

Um Algoritmo de Clusterização Espectral para Agrupamento Inteligente em Redes sem Fio Densas

Bruna T. Guedes¹, Diego G. Passos^{1,2}, Fernanda G. O. Passos^{2,3}

¹Instituto de Computação - IC, Universidade Federal Fluminense – Brasil

²ISEL, Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa – Portugal

³Atlântica - Instituto Universidade – Portugal

brunatoledo@midia.com.uff.br, dpassos@ic.uff.br, fernanda@midia.com.uff.br

Abstract. *The Restricted Access Window (RAW) was introduced in IEEE 802.11ah to improve the performance of Dense Wireless Networks (DWN). RAW restricts the number of stations that can access the channel by separating them into groups. This work proposes two clustering strategies for forming RAW groups particularly targeting the reduction of hidden terminals within groups, increasing the performance of DWN. The first strategy uses K-Means assuming the availability of the geographical positions of the stations, while the second relaxes that assumption by using spectral clustering on Received Signal Strength Indicator (RSSI) information. Extensive simulations indicate that the proposed strategies significantly outperform the standard grouping used in IEEE 802.11ah and the Hidden Matrix-based Regrouping (HMR) in terms of collision rate, throughput and delay.*

Resumo. *O mecanismo de Restricted Access Window (RAW) foi introduzido pelo IEEE 802.11ah para melhorar o desempenho das Redes Sem Fio Densas (ou Dense Wireless Networks, DWN). O RAW restringe o número de estações que podem acessar o canal, separando-as em grupos. Este trabalho propôs duas estratégias de agrupamento para formação dos grupos RAW, com o objetivo principal de diminuir os terminais ocultos dentro de cada grupo, aumentando o desempenho da DWN. A primeira estratégia utiliza o K-Means assumindo a disponibilidade da posição geográfica das estações, enquanto a segunda relaxa esta hipótese e utiliza o agrupamento espectral sobre a informação de Received Signal Strength Indicator (RSSI). Simulações extensivas indicam que as estratégias propostas superam significativamente o agrupamento padrão do IEEE 802.11ah e o algoritmo Hidden Matrix-based Regrouping (HMR) em termos de taxa de colisão, vazão e atraso.*

1. Introdução

Nos últimos anos, houve uma expansão do uso de dispositivos móveis com comunicação sem fio. As redes sem fio se tornaram uma alternativa amplamente utilizada, substituindo, em muitos cenários, as redes convencionais com fio [Cisco 2020]. Esse aumento da demanda por comunicação sem fio, aliado à escassez de recursos espectrais, trouxe uma série de desafios ao desempenho destas redes.

Particularmente, para lidar com a alta probabilidade de colisões em redes muito densas, a emenda IEEE 802.11ah introduziu o mecanismo *Restricted Access Window*

(RAW) [Raeesi et al. 2014], que divide as estações em grupos, de acordo com seus AIDs (*Association IDentification*), e atribui intervalos de tempo a cada grupo: apenas as estações de um grupo têm permissão para acessar o canal durante o intervalo de tempo de seu grupo [Tian et al. 2016]. Embora esse agrupamento reduza os níveis de contenção, ele não considera os terminais ocultos, permitindo que eles coexistam em um mesmo grupo, o que impacta negativamente o desempenho.

Esta dissertação mostra como o algoritmo K-Means pode ser usado para agrupar as estações geograficamente mais próximas, diminuindo a probabilidade de terminais ocultos serem colocados no mesmo grupo e melhorando o desempenho de uma DWN. Embora os resultados das simulações demonstrem que o método é promissor, na prática geralmente é difícil obter e manter a localização geográfica das estações da rede, tornando sua implantação muitas vezes inviável. Por este motivo, propõe-se também uma segunda abordagem baseada em informações mais facilmente disponíveis para as estações da rede, particularmente o *Received Signal Strength Indicator* (RSSI). Este valor pode ser medido em quadros recebidos na maioria das NICs (*Network Interface Card*) sem fio típicas e tende a se correlacionar com a distância entre as estações. O uso do RSSI para gerar os grupos RAW, no entanto, não é trivial. Diferentemente das posições geográficas, que são uma propriedade da estação, o RSSI é um atributo de um enlace ou par de estações, impedindo o uso do algoritmo K-means tradicional. Para contornar este problema, foi empregada uma técnica de agrupamento espectral, seguida de um pós-processamento que busca equalizar o tamanho dos grupos.

Ambas as propostas desta dissertação foram avaliadas extensivamente através de simulações de redes Wi-Fi densas. Nestas simulações, as propostas foram comparadas tanto ao algoritmo de agrupamento padronizado do IEEE 802.11ah, quanto a uma das principais propostas da literatura, o *Hidden Matrix-based Regrouping* (HMR) [Yoon et al. 2016]. Os resultados mostram que os algoritmos propostos neste trabalho têm desempenho significativamente superior ao agrupamento padronizado e ao HMR, independentemente do número de nós, da quantidade de grupos RAW utilizados ou da densidade da rede.

Em suma, este trabalho contribui para o estado da arte apresentando uma nova estratégia prática para agrupamento RAW que não necessita das posições geográficas das estações e com o objetivo de diminuir a quantidade de terminais ocultos. A extensiva avaliação realizada neste trabalho também revelou uma série de características importantes do uso do mecanismo RAW, como a importância da geração de grupos balanceados.

2. Trabalhos Relacionados

Em DWNs, entender e melhorar o desempenho do protocolo CSMA/CA (*Carrier Sense Multiple Access with Collision Avoidance*) é um tópico importante e bem estudado. Bianchi *et al.*, por exemplo, apresentam um modelo de desempenho do CSMA/CA e mostram que as colisões aumentam rapidamente com o crescimento do número de estações concorrentes — mesmo na ausência de terminais ocultos — reduzindo assim o desempenho da rede em termos de vazão [Bianchi et al. 1996]. Em [Yoon et al. 2016], os autores propõem um algoritmo para diminuir o número de terminais ocultos em uma rede. O algoritmo, chamado de *Hidden Matrix-based Regrouping*, ou simplesmente HMR, move as estações entre grupos para diminuir o número de terminais ocultos a partir do agrupa-

mento gerado pelo RAW do padrão IEEE 802.11ah. Primeiro, um grupo é selecionado e uma lista de terminais ocultos neste grupo é criada. A estação que possui mais terminais ocultos é selecionada para ser movida para outro grupo, desde que não seja um terminal oculto de nenhuma estação desse grupo — caso contrário, outro grupo de destino será escolhido. Se todos os grupos foram considerados alvos potenciais e em todos eles a estação selecionada possui terminais ocultos, a estação permanece no grupo atual e é removida da lista. Este processo é repetido para todas as estações da lista de terminais ocultos do grupo atual e, a seguir, o mesmo procedimento é aplicado a todos os grupos restantes. Nota-se que o HMR é um pré-processamento heurístico que já parte de um agrupamento potencialmente ruim, podendo eventualmente não encontrar boas soluções, conforme ilustrado na Seção 4.

3. Novas estratégias de agrupamento RAW baseadas no K-Means

A primeira estratégia de agrupamento RAW proposta nesta dissertação foi utilizar o algoritmo de agrupamento K-Means para agrupar as estações evitando terminais ocultos em um mesmo grupo. Embora haja outros algoritmos na literatura, como o agrupamento hierárquico, o K-Means obtém melhores resultados com grandes conjuntos de dados [Abbas 2008]. O agrupamento é baseado nas posições geográficas das estações. Assim, espera-se reunir estações próximas em um mesmo grupo, enquanto as mais distantes ficam em grupos diferentes. A consequência é uma redução na probabilidade de existirem terminais ocultos em um mesmo grupo e, portanto, um aumento no desempenho da DWN.

No algoritmo K-Means, os dados de entrada são o número k de *clusters* desejados e o conjunto de estações com suas respectivas posições geográficas. O algoritmo começa definindo as posições dos centroides iniciais dos *clusters* como sendo a posição geográfica de estações aleatoriamente selecionadas. Em seguida, é calculada a distância das estações para cada centroide. A estação será atribuída para o *cluster* correspondente ao centroide mais próximo. Depois disso, os centroides de cada *cluster* são atualizados de acordo com as posições das estações que o compõem. O processo é repetido até que não haja mais mudança na posição dos centroides.

3.1. Agrupamento espectral e motivação

Na prática, é difícil de obter e manter a localização geográfica de estações associadas a um ponto de acesso. Mesmo que as posições iniciais sejam conhecidas, se uma estação se move, o padrão IEEE 802.11 não prevê qualquer tipo de notificação ao ponto de acesso. Assim, a implementação do agrupamento baseado no algoritmo K-means como descrito na seção anterior não é prática, a não ser em cenários muito particulares.

Uma alternativa é a técnica de agrupamento espectral [Spielman 2012], que é capaz de criar o agrupamento de estações desejado a partir de informações como o *Received Signal Strength Indication* (RSSI). A partir do RSSI, é possível construir um grafo inter-relacionando as estações, de forma que duas estações sejam mais fortemente conectadas no grafo à medida que o RSSI entre elas seja mais alto. Como o RSSI está, geralmente, relacionado à distância — tende a ser maior para estações que estão mais próximas e mais baixo para estações mais distantes —, o comportamento esperado é que a saída do agrupamento espectral calculada sobre a informação RSSI seja semelhante à do algoritmo

K-Means sobre as posições geográficas. Por outro lado, em cenários mais complexos e realistas de propagação, em que obstáculos podem ter grande influência na potência do sinal recebido, o RSSI pode ser até mais determinante que a distância geográfica nas relações entre terminais ocultos.

Um fluxograma da estratégia baseada no agrupamento espectral é mostrado na Figura 1. A entrada é um grafo constituído pelas estações da rede (vértices) e os enlaces (arestas) que as conectam. Cada aresta tem um peso associado representando o RSSI do respectivo enlace. Na primeira etapa, é executado o algoritmo de agrupamento espectral, denominado K-Means Espectral. Ele é composto pela decomposição espectral e pela execução do algoritmo K-Means. Na decomposição espectral, o grafo é representado por uma matriz a partir da qual são calculados autovalores e autovetores. Um subconjunto dos autovetores são selecionados e utilizados para derivar coordenadas equivalentes das estações para a execução do K-Means. Sobre a saída do K-Means, é executada uma função de pós-processamento que faz um o balanceamento das estações entre os *clusters*. Como resultado, são obtidos os grupos formados de acordo com a informação do RSSI. Estas etapas são detalhadas a seguir.

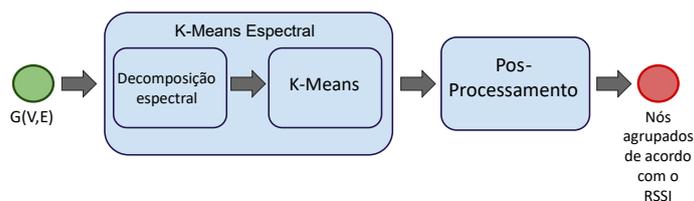


Figura 1. Fluxo da estratégia baseada no agrupamento espectral.

3.1.1. K-Means Espectral

A primeira etapa é a decomposição espectral que começa com a construção da matriz de adjacência A do grafo: as linhas e colunas correspondem aos vértices, enquanto o valor de cada entrada representa a existência — ou a falta — da aresta correspondente [Cao et al. 2015]. Em seguida, é calculada a matriz de graus D , que é uma matriz diagonal cujas entradas denotam, no caso de um grafo ponderado, a soma dos pesos das arestas que esse nó possui [Spielman 2012]. Um conceito central do agrupamento espectral é a matriz laplaciana, que captura relações de similaridade do grafo original [Chin et al. 2010]. Matematicamente, a matriz laplaciana L é definida como:

$$L = D - A. \quad (1)$$

O próximo passo é determinar autovetores e autovalores de L , pois eles capturam várias propriedades do grafo original [Chin et al. 2010]. Uma das propriedades, por exemplo, é a conectividade do grafo que é definida pelo segundo menor autovalor de L : se esse autovalor for positivo, o grafo estará conectado [Lucińska and Wierzchoń 2018, Rittenhouse 2008]. Os autovetores calculados a partir da matriz laplaciana podem ser organizados em uma nova matriz, onde as linhas representam pontos em um espaço vetorial — que, por sua vez, representam as estações do grafo original — e as colunas formadas pelos autovetores representam as coordenadas das estações [Hefeda et al. 2012].

Após ser feita a decomposição espectral, o algoritmo K-Means é executado para o agrupamento das estações. O conjunto de dados de entrada para o K-Means é um subconjunto da matriz de autovetores da matriz laplaciana. Para obter esse subconjunto, apenas os $k - 1$ autovetores associados aos $k - 1$ menores autovalores não nulos são selecionados.

3.1.2. Função de pós-processamento

Ao final do processamento do K-Means, os grupos nem sempre serão balanceados — *i.e.*, com o mesmo número de estações em cada grupo. Porém, para o mecanismo RAW, é importante que os grupos sejam balanceados, pois agrupamentos altamente desbalanceados levam a grandes grupos com alta competição pelo uso do meio e, portanto, mais colisões [Guedes et al. 2019]. Por este motivo, nossa proposta inclui uma função de pós-processamento onde o balanceamento é realizado movendo certas estações de *clusters* maiores — referidos como *doadores* — para os menores — referidos como *receptores*. No entanto, essa tarefa não é trivial, pois não se deve mover um nó para um determinado *cluster* se outras estações neste *cluster* estiverem ocultas a ele — pois isso também levaria a mais colisões. Assim, em algumas situações não será possível ter um equilíbrio perfeito devido à possibilidade de criação de terminais ocultos.

Para evitar a inclusão de terminais ocultos nos grupos durante o pós-processamento, um nó só é transferido se não estiver oculto a nenhum outro nó do *cluster* receptor, ou seja se todas as arestas entre o nó doador e as estações do *cluster* receptor possuírem peso acima do *CCAThres* (*Clear Channel Assessment Threshold*). Caso contrário, outro nó neste grupo doador é selecionado e a mesma verificação é executada novamente. Esse procedimento é repetido até que um nó do *cluster* doador seja eleito para migrar para o *cluster* receptor ou não haja mais estações para analisar no *cluster* doador. As migrações tentam equilibrar os *clusters* para que atinjam o número ideal de estações ou a melhor distribuição possível, evitando a criação de terminais ocultos. No entanto, a ordem de realização das trocas pode fazer diferença nessa distribuição, e nem sempre será possível obter o melhor equilíbrio.

4. Resultados de simulações

Para avaliação das estratégias de agrupamento propostas — denominadas K-Means Espectral e K-Means Original —, foram realizadas simulações comparativas ao método de agrupamento padrão usado no IEEE 802.11ah — denominado Arbitrário — e ao algoritmo HMR. Também foram feitas simulações sem utilizar nenhum tipo de agrupamento que proveem um *baseline* de desempenho. Para simplificar a integração das simulações de redes com os diversos algoritmos de agrupamento avaliados, um simulador foi implementado, na linguagem Python, para realização dos testes. Esta seção resume alguns dos principais resultados documentados na dissertação.

A Figura 2 mostra as taxas de colisão em função do número de estações associadas ao ponto de acesso para quantidades de grupos $k \in \{5, 10, 15\}$. As barras de erro denotam intervalos de confiança de 95%. Os agrupamentos baseados no K-Means Original e K-Means Espectral tiveram um desempenho melhor do que o agrupamento Arbitrário, apresentando menores taxas de colisão. Essa diferença é acentuada em cenários

com menos estações ou mais grupos. Para um exemplo extremo, com 100 estações e 15 grupos, o agrupamento Arbitrário resultou em uma taxa de colisão de 67%, enquanto o K-Means Original resultou em 7% e o K-Means Espectral resultou em 9%. À medida que o número de estações aumenta, o número de grupos disponíveis torna-se insuficiente para lidar de forma satisfatória com o nível de competição, independentemente da estratégia de agrupamento. Como consequência, a diferença de desempenho é reduzida.

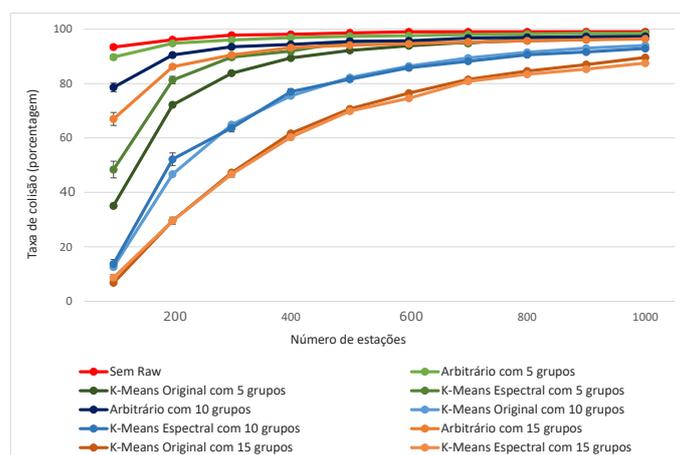


Figura 2. Taxa de colisão versus número de estações.

Também é importante destacar que o K-Means Espectral teve um desempenho muito semelhante ao K-Means Original para a maioria dos cenários avaliados. Diferenças de desempenho significativas entre as duas abordagens só podem ser percebidas quando apenas cinco grupos estavam disponíveis, o que torna mais difícil evitar que terminais ocultos sejam colocados nos mesmos grupos. Isso é digno de nota porque o K-Means Original tem à sua disposição as posições geográficas precisas das estações, uma informação que normalmente não está disponível.

Na Seção 2 foi descrito o algoritmo HMR [Yoon et al. 2016], uma solução da literatura para diminuir os terminais ocultos com o objetivo de melhorar o desempenho da rede. Neste algoritmo, a informação do quadro PS-POLL é usada para verificar se as estações são terminais ocultos entre si. As estações enviam esse quadro PS-POLL para o ponto de acesso no momento em que elas estão preparadas para receber os dados [Arnosti 2020]. Para comparar o desempenho dos algoritmos K-Means Original e espectral com o algoritmo HMR, em relação às métricas de colisão, atraso e vazão, foi implementada uma versão do algoritmo HMR considerando a mesma forma de identificação dos terminais ocultos da nossa proposta, na qual foi usado como critério o limiar de detecção de portadora.

A Figura 3 mostra a vazão agregada em função da carga oferecida da rede. As simulações foram feitas com 400 estações e 15 grupos. Para a maioria dos casos, o algoritmo K-Means Espectral atingiu uma vazão maior em relação ao algoritmo HMR devido à diminuição dos terminais ocultos – já que mais pacotes são entregues com sucesso, aumentando a vazão. Conforme o número de estações aumenta, a eficiência do agrupamento feito pelo HMR piora, comparado ao algoritmo K-Means Espectral. À medida que a quantidade de estações aumenta, a quantidade de terminais ocultos nos grupos também aumenta quando o algoritmo HMR foi executado. O K-Means Espectral conseguiu for-

mar grupos com uma quantidade menor de terminais ocultos em relação ao HMR devido ao agrupamento ser feito utilizando o RSSI.

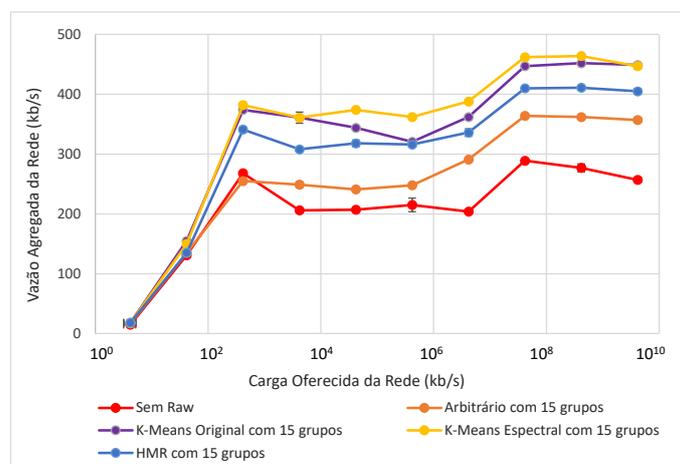


Figura 3. Vazão agregada da rede versus carga oferecida da rede. Comparação entre o desempenho do HMR e K-Means Espectral nas simulações com 400 estações.

5. Publicações

Esta dissertação de mestrado resultou na publicação de dois artigos científicos:

1. Guedes, B. T., Passos, F. G., and Passos, D. (2023). A spectral clustering algorithm for intelligent grouping in Dense Wireless Networks. *Computer Communications*, 198:117-127.
2. Guedes, B. T., Passos, F. G., and Passos, D. (2019). The role of artificial intelligence in current and future dense wireless networks. In *2019 1st Blockchain, Robotics and AI for Networking Security Conference (BRAINS)*, pages 1–8. IEEE.

6. Conclusão

Esta dissertação propôs e avaliou duas estratégias de agrupamento RAW baseadas em algoritmos de agrupamento. A primeira, baseada no algoritmo K-means, tem desempenho muito expressivo em relação ao agrupamento padrão do IEEE 802.11ah e ao estado da arte da área, mas necessita de informações de posicionamento geográfico das estações, que nem sempre podem ser obtidas. A segunda, baseada em agrupamento espectral utilizando o RSSI, tem desempenho em geral ligeiramente inferior à primeira — embora ainda significativamente superior ao estado da arte —, mas pode ser implementada facilmente com base nas NICs Wi-Fi existentes. Em resumo, a estratégia de agrupamento espectral proposta neste trabalho tem implementação viável na prática e é capaz de melhorar significativamente o desempenho de redes sem fio densas.

Disponibilidade da Dissertação

A dissertação pode ser acessada em <http://www.ic.uff.br/PosGraduacao/frontend-tesesdissertacoes/download.php?id=1045.pdf&tipo=trabalho>

Referências

- Abbas, O. A. (2008). Comparisons between data clustering algorithms. *International Arab Journal of Information Technology (IAJIT)*, 5(3).
- Arnosti, S. (2020). Estratégias para agrupamento de nós e alocação dinâmica de recursos em redes IEEE 802.11ah. Master's thesis, Universidade Estadual de Campinas, Campinas, SP.
- Bianchi, G., Fratta, L., and Oliveri, M. (1996). Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 wireless LANs. In *Proceedings of PIMRC'96-7th International Symposium on Personal, Indoor, and Mobile Communications*, volume 2, pages 392–396. IEEE.
- Cao, S., Lu, W., and Xu, Q. (2015). Grarep: Learning graph representations with global structural information. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 891–900.
- Chin, C.-F., Shih, A. C.-C., and Fan, K.-C. (2010). A novel spectral clustering method based on pairwise distance matrix. *J. Inf. Sci. Eng.*, 26(2):649–658.
- Cisco, U. (2020). Cisco annual internet report (2018–2023) white paper. *Cisco: San Jose, CA, USA*.
- Guedes, B. T., Passos, F. G., and Passos, D. (2019). The role of artificial intelligence in current and future dense wireless networks. In *2019 1st Blockchain, Robotics and AI for Networking Security Conference (BRAINS)*, pages 1–8. IEEE.
- Hefeeda, M., Gao, F., and Abd-Almageed, W. (2012). Distributed approximate spectral clustering for large-scale datasets. In *Proceedings of the 21st international symposium on High-Performance Parallel and Distributed Computing*, pages 223–234.
- Lucińska, M. and Wierzchoń, S. T. (2018). Clustering based on eigenvectors of the adjacency matrix. *International Journal of Applied Mathematics and Computer Science*, 28(4).
- Raeesi, O., Pirskanen, J., Hazmi, A., Levanen, T., and Valkama, M. (2014). Performance evaluation of IEEE 802.11 ah and its restricted access window mechanism. In *2014 IEEE international conference on communications workshops (ICC)*, pages 460–466. IEEE.
- Rittenhouse, M. L. (2008). *Properties and Recent Applications in Spectral Graph Theory*. PhD thesis, Virginia Commonwealth University.
- Spielman, D. (2012). Spectral graph theory. *Combinatorial scientific computing*, 18.
- Tian, L., Famaey, J., and Latré, S. (2016). Evaluation of the IEEE 802.11 ah restricted access window mechanism for dense IoT networks. In *2016 IEEE 17th international symposium on a world of wireless, mobile and multimedia networks (WoWMoM)*, pages 1–9. IEEE.
- Yoon, S., J.Seo, and Bahk, S. (2016). Regrouping algorithm to alleviate the hidden node problem in 802.11 ah networks. *Computer Networks*, 105:22–32.