

Avaliação de Técnicas de Localização Indoor por Fingerprint de RSSI com Simulações no NS-3*

Marcelo Zamith¹, Marcel William Rocha da Silva¹

¹Universidade Federal Rural do Rio de Janeiro (UFRRJ)
Departamento de Ciência da Computação (DCC/IM) - Campus Nova Iguaçu

{mzamith,marcelsilva}@ufrrj.br

Abstract. *Indoor localization through RSSI fingerprinting has been a widely studied topic in recent years. A common point in most works is that the performance evaluation of the proposed techniques is carried out through practical experiments. Despite being a good way to evaluate performance and validate proposals in real environments, practical experiments limit the assessment of the impact that environmental characteristics have on the performance of localization techniques. In addition, the use of different scenarios makes it difficult to compare proposals. This work proposes a new way to evaluate the performance of RSSI fingerprint localization techniques using NS-3 simulations. With the NS-3 tools it is possible to create scenarios that represent indoor environments with different characteristics and to evaluate the impact of these characteristics on the localization performance.*

Resumo. *A localização indoor através de fingerprint de RSSI foi um tema amplamente estudado nos últimos anos. Um ponto comum na maioria dos trabalhos é que a avaliação de desempenho das técnicas propostas é realizada através de experimentos práticos. Apesar de ser uma boa forma de avaliar desempenho e validar as propostas em ambientes reais, os experimentos práticos limitam a avaliação do impacto que as características do ambiente têm sobre o desempenho das técnicas de localização. Além disso, o uso de cenários distintos dificulta a comparabilidade entre propostas. Este trabalho propõe uma nova forma de avaliar desempenho de técnicas de localização por fingerprint de RSSI usando simulações com o NS-3. Com as ferramentas do NS-3 é possível criar cenários que representam ambientes indoor com características distintas e avaliar o impacto dessas características no desempenho da localização.*

1. Introdução

Nas últimas duas décadas, a sociedade vem presenciando um aumento significativo da quantidade de dispositivos móveis e portáteis, o que fez crescer a quantidade de dispositivos que se comunicam através de transmissões sem fio. Neste contexto, surgiram novas aplicações pervasivas, como as desenvolvidas para ambientes inteligentes e para a Internet das Coisas, que dependem de informação de contexto do ambiente para o seu funcionamento. Um tipo de informação muito usada por essas novas aplicações é a localização geográfica dos dispositivos, que viabiliza desde aplicações de navegação até aplicações de entretenimento e segurança [Basiri et al. 2017].

*Este trabalho recebeu recursos financeiros da FAPERJ.

Geralmente, o GPS é a forma mais usada por dispositivos móveis para determinar sua localização geográfica. Usando um receptor para sinais transmitidos por satélites geoestacionários, um dispositivo móvel pode calcular a distância para os satélites visíveis no céu e estimar, com uma boa precisão, a sua localização na superfície terrestre. Apesar de ser uma técnica amplamente utilizada, a localização através de GPS possui limitações. A existência de barreiras físicas entre o receptor GPS e os satélites prejudica na recepção dos sinais e pode tornar imprecisa, ou até mesmo impossível, a estimativa da localização [Sadowski and Spachos 2018]. Esta situação é mais comum em ambientes internos (*indoor*), onde o teto e as paredes funcionam como barreiras para os sinais GPS.

Uma alternativa para a localização *indoor* são as técnicas baseadas nos níveis de sinais de redes sem fio [Zafari et al. 2019], em especial a técnica de *fingerprinting*. Esta técnica consiste em utilizar os níveis de potência dos sinais recebidos (*Received Signal Strength Indication* - RSSI) a partir de transmissões de equipamentos fixos (âncoras) como uma “impressão digital” (*fingerprint*) daquele local. No caso de redes Wi-Fi, os âncoras são os pontos de acesso (Access Points - APs) presentes no ambiente. Para cada local do ambiente em que uma estação Wi-Fi estiver localizada, haverá um padrão formado pelos RSSIs de cada AP. Desta forma, o problema da localização *indoor*, pode ser solucionado como um problema de reconhecimento de padrões. Usando técnicas de aprendizado de máquina, um classificador pode ser treinado para identificar o local mais provável de onde um *fingerprint* (FP) foi coletado.

Por ser uma técnica amplamente estudada na literatura, muitas variações do modelo clássico de sistema de localização baseado em *fingerprint* descrito anteriormente já foram propostas. Diversos trabalhos já avaliaram o uso de diferentes algoritmos de aprendizado de máquina [Bahl and Padmanabhan 2000, Haeberlen et al. 2004, Brunato and Battiti 2005, Jedari et al. 2015], novas formas de gerar datasets de treinamento [Jiang et al. 2012] e novos formatos de FP [Ferreira et al. 2020, Jiang et al. 2012]. De acordo com nosso levantamento bibliográfico, um ponto em comum entre os trabalhos presentes na literatura é que a grande maioria realiza a avaliação de desempenho das técnicas de localização por *fingerprint* através de experimentos práticos ou utilizando *datasets* de *fingerprints* previamente coletados em ambientes reais. Apesar do experimento prático possuir suas vantagens, como a comprovação de que a técnica funciona bem em um ambiente real, os experimentos práticos também limitam a avaliação de desempenho às características daquele único cenário utilizado para coletar os FPs. Desta forma, o impacto de algumas características do ambiente de experimentação, que podem influenciar no desempenho da localização, não são avaliados. Por exemplo, seria interessante saber como a quantidade e tamanho dos cômodos do ambiente influenciam na técnica de localização. Também seria interessante avaliar o impacto da densidade de pontos de referência (PRs) e de APs presentes no ambiente. Essas características quase sempre são constantes nos experimentos apresentados.

Além do problema citado anteriormente, há também uma dificuldade de comparar resultados de diferentes trabalhos. Cada artigo utiliza um *dataset* coletado em um ambiente diferente, com características diferentes e nem sempre o *dataset* de FPs é disponibilizado. Esses artigos muitas das vezes costumam enumerar o erro da localização da técnica proposta, mas esse erro pode ser diretamente dependente do cenário em que a técnica foi avaliada.

Este trabalho tem como objetivo propor uma nova forma de avaliação de técnicas de localização *indoor*, utilizando simulações. Através de ferramentas já disponíveis na última versão do simulador NS-3 [NS-3 2022b], apresentamos uma nova técnica para a obtenção de *datasets* de FP através da: criação de cenários *indoor*, geração de pacotes, e a medição de RSSI. Esta técnica fornece um arcabouço para a avaliação do impacto que características do ambiente têm no desempenho da localização *indoor*. Tais características do cenário, que geralmente são mantidas constantes em experimentos práticos, podem ser facilmente modificadas através de simulações.

Para avaliar a metodologia proposta, este trabalho apresenta também um estudo de caso para avaliar o impacto de características como densidade de PRs, quantidade de APs e tamanho/quantidade de cômodos no desempenho da localização por *fingerprint* de RSSI com o algoritmo kNN. Os resultados mostram que essas características do ambiente, que geralmente são negligenciadas nas avaliações de desempenho feitas na literatura, de fato causam impacto na precisão da localização por *fingerprint*.

O restante do trabalho está organizado da seguinte maneira. A Seção 2 explica com mais detalhes a técnica de localização por *fingerprint* e apresenta trabalhos relacionados. A Seção 3 explica o funcionamento do modelo de propagação *indoor* com cômodos do NS-3. Na Seção 4 apresenta-se a metodologia proposta para a criação dos cenários de experimentação de localização por *fingerprint*. Na Seção 5 apresentamos o estudo de caso e os resultados obtidos. Por fim, a Seção 6 apresenta as conclusões e trabalhos futuros.

2. Localização por *Fingerprint* de RSSI e Trabalhos Relacionados

Para entender como funcionam os sistemas de localização por *fingerprint* de RSSI, primeiro será explicado o que é um *fingerprint*. A Figura 1 mostra um cenário com três APs, AP_1 , AP_2 e AP_3 , e duas estações, A e B. Para cada estação do exemplo pode-se visualizar o *fingerprint* (FP) formado pelo RSSI com que cada estação recebe sinais oriundos dos APs, que é um vetor com o seguinte formato $FP = \{RSSI_{AP_1}, RSSI_{AP_2}, RSSI_{AP_3}\}$.

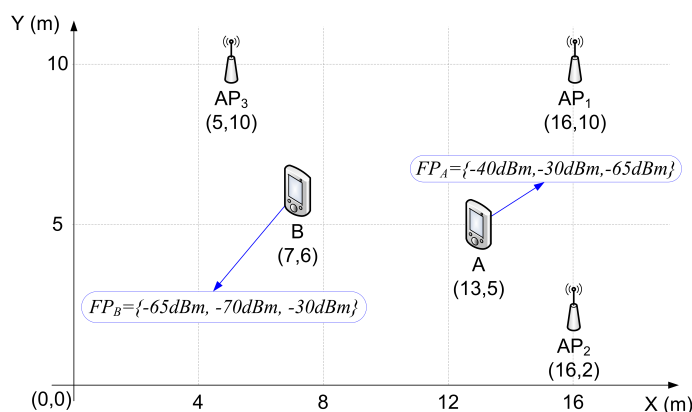


Figura 1. Exemplo de *fingerprint* de duas estações em ambiente com três APs.

De acordo com as características do ambiente, cada local do ambiente terá seu FP próprio. Desta forma, o problema da localização torna-se um problema de classificação ou reconhecimento de padrões. Para isso, os sistemas de localização baseados em *fingerprint* costumam ser constituídos em duas etapas. A primeira etapa é *offline* e consiste

em mapear pontos de referência (PRs) no ambiente através de suas coordenadas e coletar amostras de FP nesses pontos. As amostras coletadas são registradas em um *dataset* de treinamento, onde os nomes dos PRs em que os FPs foram coletados servem de rótulo. Esse *dataset* é então usado para treinar um modelo de aprendizado de máquina, que funcionará como um classificador para as novas amostras de FP. Na segunda etapa, *online*, um dispositivo que deseja se localizar pode coletar um FP do local em que está localizado e utilizar o modelo de aprendizado de máquina previamente treinado para classificar essa nova amostra. O resultado da classificação indica um PR, e por consequência, a localização mais provável do dispositivo que coletou aquela amostra de FP.

Conforme mencionado na Seção 1, diversos trabalhos já abordaram o problema da localização por *fingerprint* de RSSI. Um dos primeiros trabalhos sobre o tema é [Bahl and Padmanabhan 2000], que propôs a técnica de *fingerprint* para a localização dos dispositivos Wi-Fi através do algoritmo kNN. Posteriormente, vários outros trabalhos visaram explorar novos algoritmos de classificação, como por exemplo a inferência bayesiana [Haeberlen et al. 2004], máquina de vetores de suporte (*Support Vector Machine - SVM*) [Brunato and Battiti 2005], e *random forest* [Jedari et al. 2015].

Outros trabalhos relacionados ao tema exploram novos formatos para o vetor de FP. Em [Ferreira et al. 2020], propõe-se o uso de quartis das amostras de RSSI para compor o vetor de FPs ao invés do uso da média do RSSI. Para avaliar o desempenho da proposta o trabalho realizou simulações e experimentos reais. Entretanto, em ambos os casos foi avaliado apenas um cenário simples, de apenas uma sala com pouco mais de $12 m^2$.

Essa restrição a respeito do cenário utilizado nas avaliações de desempenho da localização baseada em *fingerprints* é bem comum nos trabalhos presentes na literatura. [Singh et al. 2021] apresenta uma vasta revisão bibliográfica sobre o tema, onde sintetiza em uma única tabela informações sobre 31 trabalhos publicados entre os anos de 2011 e 2021. Uma das informações que constam na tabela diz respeito a área utilizada nos experimentos para a avaliação de desempenho. Neste ponto há uma grande variação, com trabalhos avaliando suas propostas em apenas uma sala com algumas dezenas de metros quadrados, e outros em áreas que consistem em prédios inteiros com múltiplos andares e milhares de metros quadrados.

[Carvalho 2021] também apresenta um levantamento de trabalhos que propõem técnicas de localização por *fingerprint* de RSSI, mas restrito apenas aos trabalhos que utilizam kNN e suas variantes. Neste trabalho, além da área de experimentação, o levantamento bibliográfico listou a quantidade de APs, e a quantidade e a densidade de PRs. Aqui novamente pode-se perceber uma variação grande nos cenários, com a quantidade de APs variando de poucas unidades até centenas de APs e a densidade de PRs variando de um PR a cada $100 m^2$ até mais de um PR por m^2 .

Esses levantamentos mostram que as técnicas propostas na literatura para o problema da localização por *fingerprint* de RSSI estão sendo avaliadas em experimentos realizados em cenários muito distintos. Eventualmente uma técnica que seja mais eficiente, por exemplo, em um cenário com uma densidade de PRs alta poderia apresentar um desempenho ruim num ambiente de experimentação com uma densidade baixa de PRs. A avaliação do impacto dessas características do cenário no desempenho das técnicas acaba não sendo

avaliada pelos autores através de experimentos práticos, uma vez que seria necessário um esforço grande para criar ambientes de experimentação em cenários de características variadas e que podem nem estar disponíveis para os autores. Outra possibilidade seria utilizar os *datasets* coletados em outros trabalhos que utilizam ambientes de experimentação diversos para avaliar as técnicas. Entretanto, nem sempre os *datasets* criados para os trabalhos estão disponíveis publicamente. [Singh et al. 2021] indica em seu levantamento que apenas 8 dos 31 trabalhos citados disponibilizam ou usam *datasets* públicos. Sendo assim, este trabalho apresenta uma nova metodologia para avaliar técnicas de localização por *fingerprint* através de simulações que representam características de ambientes reais usando o NS-3.

3. Modelo de Propagação *Indoor* do NS-3

O NS-3 [NS-3 2022b] é um simulador de redes por eventos discretos amplamente utilizado em trabalhos científicos. O simulador é um software modular escrito em C++ que permite criar simulações de redes através da instanciação de nós (nodes) e canais de comunicação. Os nós são formados pela interconexão de diversos objetos que implementam as diversas camadas da pilha de protocolos de comunicação e uma aplicação para a geração de tráfego. Já os canais são objetos que interconectam as interfaces de camada física dos nós. Para simulações de nós sem fio há um módulo de mobilidade que permite definir a localização geográfica dos nós e fazer com que os nós se movimentem durante a simulação. Há também a possibilidade de escolher diferentes modelos de propagação para ser usados pelo objeto canal para o cálculo da atenuação dos sinais transmitidos de acordo com a localização dos nós e das características do canal e do cenário. A Figura 2 mostra um exemplo dos nós de uma simulação sem fio interconectados por um objeto canal.

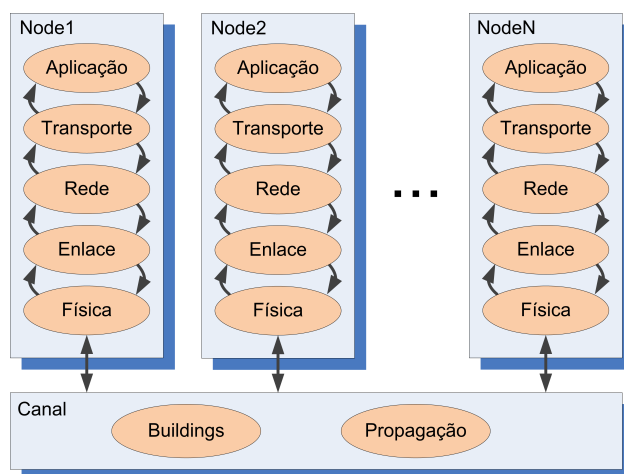


Figura 2. Abstração de uma simulação no NS-3: nós interconectados por um canal

Um dos módulos do NS-3 usados nas simulações desse artigo é o módulo *Buildings* [NS-3 2022a]. Este módulo implementa a possibilidade de criar construções (prédios) no cenário de simulação, os quais são representados por caixas de dimensões configuráveis. Para cada prédio, é possível especificar a quantidade de andares e também a quantidade de salas de cada andar. A representação disponibilizada pelo

módulo é simplificada. As salas presentes em um andar do prédio são sempre iguais e dispostas no formato de uma grade regular $n \times m$.

Além da possibilidade de criar prédios, o módulo `Buildings` também disponibiliza um modelo de propagação específico para este cenário, denominado `HybridBuildingsPropagationLossModel`. Este modelo na realidade é uma combinação de vários outros modelos de propagação existentes na literatura e alguns outros modelos de atenuação de sinais próprios para o ambiente com construções. O modelo de propagação a ser usado para o cálculo da atenuação dos sinais é escolhido de acordo com a condição em que o par TX-RX se encontram. Os modelos usados são Okumura-Hata [Rappaport et al. 1996], ITU-R P.1411 [ITU 2021b] e ITU-R P.1238 [ITU 2021a].

Além do modelo de propagação determinado pela situação, o modelo de propagação do módulo `Buildings` inclui modelos específicos para alguns fenômenos típicos de um cenário com prédios. A perda por penetração na parede externa dos prédios (*External Wall Loss* - EWL) é representada por um valor fixo de atenuação em dB toda vez que ocorre uma comunicação *indoor-outdoor* e vice e versa. Há também a representação da perda por penetração nas paredes internas do prédio (*Internal Wall Loss* - IWL). Essa perda é modelada como um valor fixo de atenuação em dB multiplicado pela quantidade de paredes existentes entre o par TX-RX. Para estimar a quantidade de paredes, como os cenários *indoor* são formados por uma grade regular de salas, utiliza-se a distância de *manhattan* entre TX e RX dada em números de salas. Além desses valores há também um modelo de ganho por altura (*Height Gain* - HG), que acrescenta 2 dB de ganho toda vez que o transmissor está em um andar acima do primeiro. Por fim, em todos os casos, o modelo adiciona também um valor aleatório em dB ao sinal para contabilizar pelo fenômeno de *shadowing*. Esse valor aleatório é sorteado de uma distribuição log-normal apenas uma vez para cada par TX-RX em cada simulação.

Os cenários de simulação criados para este trabalho utilizam o módulo `Buildings` para a criação de cenários *indoor*. Nesta situação, em que TX e RX encontram-se dentro de um mesmo andar de um prédio, o módulo escolhe sempre o modelo de propagação ITU-R P.1238, com *shadowing* log-normal e com as perdas por penetração nas paredes (IWL). O modelo ITU-R P.1238 possui ainda um parâmetro de configuração que é o tipo de ambiente *indoor*: residencial, escritório ou comercial.

Além do modelo de atenuação por percurso, adicionou-se às simulações um modelo de desvanecimento rápido (*fast fading*) Nakagami-m. Este modelo, que também está disponível no NS-3, inclui uma componente de variação momentânea dos sinais propagados, muito comum de ocorrer em ambientes *indoor*. Desta forma, a cada pacote transmitido no simulador, uma variação aleatória no nível do sinal é somada à perda no percurso já calculada pelo módulo `Buildings`. O modelo de Nakagami-m, possui um parâmetro de configuração, m , que influencia na magnitude da variação do sinal propagado.

4. Cenários de simulação

O uso da simulação para avaliar o desempenho das técnicas de localização por *fingerprint* consiste em gerar cenários formados por nós âncora, APs do padrão IEEE 802.11a, que transmitem pacotes periodicamente e posicionar outros nós, uma estação em cada ponto de referência (PR), para medir o nível de sinal dos pacotes recebidos com sucesso. Com isso, representa-se na simulação o mesmo que seria feito em um ambiente real para a

medição dos níveis de RSSI dos pacotes recebidos e a posterior composição de *fingerprints* (FPs) agrupando essas medidas de RSSI de cada PR.

O NS-3 permite o uso de diferentes agentes para a geração de pacotes num nível de camada de aplicação. Entretanto, para esta simulação optou-se por utilizar os pacotes de *beacon* gerados pela camada MAC do IEEE 802.11. Desta forma, cada AP da simulação foi instanciado contendo objetos da camada física até a camada MAC da pilha de protocolos, sem as camadas superiores. As estações instanciadas nos PRs para medir o nível de RSSI dos *beacons* transmitidos pelos APs foram configuradas de forma semelhante aos APs, mas com a geração de *beacons* desligada.

Para realizar a medição do nível de sinal nas estações posicionadas nos PRs, utilizou-se uma função de *callback* que era chamada a cada evento de recepção com sucesso pela camada física (*Phy/State/RxOk*). Este método era responsável por gravar em um arquivo de *trace* uma linha por pacote recebido, contendo: o *timestamp* do evento, endereço MAC do transmissor, endereço MAC do receptor e o RSSI do *beacon*. De posse dessas informações foi possível numa etapa posterior a simulação converter os RSSIs em FPs de cada PR para o treinamento e teste com um classificador.

O objetivo das simulações foi gerar cenários variados que viabilizassem a análise do impacto de características do cenário no desempenho de técnicas de localização. Usando o módulo *Buildings* foi possível criar cenários *indoor* variando a quantidade e tamanho das salas, a densidade de PRs e a quantidade de APs. Em todas as simulações os cenários consistiam no interior de um andar de um prédio de tamanho fixo de 20×20 metros, com altura de 3 metros. Neste ambiente variou-se a quantidade de salas internas através da escolha das dimensões n e m da grade regular. Para o mapeamento dos PRs, optou-se por uma grade regular de pontos de referência cobrindo de maneira uniforme e equidistante todo o ambiente. O tamanho dessa grade de PRs também foi variado.

A Figura 3 ilustra dois exemplos de cenários contendo as paredes que delimitam as salas, os PRs e os APs. A Figura 3(a) apresenta um esquema composto por 9 salas (3×3) e cada uma grade de PRs 9×9 , ao passo que a Figura 3(b) mostra um esquema com 25 salas (5×5) e uma grade de PRs 10×10 . É importante ressaltar que os PRs estão igualmente espaçados, independente das salas. Os APs foram posicionados de forma aleatória pelo andar, garantindo uma distância mínima de 1 metro entre APs. Tanto os APs quanto os nodes posicionados em cada PRs estavam a uma altura de 1 metro.

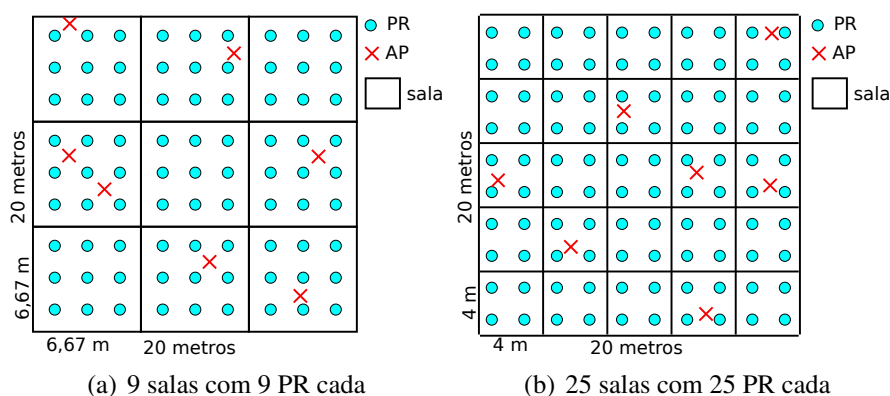


Figura 3. Esquema das salas com PRs - PR e seus APs

Em todas as simulações utilizou-se o modelo de propagação HybridBuildings combinado com o modelo Nakagami-m de desvanecimento rápido. Para o modelo HybridBuildings a constante de perda por penetração nas paredes (IWL) foi configurada em 5 dB, representando paredes de alvenaria finas ou divisórias [Rappaport et al. 1996], e o ambiente foi definido como residencial. Para o modelo Nakagami-m, utilizou-se o parâmetro $m = 1$.

5. Estudo de Caso e Resultados Obtidos

Com o intuito de demonstrar o potencial da técnica proposta, foi realizado um estudo de caso para a avaliação de desempenho de um sistema de localização *indoor* clássico, baseado em *fingerprint* de RSSI utilizando o RSSI médio e o algoritmo de classificação kNN. A análise realizada utilizou as técnicas apresentadas na Seção 4 para criar cenários de simulação diversos e coletar medidas do RSSI dos pacotes recebidos nos vários PRs, mas também em locais aleatórios. Em seguida, a partir dos RSSIs coletados utilizou-se um programa em linguagem Python para converter os RSSIs em amostras de FP rotuladas, o que permitiu treinar e testar um classificador kNN oriundo da biblioteca scikit-learn.

Os cenários foram criados nos moldes do que foi apresentado na Seção 4 e para esse estudo de caso, foram variados os seguintes parâmetros: (i) quantidade de APs, (ii) tamanho da *grid* de PRs, e (iii) tamanho da *grid* de salas. A Tabela 1 mostra um resumo dos valores utilizados para cada um desses parâmetros nas simulações.

Parâmetro	Valores usados na simulação
Quantidade de APs	3, 5 e 7
Grid de PRs	6×6 , 8×8 e 10×10
Grid de salas	3×3 , 4×4 e 5×5

Tabela 1. Valores variados nos parâmetros do cenário de simulação

Para cada combinação dos parâmetros da Tabela 1 foram gerados 60 cenários com o posicionamento aleatório dos APs. Cada um desses cenários deu origem a uma simulação onde cada AP transmitia uma quantidade de 1000 *beacons*. O RSSI de cada *beacon* foi medido pelos nodes posicionados nos PRs e registrado em arquivo. Na etapa seguinte de análise, usando o *script* em Python, as medidas de RSSI de cada PR foram agrupadas em janelas de tempo com duração necessária para a transmissão de 10 *beacons*. Desta forma, para cada janela de tempo, calculou-se o RSSI médio de cada AP, os quais posteriormente eram organizados em um vetor de RSSIs médios para dar origem ao FP daquela janela de tempo, conforme exemplo da Figura 1.

Os FPs então foram usados como entrada para treinar um classificador kNN usando o parâmetro $k = 1$. Logo, como resultado o classificador indicava apenas o rótulo mais provável para cada amostra de teste. Para o processo de teste utilizou-se o método do *holdout*, onde o *dataset* de FPs foi subdividido em 30% para teste e 70% para treino. Através dessa análise foi possível utilizar o modelo treinado para estimar a localização de cada uma das amostras do conjunto de teste.

A métrica de desempenho que será apresentada a seguir foi o erro médio absoluto. Para cada amostra de teste, foi calculada a distância entre a coordenada real da amostra e a coordenada do PR no qual a amostra foi classificada. Esse cálculo dá origem ao erro

de localização. Através do cálculo da média do erro das diversas amostras do conjunto de teste foi possível obter para cada um dos cenários o erro médio absoluto da localização. Os resultados apresentados a seguir são uma média das 60 variações de posicionamento aleatório dos APs, com um intervalo de confiança de 95%.

O primeiro conjunto de gráficos da Figura 4 mostram o erro médio da localização para cada combinação de parâmetros da Tabela 1. A primeira observação que pode ser feita sobre esses resultados é que o aumento da quantidade de APs reduz significativamente o erro médio da localização. Isso é um resultado esperado, uma vez que a quantidade de APs determina a quantidade de *features* presentes em cada FP. De maneira genérica, com n APs o vetor de FP terá n valores de RSSI, um por AP, e cada valor de RSSI presente no FP funciona como uma *feature* para o classificador. A maior quantidade de *features* por FP torna mais eficiente o processo de diferenciação entre os PRs.

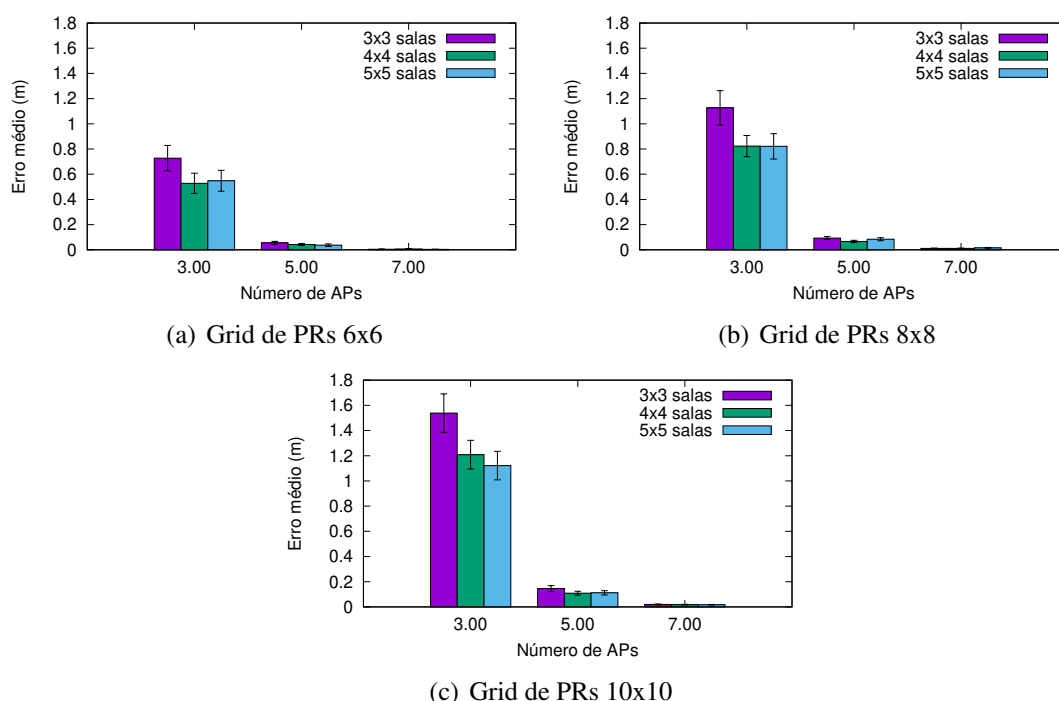


Figura 4. Erro médio da localização usando FPs coletados apenas nos PRs

Analisando ainda os gráficos da Figura 4, pode-se perceber que o aumento do tamanho da *grid* de salas reduz o erro na localização. Isso é um resultado contraintuitivo, uma vez que com o aumento da densidade de salas no prédio, o esperado seria que a atenuação dos sinais fosse mais intensa e o desempenho da localização piorasse. Entretanto, o aumento da densidade de salas torna mais provável a existência de paredes entre dois PRs adjacentes. Isso faz com que a diferenciação entre os RSSIs dos FPs de PRs adjacentes seja mais pronunciada, fazendo com que o classificador “aprenda” essa diferenciação. Dessa forma, a localização pode se beneficiar da existência de paredes, que auxiliam na diferenciação de FPs de PRs próximos.

Uma última observação a ser feita sobre os resultados da Figura 4 diz respeito a influência do tamanho da *grid* de PRs. Aqui novamente vemos um resultado interessante, com o aumento da densidade de PRs há um aumento do erro médio da localização. Nesse

ponto dois fatores influenciam no erro. Com uma densidade baixa de PRs, cada erro de classificação causará um erro grande, pois os PRs estão mais separados espacialmente. Mas com uma densidade alta de PRs a separação espacial entre eles será menor, fazendo com que os FPs de PRs adjacentes sejam muito parecidos, com valores de RSSI próximos. Desta forma, podemos dizer que há um *trade-off* entre a taxa de erros de classificação e o erro absoluto de cada classificação errada. Apesar disso, pelos resultados dos experimentos realizados neste trabalho, o impacto da taxa de erros de classificação foi mais relevante. Por isso, com densidades mais altas de PRs os erros de classificação tornaram-se mais frequentes e o erro médio da localização foi maior.

O próximo conjunto de gráficos da Figura 5 mostram o erro médio da localização para FPs coletados em 50 pontos aleatórios dos mesmos cenários de simulação descritos anteriormente. Para isso, 50 nodes foram posicionados em locais aleatórios do ambiente para a coleta do FPs naqueles locais durante as simulações. Assim, foi possível usar esses FPs de locais aleatórios como o conjunto de teste e os FPs coletados nos PRs como o conjunto de treinamento do classificador. A ideia desse experimento era avaliar o erro em uma situação real de uso do sistema de localização, onde os dispositivos que querem se localizar não estão exatamente nos PRs. Uma primeira observação dos gráficos mostra que o erro médio da localização foi superior. Isso é natural, já que os dispositivos não estão mais localizados exatamente nos PRs e uma classificação, mesmo que correta, irá retornar a localização como sendo em algum PR. Logo, o próprio método de localização já induz um erro de base.

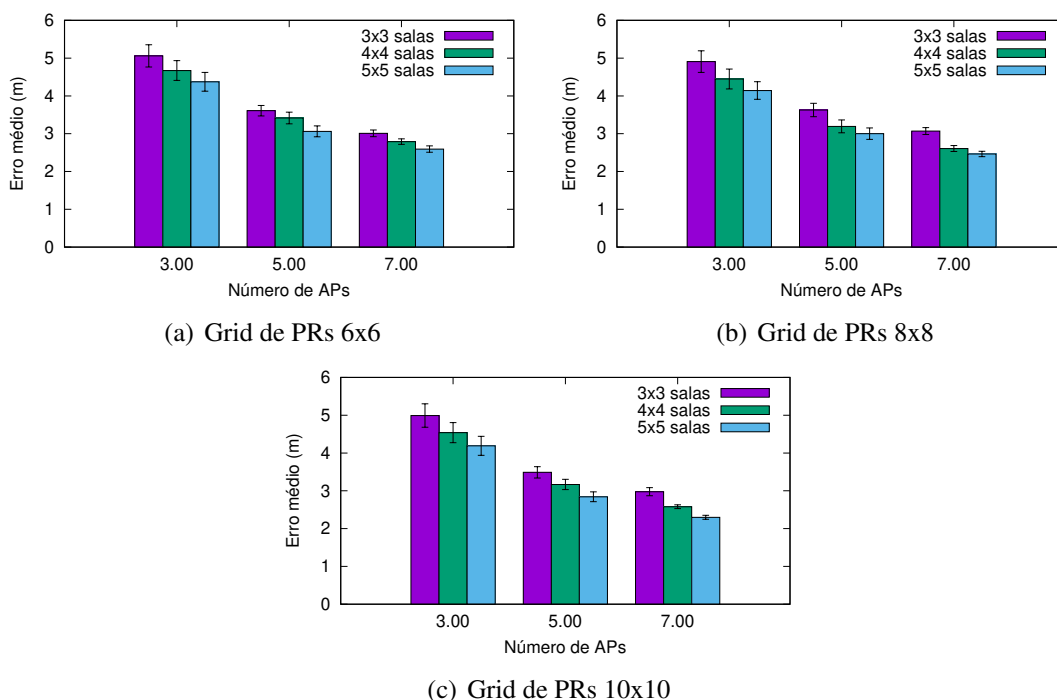


Figura 5. Erro médio da localização usando FPs coletados em locais aleatórios

Observando ainda os gráficos da Figura 5, pode-se perceber que as observações feitas sobre os gráficos da Figura 4 a respeito da quantidade de APs no cenário e da quantidade de salas se mantém as mesmas. Sendo que para os FPs em localizações aleatórias fica mais pronunciada a melhoria de desempenho em cenários com maior densidade de

salas. Apesar disso, o impacto da densidade de PRs não foi tão evidente. Uma possível justificativa para esse resultado é que quando um local é escolhido aleatoriamente, há grande probabilidade do local estar a distâncias parecidas dos PRs mais próximos. Logo, um erro de classificação que escolha o PR que não seja o mais próximo terá um impacto suavizado no erro de localização. Por isso, escolher PRs mais espaçados ou mais próximos influencia pouco no erro de localização neste caso.

6. Conclusões e Trabalhos Futuros

Os sistemas de localização *indoor* fornecem informação de contexto importante para diversas novas aplicações, inclusive para as aplicações de Internet das Coisas. Diversos trabalhos já estudaram o assunto, mas sempre avaliando o desempenho das técnicas propostas através de experimentos em cenários específicos. Essas avaliações geralmente fornecem conclusões limitadas no que diz respeito ao impacto das características do ambiente no desempenho da técnicas propostas.

Neste trabalho, foi apresentada uma metodologia de avaliação de desempenho das técnicas de localização por *fingerprint* de RSSI através de simulações com o NS-3. Usando modelos já existentes no NS-3 foi possível criar cenários que permitem avaliar o impacto da quantidade de APs, da densidade de pontos de referência (PR) e também do tamanho e da quantidade de salas presentes no ambiente. Para avaliar a técnica proposta, foi realizado um estudo de caso para a avaliação de uma sistema de localização baseado no classificador kNN. Os resultados mostram que de fato as características do ambiente influenciam no desempenho da técnica de localização. Além disso, pode-se tirar algumas conclusões não intuitivas, como: o aumento da densidade de salas ajuda no processo de classificação dos FPs, e o aumento da densidade de PRs causa um aumento do erro da localização por tornar os erros de classificação mais prováveis.

Como trabalhos futuros pretende-se realizar avaliações de desempenho extensas para avaliar quais características do ambiente impactam em cada tipo de técnica de localização *indoor* por *fingerprint* de RSSI. Além disso, pretende-se modificar os cenários de simulação para a criação de cenários mais realistas, contendo salas de tamanho variável e corredores, de forma parecida com o que foi feito por [Bugarcic et al. 2021]. Esses cenários mais próximos da realidade podem ser usados inclusive para gerar sinteticamente *datasets* de treinamento para sistemas de localização reais, simplificando a etapa *offline*.

Referências

- Bahl, P. and Padmanabhan, V. N. (2000). RADAR: An in-building RF-based user location and tracking system. In *Proceedings IEEE International Conference on Computer Communications (INFOCOM)*, volume 2, pages 775–784.
- Basiri, A., Lohan, E. S., Moore, T., Winstanley, A., Peltola, P., Hill, C., Amirian, P., and Figueiredo e Silva, P. (2017). Indoor location based services challenges, requirements and usability of current solutions. *Computer Science Review*, 24:1–12.
- Brunato, M. and Battiti, R. (2005). Statistical learning theory for location fingerprinting in wireless LANs. *Computer Networks*, 47(6):825–845.

- Bugarcic, P. D., Jevtic, N. J., and Malnar, M. Z. (2021). An extension of building model for indoor communication in NS-3 simulator. In *2021 29th Telecommunications Forum (TELFOR)*, pages 1–4.
- Carvalho, A. d. S. (2021). Localização em Ambientes Internos utilizando redes IEEE 802.11 e algoritmo WKNN. Master’s thesis, Universidade Federal do Amazonas, Brasil.
- Ferreira, D., Souza, R., and Carvalho, C. (2020). QA-kNN: Indoor localization based on quartile analysis and the knn classifier for wireless networks. *Sensors*, 20(17):4714.
- Haeberlen, A., Flannery, E., Ladd, A. M., Rudys, A., Wallach, D. S., and Kavraki, L. E. (2004). Practical robust localization over large-scale 802.11 wireless networks. In *Proceedings of the 10th annual international conference on Mobile computing and networking*, pages 70–84.
- ITU (2021a). RECOMMENDATION ITU-R P.1238-11 Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 450 GHz. Disponível em <https://www.itu.int/rec/R-REC-P.1238/en>.
- ITU (2021b). Recommendation ITU-R P.1411-11 Propagation data and prediction methods for the planning of short-range outdoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 100 GHz. Disponível em <https://www.itu.int/rec/R-REC-P.1411/en>.
- Jedari, E., Wu, Z., Rashidzadeh, R., and Saif, M. (2015). Wi-Fi based indoor location positioning employing random forest classifier. In *2015 international conference on indoor positioning and indoor navigation (IPIN)*, pages 1–5. IEEE.
- Jiang, Y., Pan, X., Li, K., Lv, Q., Dick, R. P., Hannigan, M., and Shang, L. (2012). ARIEL: Automatic Wi-Fi based room fingerprinting for indoor localization. In *Proceedings of the 2012 ACM conference on ubiquitous computing*, pages 441–450.
- NS-3 (2022a). Buildings Module - Model Library. <https://www.nsnam.org/docs/models/html/buildings.html>.
- NS-3 (2022b). ns-3 | a discrete-event network simulator for internet systems. <https://www.nsnam.org/>.
- Rappaport, T. S. et al. (1996). *Wireless communications: principles and practice*, volume 2. New Jersey: prentice hall PTR.
- Sadowski, S. and Spachos, P. (2018). RSSI-Based Indoor Localization With the Internet of Things. *IEEE Access*, 6:30149–30161.
- Singh, N., Choe, S., and Punmiya, R. (2021). Machine Learning Based Indoor Localization Using Wi-Fi RSSI Fingerprints: An Overview. *IEEE Access*.
- Zafari, F., Gkelias, A., and Leung, K. K. (2019). A survey of indoor localization systems and technologies. *IEEE Communications Surveys Tutorials*, 21(3):2568–2599.