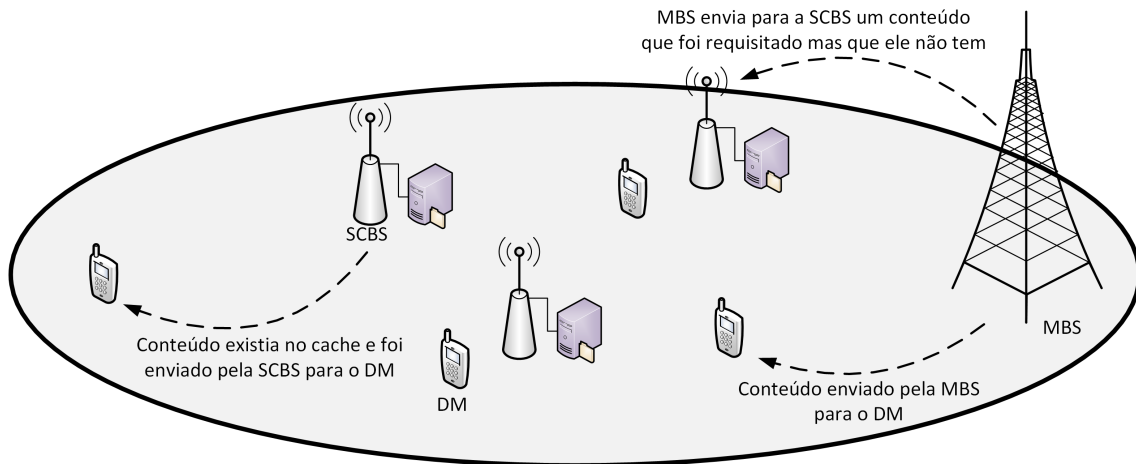


Figura 1. Estrutura da SCN utilizada.

requisições dos usuários e são utilizados como insumo para o cálculo de similaridade e formação dos agrupamentos.

3.2. Função de Similaridade

A função de similaridade tem o objetivo de quantificar a semelhança entre diferentes usuários de acordo com seus interesses.

A similaridade entre os usuários é medida através da Equação 1. A primeira parte representa o Índice de Jaccard [Jaccard 1912], comumente utilizado para medir a similaridade através das características semelhantes e distintas entre dois indivíduos. Ela foi utilizada para medir a similaridade de acordo com os conteúdos em comum que dois nós (i, j) tem interesse. Já a segunda parte é calculada utilizando a lei de similaridade do cosseno em relação à frequência de requisições de conteúdos em comum.

Desta forma, C_i e C_j representam os conjuntos de conteúdos que foram solicitados por i e por j , respectivamente, e F_i e F_j representam a frequência, em i e em j , em que os conteúdos comuns foram requisitados. O β define um peso entre as partes da equação e permite que seja avaliado o impacto delas no cálculo de similaridade.

$$S_{i,j} = \beta \left(\frac{|C_i \cap C_j|}{|C_i \cup C_j|} \right) + (1 - \beta) \left(\frac{F_i \cdot F_j}{\|F_i\| \|F_j\|} \right) \quad (1)$$

Vale ressaltar, que não estamos avaliando a similaridade entre usuário e *cache*, nem entre dois conteúdos, mas a similaridade entre usuários de acordo com seus interesses por conteúdo. Além disso, neste trabalho, o cenário “com similaridade” representa o uso da função de similaridade na formação dos agrupamentos. Enquanto no cenário “sem similaridade” os membros são adicionados de acordo com sua identificação. Por exemplo, os nós de 1 a 10 formam o primeiro agrupamento, os nós de 10 a 20 formam o segundo, e assim por diante. Isso não quer dizer que não exista similaridade entre os usuários, mas que a similaridade não está sendo levada em consideração na formação dos grupos.

3.3. Algoritmo de Agrupamento

O algoritmo de agrupamento (Algoritmo 1) é executado na MBS e tem a função de estabelecer as associações entre nós que tenham interesses similares. Esse trabalho avalia o efeito de apenas uma execução do algoritmo.

Os agrupamentos são determinados utilizando as tabelas de requisições dos nós como insumo. A partir de um determinado instante de tempo t , cada usuário envia uma única vez um pacote de controle contendo a sua tabela de requisições. Cada tabela de requisição recebida pela MBS é adicionada a uma matriz $N \times C$, sendo N a quantidade de nós da rede e C a quantidade de conteúdos disponíveis. Ao receber a tabela de requisição, a MBS verifica o ID do nó que enviou a requisição e adiciona na linha da matriz referente àquele ID. Depois de receber as requisições de todos os nós da rede, é inicializado o cálculo da similaridade. Para saber quando todas as requisições foram recebidas, foi utilizado um contador inicializado em zero e incrementado até a quantidade de nós.

Algoritmo 1: Algoritmo de Agrupamento

Entrada: Matriz $N \times C$

```
1 início
2   para cada  $i \in N$  faça
3     para cada  $j \in N$  faça
4       para cada  $w \in C$  faça
5         Calcule os valores de  $C_i \cap C_j$ ,  $C_i \cup C_j$ ,  $F_i \cdot F_j$  e  $\|F_i\| \|F_j\|$ ;
6       fim
7     Calcule a similaridade entre  $i$  e  $j$ ;
8     Construa uma matriz  $N \times N$  com a similaridade dos nós;
9   fim
10  fim
11  para cada  $z \in s$  faça
12    Aplique o método de ordenação por seleção em cada linha da matriz;
13    Calcule a média aritmética entre os  $k$  vizinhos mais similares de cada
        nó;
14    Identifique o nó com maior média de similaridade;
15    Forme um agrupamento com o nó com maior média e seus  $k$  vizinhos
        mais similares;
16    Envie mensagem de controle para os nós do agrupamento formado
        identificando o servidor de distribuição que eles devem se associar;
17    Zere os valores de similaridade dos nós que já foram agrupados;
18  fim
19 fim
```

Saída: s agrupamentos, sendo s a quantidade de SCBS

O cálculo da similaridade é realizado par a par entre cada nó da rede, gerando uma matriz $N \times N$. O índice de cada linha da matriz corresponde ao ID de um dos nós que compõe a rede e a linha representa a similaridade entre um nó e os demais. A linha 1, por exemplo, equivale à similaridade entre o nó 1 e todos os outros nós da rede $\{1, 2, 3, \dots, N\}$. Além disso, a diagonal principal dessa matriz é uma diagonal nula, pois representa a similaridade entre um nó e ele mesmo.

Após a formação da matriz contendo a similaridade entre todos os nós da rede, é feita uma ordenação decrescente entre os elementos de cada linha da matriz, para que os nós sejam ordenados de acordo com a similaridade. Cada elemento da matriz é uma *struct* formada pelo ID do nó e pela similaridade, isso garante que, mesmo após a ordenação, sejam mantidos os IDs dos nós que mudaram de posição.

Com a matriz de similaridade ordenada, o próximo passo é calcular a média de similaridade entre um nó e os seus k vizinhos mais próximos (do inglês, k *Nearest Neighbor*), obtendo assim a média de similaridade de cada nó da rede. Em seguida, o nó com maior média é identificado e inicia-se a formação do agrupamento.

O agrupamento é formado pelo nó que tem a maior média e os seus k primeiros vizinhos, que representam os vizinhos com maior similaridade. O valor de k é calculado por $(N/S)-1$. Sendo que N é a quantidade de nós, S é a quantidade de SCBS e a subtração por 1 é justificada pelo fato de que o primeiro nó do agrupamento já foi determinado (nó com maior média de similaridade). Esse cálculo garante que os nós sejam distribuídos de maneira uniforme entre as SCBS. A distribuição uniforme foi feita com o intuito de equilibrar a quantidade de usuários por SCBS, para que a quantidade de usuários por antenna não interferisse nos resultados do efeito da equação de similaridade. Por exemplo, supõe-se que uma antenna com dois usuários conectados teria uma taxa de acerto maior do que uma antenna com 20 nós conectados, ainda que fossem bastante similares.

Ao identificar os membros do agrupamento, a MBS envia um pacote de controle para cada usuário identificando a SCBS que ele deve se associar. A partir desse momento, os DM fazem requisições apenas para as SCBS, deixando de requisitar à MBS.

No Algoritmo 1 a complexidade do cálculo de similaridade é de $O(CN^2)$ enquanto a realização do agrupamento tem complexidade $O(N^3)$. Sendo C a quantidade de conteúdos e N a quantidade de usuários.

3.4. Formas de Comunicação

Após a formação dos agrupamentos, existem duas possibilidades de comunicação. Partindo do pressuposto de que um DM envia um conteúdo nomeado para a SCBS1 e o único nó da rede que possui o pacote de dados associado a esse conteúdo é a MBS, é possível que a comunicação ocorra de duas maneiras:

- Comunicação Direta, conforme Figura 2: ao receber um pacote de interesse de um dispositivo móvel (1), SCBS1 verifica a sua *cache* (2) e percebe que não existe um conteúdo relacionado ao interesse. Desta forma, é adicionada uma entrada na PIT (3) e o pacote de interesse é encaminhado diretamente a MBS (4). A MBS acessa o *cache* para obter o conteúdo solicitado (5,6) e envia o pacote de dados para SCBS1 (7), SCBS1 armazena o pacote de dados no seu *cache* (8), apaga a entrada da PIT (9) e encaminha o pacote de dados para o dispositivo móvel solicitante (10);
- Comunicação Indireta, conforme Figura 3: o DM envia um pacote de interesse para SCBS1 (1), como SCBS1 não tem o conteúdo referente ao interesse, ela encaminha o interesse à SCBS2 (2) e a SCBS2 realiza o mesmo processo de verificação e encaminhamento (3). Ao chegar na SCBS3, o pacote de interesse já percorreu todos os nós da rede daquele nível, portanto, a única entidade que pode resolver

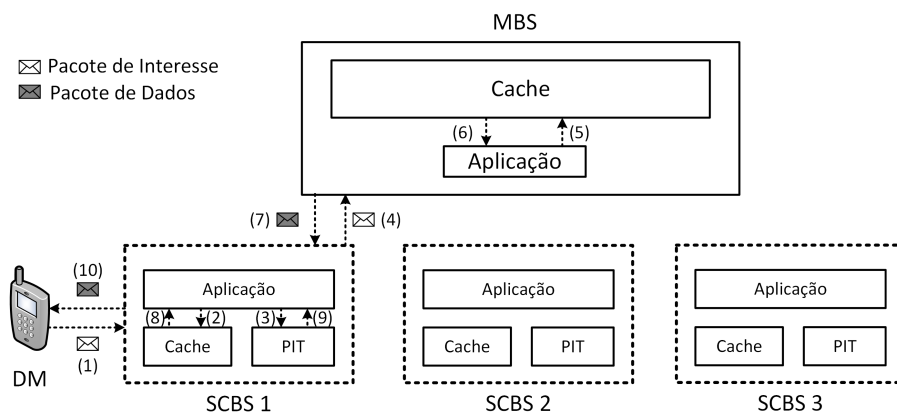


Figura 2. Comunicação Indireta.

o interesse é a MBS. Logo, o servidor SCBS3 envia o pacote de interesse para a MBS (4) e a MBS envia o pacote de dados diretamente para SCBS1 (5). Isso é possível porque mesmo o pacote sendo encaminhado entre as SCBS, o endereço da primeira SCBS solicitante é mantido nas mensagens. Por fim, a SCBS1 encaminha o pacote de dados para o DM (6).

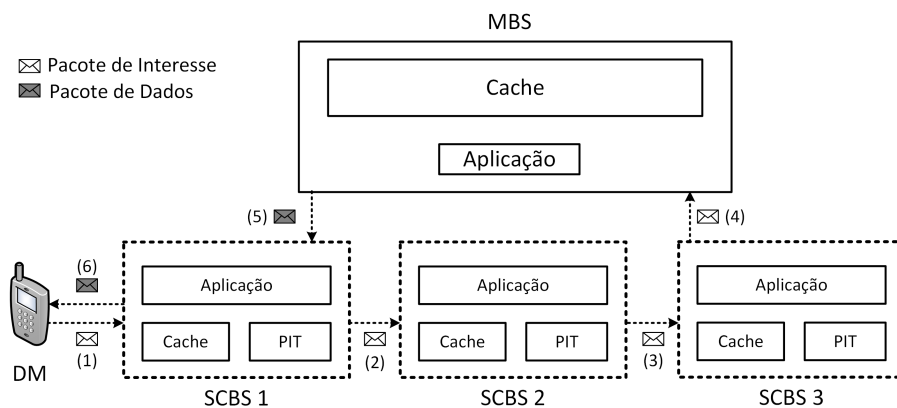


Figura 3. Comunicação Direta.

4. Experimentos

Para realizar os experimentos, foi necessário determinar cenários de avaliação, métricas, fatores e parâmetros. Todos esses aspectos serão abordados nessa seção.

4.1. Cenários, Métricas, Fatores e Parâmetros

Os fatores utilizados podem ser vistos na primeira coluna da Tabela 1. A similaridade representa a utilização da equação de similaridade e o β representa o peso usado na equação. A comunicação retrata a direção no encaminhamento dos pacotes, o tamanho do *cache* indica espaço disponível para armazenamento, e o α representa o parâmetro da Distribuição de Zipf, que é responsável por definir a popularidade dos conteúdos.

De acordo com esses fatores foram definidos três cenários de avaliação: o Grupo 1 compara o impacto do parâmetro β e do Tamanho da *Cache* em relação à taxa de acerto;

Tabela 1. Fatores e Cenários de Avaliação.

Fatores	Grupo 1	Grupo 2	Grupo 3
Similaridade	Sim - Não	Sim - Não	Sim - Não
Comunicação	Direta	Direta	Indireta - Direta
Tamanho do <i>Cache</i>	10 - 20 - 30	20	20
β (Funç. Similaridade)	0 - 0.5 - 1	0 - 0.5 - 1	0.5
α (Distr. Zipf)	1	0.6 - 0.8 - 1	1

o Grupo 2 avalia o impacto do parâmetro Zipf e do parâmetro β na taxa de acerto; e, por fim, o Grupo 3 analisa a utilização da similaridade e da hierarquia em relação à taxa de acerto e ao descarregamento de dados.

Os experimentos foram avaliados de acordo com duas métricas: taxa de acerto na SCBS e descarregamento de dados na MBS. O descarregamento de dados visa medir a quantidade de requisições que deixaram de ser feitas na MBS e pode ser representado pela Equação 2, na qual $rMax$ representa os interesses que foram requisitados quando todas as requisições eram feitas diretamente na MBS e $rMin$ são as requisições recebidas após a formação dos agrupamentos.

$$off = 100 \times \frac{rMax - rMin}{rMax} \quad (2)$$

Já a taxa de acerto mede a quantidade de vezes que uma SCBS recebeu um pacote de interesse e tinha o pacote de dados no *cache*. Essa taxa é medida de acordo com a Equação 3, na qual $match$ representa os interesses que foram resolvidos ao chegar na SCBS e $iRec$ o total de todos os interesses que a SCBS recebeu. A taxa de acerto total é calculada através da média aritmética da taxa de acerto das três SCBS.

$$hr = 100 \times \frac{match}{iRec} \quad (3)$$

4.2. Ambiente de Simulação

A solução proposta foi implementada no simulador OMNeT++ utilizando o *framework* INET, conforme parâmetros da Tabela 2.

A área simulada é de 600x800m onde estão distribuídos 30 dispositivos móveis que se movimentam de acordo com o modelo de mobilidade *Mass Mobility* e requisitam conteúdo seguindo o modelo de Poisson. O meio físico sem fio tem propagação do tipo *Constant Speed Propagation*, frequência de 2.4GHz, perda do tipo *Free Space Path Loss* e ruído de -110dBm, todos esses parâmetros já vieram especificados no INET. Foi criada uma MBS que alcança toda área de simulação e foram criadas três SCBS cujo alcance é de 300m. Na simulação, cada SCBS é associada a 10 dispositivos móveis, de forma imperativa, de acordo com a estratégia utilizada. Existem 100 conteúdos disponíveis na rede e, para determinar perfis diferentes para os usuários, foi selecionado um conjunto de conteúdo para cada dispositivo. Para cada subconjunto de conteúdo é aplicado o modelo de popularidade de *Zipf*. O tamanho do *cache* da SCBS varia entre 10% e 30% da

Tabela 2. Parâmetros da Simulação.

Parâmetro	Valor
Quantidade de dispositivos móveis na rede	30
Quantidade de MBS	1
Quantidade de SCBS	3
Quantidade de conteúdos disponíveis	100
Política de descarte de <i>cache</i>	LFU
Modelo de Mobilidade	<i>MassMobility</i>
Tipo de Propagação	<i>ConstantSpeedPropagation</i>
Frequência	2.4GHz
Perda	<i>FreeSpacePathLoss</i>
Ruído	-110dBm

quantidade de conteúdos e a estratégia de descarte utilizada é a LFU. O parâmetro β , que determina o peso na equação de similaridade, é variado entre 0 e 1 e são comparados cenários com comunicação direta e indireta. Cada experimento foi executado 10 vezes com um Intervalo de Confiança de 95%.

5. Análise dos Resultados

Nesta seção são apresentados os resultados e as análises dos diferentes cenários em relação à taxa de acerto e de descarregamento de dados de acordo com as variações de fatores apresentadas na Tabela 1.

5.1. Grupo 1: Relação entre Tamanho do *Cache*, β e Taxa de Acerto

A Figura 4 expõe o comportamento da taxa de acerto em relação ao uso da função de similaridade e da variação do tamanho do *cache*. Em relação ao uso da equação de similaridade, para todos os tamanhos de *cache* (10, 20, 30), é possível observar que quando apenas a quantidade de conteúdos em comum é levada em consideração ($\beta = 1$), a taxa de acerto é superior a quando apenas a frequência é utilizada ($\beta = 0$). Isso pode ser justificado devido ao fato de que ao contabilizar apenas a frequência, é possível que dois usuários sejam considerados bastante similares mesmo tendo um único conteúdo em comum. Por exemplo, quando dois nós tem um conteúdo em comum, ainda que requisitado uma única vez, de acordo com a função da similaridade do cosseno, esses dois nós terão similaridade igual a 1. Dessa forma, como o Índice de Jaccard leva em consideração a relação entre os conteúdos requisitados em comum e todos os conteúdos requisitados pelos dois nós, os agrupamentos baseados nessa relação de similaridade conseguem alcançar resultados superiores. Apesar disso, os melhores resultados foram obtidos quando ambos ($\beta = 0.5$) foram utilizados, evidenciando a importância de uma equação de similaridade que considere esses dois fatores.

Vale ressaltar que ao levar em consideração apenas a frequência de conteúdos ($\beta = 0$) calculada pela similaridade do cosseno, não estamos replicando o cenário descrito em [ElBamby et al. 2014], visto que o autor calcula similaridade usuário-*cache* e neste trabalho é calculada a similaridade usuário-usuário.

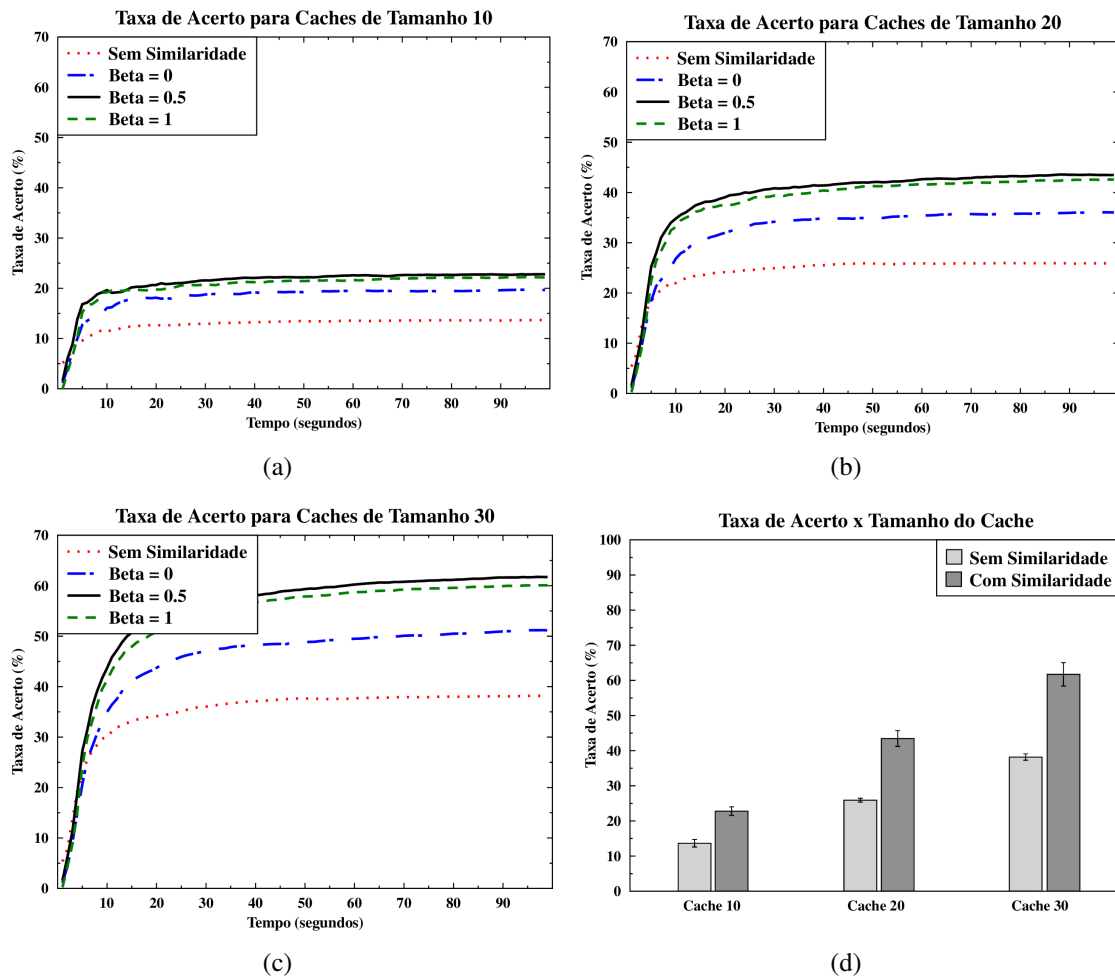


Figura 4. Relação entre o Tamanho do *Cache* e a Taxa de Acerto.

A distinção entre as taxas de acerto dos gráficos das Figuras 4(a) 4(b) 4(c) se dá devido ao aumento do tamanho do *cache*. No entanto, ao relacionar os tamanhos do *cache* 10 e 20, conforme a Figura 4(d), observa-se que a taxa de acerto obtida sem similaridade com um *cache* = 20 alcança valores próximos ao melhor valor obtido com uso de similaridade em um *cache* = 10. Ao se comparar os *caches* de tamanho 20 e 30, percebe-se que com o uso da similaridade em um *cache* = 20 é possível obter valores superiores ao cenário sem similaridade em um *cache* = 30. Isso evidencia que a utilização de agrupamentos levando em consideração a similaridade dos usuários torna, de fato, o gerenciamento de *cache* mais eficiente e possibilita economia de recursos, visto que os mesmos valores de taxa de acerto podem ser obtidos com *caches* de tamanho inferior.

5.2. Grupo 2: Relação entre o parâmetro α da Lei de Zipf e a Taxa de Acerto

A Figura 5(a) evidencia que mesmo com diferentes valores para o parâmetro Zipf, o comportamento da função de similaridade permanece o mesmo: os resultados com valor de $\beta = 0.5$ ou $\beta = 1$ são superiores a quando considera-se apenas a frequência. No entanto, é possível observar na Figura 5(b) que, apesar do comportamento da função de similaridade permanecer o mesmo, os valores de taxa de acerto alcançados são distintos. Quanto menor o valor do parâmetro Zipf, maior a taxa de acerto obtida. Isso acontece porque

quando o parâmetro Zipf é menor, há uma maior quantidade de conteúdos populares que, conseqüentemente, são requisitados mais vezes. Como a política de descarte utilizada é a LFU, os conteúdos mais populares ficam armazenados por mais tempo, satisfazendo as requisições e culminando no aumento da taxa de acerto.

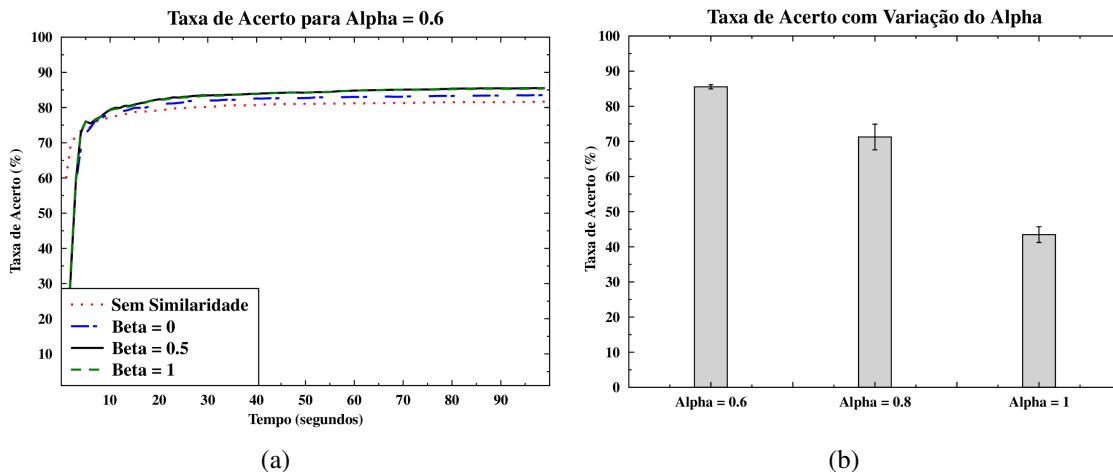


Figura 5. Efeito do parâmetro Zipf na Taxa de Acerto.

5.3. Grupo 3: Relação entre Similaridade e Comunicação na Taxa de Acerto e no Descarregamento de Dados

Na Figura 6(a) é possível observar o comportamento entre similaridade e comunicação em relação à taxa de descarregamento. A taxa de descarregamento de aproximadamente 22% obtida no cenário sem similaridade e com comunicação direta, decorre da adição das três SCBS. O cenário com comunicação indireta e com similaridade consegue obter resultados superiores ao cenário com comunicação direta e sem similaridade, evidenciando que a utilização de agrupamentos influencia mais no descarregamento do que a forma de comunicação. Os melhores resultados foram obtidos com comunicação indireta e similaridade.

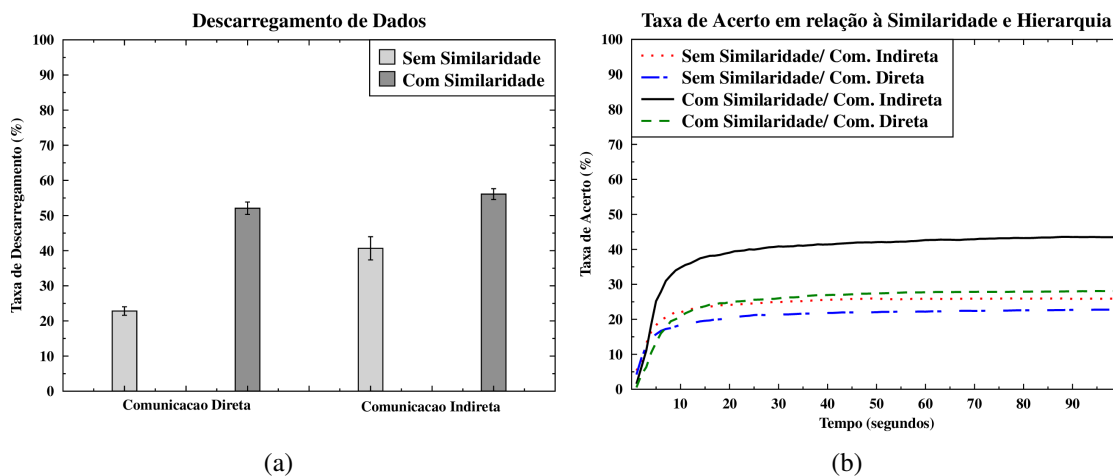


Figura 6. Influência da Similaridade e da Hierarquia na Taxa de Acerto e de Descarregamento de Dados.

Em relação a avaliação da taxa de acerto nos cenários com e sem similaridade e comunicação direta e indireta, é possível observar, de acordo com a Figura 6(b), que os cenários de comunicação direta obtêm taxa de acerto superior aos cenários de comunicação indireta. Isso acontece porque quando a comunicação é indireta, as SBCS recebem pacotes de interesse que não pertencem aos usuários do grupo que está associado a elas, portanto, a probabilidade delas terem o conteúdo armazenado em *cache* é menor. Apesar disso, o grupo de usuários associado a cada SCBS não sofre nenhum impacto decorrente dessa diminuição da taxa de acerto, visto que não há substituição em *cache* dos conteúdos que foram solicitados por outra SCBS. No entanto, há um aumento no tempo de resposta ao usuário decorrente da comunicação indireta, visto que pode aumentar a quantidade de saltos para alcançar o conteúdo.

Na Figura 7 foi feito um experimento com apenas 30 conteúdos, no qual todos os usuários requisitam os conteúdos disponíveis com diferentes frequências, logo, apenas a parte da equação que representa a similaridade do cosseno irá refletir a similaridade dos usuários. Assim, é possível observar que, contabilizando apenas os conteúdos em comum ($\beta = 1$), a taxa de acerto é inferior a quando a frequência é considerada ($\beta = 0$). Isso acontece porque como todos os usuários requisitam todos os conteúdos, o Índice de Jaccard não consegue detectar a similaridade entre os usuários, pois os conjuntos de conteúdo serão sempre iguais. Desta forma, ao utilizar $\beta = 1$, os resultados são semelhantes a um cenário sem similaridade.

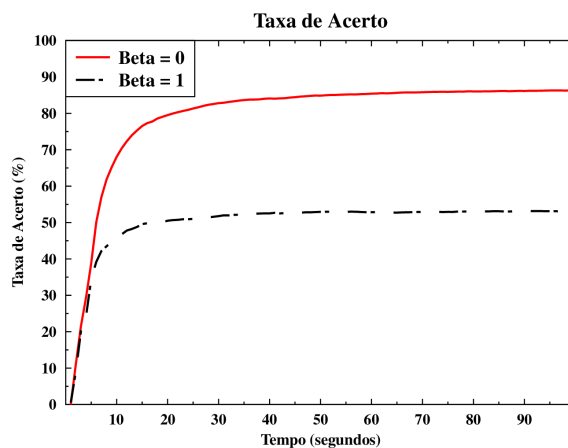


Figura 7. Representação da adaptatividade da função de similaridade.

Também já foi exposto que quando dois nós tem um único conteúdo em comum solicitado uma única vez, de acordo com a similaridade do cosseno, esses nós são caracterizados com grau de similaridade máximo. Dessa forma, fica evidente a importância da utilização de ambas as partes da função de similaridade (Equação 1) durante a formação dos agrupamentos, tornando o cálculo de similaridade adaptativo às características do ambiente e capaz de superar lacunas de cada indicador de similaridade.

6. Conclusão e Trabalhos Futuros

Os resultados obtidos demonstraram que, ao utilizar a função de similaridade proposta, foi possível obter valores de taxa de acerto superiores ao cenário aleatório mesmo com tamanho de *cache* inferior, evidenciando a economia de recursos e aumento da taxa de acerto.

Ao utilizar a comunicação indireta há um aumento no tempo de resposta ao usuário e degradação da taxa de acerto nas SCBS. Apesar disso, os melhores resultados para descarregamento de dados na MBS envolvem a comunicação indireta e o uso de similaridade.

Como trabalhos futuros, pretende-se aumentar a quantidade de conteúdos e de nós, também levando em consideração os nós que não se conectam a todas as SCBS. Além disso, será avaliada a entrada e saída de nós no ambiente para verificar a necessidade de recalcular a similaridade entre os usuários e a distribuição dos nós nas SCBS não será realizada uniformemente.

Agradecimentos

Os autores agradecem o apoio da CAPES e do CNPq.

Referências

- Aijaz, A., Aghvami, H., and Amani, M. (2013). A Survey on Mobile Data Offloading: Technical and Business Perspectives. *IEEE Wireless Communications*, 20(2):104–112.
- Bhushan, N., Li, J., Malladi, D., Gilmore, R., Brenner, D., Damnjanovic, A., Sukhavasi, R. T., Patel, C., and Geirhofer, S. (2014). Network Densification: The Dominant Theme for Wireless Evolution into 5G. *IEEE Communications Magazine*, 52(2):82–89.
- Chang, Z., Gu, Y., Han, Z., Chen, X., and Ristaniemi, T. (2016). Context-Aware Data Caching for 5G Heterogeneous Small Cells Networks. In *Communications (ICC), 2016 IEEE International Conference on*, pages 1–6. IEEE.
- Chen, Z. and Kountouris, M. (2015). Cache-enabled Small Cell Networks with Local User Interest Correlation. In *IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 680–684.
- ElBamby, M. S., Bennis, M., Saad, W., and Latva-Aho, M. (2014). Content-aware User Clustering and Caching in Wireless Small Cell Networks. In *11th International Symposium on Wireless Communications Systems (ISWCS)*, pages 945–949.
- Hajri, S. E. and Assaad, M. (2016). Caching Improvement Using Adaptive User Clustering. *arXiv preprint arXiv:1605.09602*.
- Ioannou, A. and Weber, S. (2016). A Survey of Caching Policies and Forwarding Mechanisms in Information-Centric Networking. *IEEE Communications Surveys & Tutorials*, pages 1553–877.
- Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytologist*, 11(2):37–50.
- Jacobson, V., Smetters, D. K., Thornton, J. D., Plass, M. F., Briggs, N. H., and Braynard, R. L. (2009). Networking Named Content. In *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*, pages 1–12.
- Pantisano, F., Bennis, M., Saad, W., and Debbah, M. (2015). Match to Cache: Joint User Association and Backhaul Allocation in Cache-aware Small Cell Networks. In *IEEE International Conference on Communications (ICC)*, pages 3082–3087.