ELIEL MARLON DE LIMA PINTO MOREIRA

Estratégia para Detecção de Ataques de Spoofing em Redes Sem Fio

Curitiba - PR, Brasil

2022

ELIEL MARLON DE LIMA PINTO MOREIRA

Estratégia para Detecção de Ataques de Spoofing em Redes Sem Fio

Tese apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de doutor em Informática.

Pontifícia Universidade Católica do Paraná - PUCPR

Programa de Pós-Graduação em Informática - PPGIa

Orientador: MARCELO EDUARDO PELLENZ

Curitiba - PR, Brasil

2022

AGRADECIMENTOS

É difícil para mim expressar em poucas palavras a grande gratidão que sinto pelas pessoas que fizeram parte de minha caminhada nos dez últimos anos na Pontifícia Universidade Católica do Paraná. Todas foram especiais e importantes desde o primeiro dia.

Primeiramente agradeço à Deus, fonte de toda sabedoria, pela inspiração, força nas adversidades e principalmente pelas pessoas que colocou em meu caminho para que eu tivesse a oportunidade de concretizar um grande sonho.

Gostaria de expressar minha profunda admiração ao meu orientador Prof. Marcelo E. Pellenz, Prof. Edgard Jamhour e Prof. Manoel Camillo Pena. Agradeço pelo incentivo, paciência, apoio e amizade. Serei para sempre grata por guiarem meus passos na grande aventura do conhecimento. Sem vocês este trabalho não seria possível.

Agradeço à Pontifícia Universidade Católica do Paraná por ter me acolhido tão bem durante tantos anos e pelo apoio financeiro.

Além disso, gostaria de agradecer a todos os funcionários, que provavelmente não imaginam como seus pequenos e grandes gestos significaram tanto para mim. Agradeço também aos colegas e colaboradores dos ambientes de pesquisa que foram fundamentais para a execução das etapas desse trabalho. Seja no SINE de Tubarão, na Escola Dom Joaquim ou na Unisul de Tubarão.

Finalmente, agradeço minha família pela compreensão e apoio. Especialmente às minhas filhas Elis Moreira de Lima e Lívia Moreira de Lima que foram a principal fonte de incentivo. Agradeço ainda a minha esposa Tatianne Dias Moreira que tem sido uma referência para nossas conquistas.

Obrigado!

Se eu vi mais longe, foi por estar sobre ombros de gigantes. Isaac Newton

Dedico essa tese aos meus pais, a minha esposa, a minhas filhas e aos meus diretores

"Quem atribui à crise seus fracassos e penúrias, violenta seu próprio talento e respeita mais os problemas do que as soluções".

Albert Einstein

RESUMO

Contexto: Atualmente, existe uma grande variedade de tecnologias de rádio de baixo custo que estão sendo usadas para comunicação sem fio em aplicações emergentes importantes, como smart grids, smart cities e Internet das Coisas (IoT). No entanto, a facilidade de acesso a estas novas tecnologias, como por exemplo os rádios definidos por software, trazem problemas de segurança pela facilidade em se efetuar ataques na camada física. Um usuário malicioso pode executar varreduras passivas do canal de rádio móvel e usar essas informações para lançar ataques de falsa identidade na rede sem fio. Estes ataques podem gerar instabilidade e degradação de desempenho destas redes. Neste contexto, focamos na detecção de ataques de identidade em redes locais sem fio (WLANs). Objetivos: Este trabalho propõe uma estratégia de detecção de ataques de spoofing na camada física de WLANs. A estratégia explora medidas de potência de sinal recebido dos clientes pelos Access Points (APs), juntamente com o princípio de sombreamento que ocorre durante o processo de transmissão dos dispositivos no canal de rádio móvel. As informações de camada física são processadas por algoritmos de aprendizagem de máquina para a identificação da presença de ataques. Método: Com base em modelos analíticos detalhados para o canal de rádio móvel, o algoritmo proposto combina dois classificadores para processar e analisar as amostras instantâneas de intensidade do sinal coletadas na rede. O algoritmo é otimizado para cenários em que o nó legítimo e o nó atacante estão em condições similares de propagação com relação aos APs, o que é um cenário de pior caso para detecção de ataques. Resultados: Como resultado desta pesquisa desenvolvemos duas estratégias. A primeira estratégia é um método geral que foi validado através de simulações computacionais usando uma modelagem analítica para o canal de rádio móvel e que pode ser aplicada a diferentes contextos e tecnologias de redes de comunicação sem fio. A segunda estratégia é uma abordagem simplificada aplicada no contexto de WLANs, onde informações reais da rede foram coletadas, considerando um cenário com um nó verdadeiro e um nó atacante. O método assume a existência de múltiplos APs que fazem a coleta das informações para análise. Os resultados demonstram que a abordagem proposta consegue melhorar em aproximadamente 10% a detecção de ataques, quando comparado com estratégias similares da literatura. Conclusão: Neste trabalho abordamos o problema da detecção de ataques de identidade na camada física das redes locais sem fio. Validamos a hipótese de se utilizar o fenômeno de propagação denominado sombreamento como princípio para identificar ataques. As estratégias propostas demonstrarem ser eficientes para identificação de ataques e podem ser adaptadas para outros cenários e tecnologias de rede sem fio.

Palavras-chave: Redes Sem Fio, WLANs, Ataque de Identidade, Aprendizagem de Máquina, Canal de Rádio Móvel.

ABSTRACT

Context: Currently, a wide variety of low-cost radio technologies are being used for wireless communication in important emerging applications such as smart grids, smart cities, and the Internet of Things (IoT). However, the ease of access to these new technologies, such as software-defined radios, brings security problems due to the ease of carrying out attacks on the physical layer. A malicious user can perform passive scans of the mobile radio channel and use this information to launch impersonation attacks on the wireless network. These attacks can generate instability and performance degradation in these networks. In this context, we focus on detecting identity attacks on wireless local area networks (WLANs). Objectives: This work proposes a strategy for detecting spoofing attacks at the physical layer of WLANs. The strategy exploits measurements of signal strength received from clients by Access Points (APs), along with the principle of shadowing that occurs during the transmission process of devices on the mobile radio channel. Machine learning algorithms process physical layer information to identify the presence of attacks. Method: Based on detailed analytical models for the mobile radio channel, the proposed algorithm combines two classifiers to process and analyze the instantaneous signal strength samples collected in the network. The algorithm is optimized for scenarios where the legitimate node and the attacking node are in similar propagation conditions concerning the APs, which is a worst-case scenario for attack detection. Results: As a result of this research, we developed two strategies. The first strategy is a general method validated through computer simulations using analytical modeling for the mobile radio channel and applied to different wireless communication network contexts and technologies. A second strategy is a simplified approach applied in WLANs, where real network information was collected, considering a scenario with a true node and an attacking node. The method assumes the existence of multiple APs that collect the information for analysis. The results demonstrate that the proposed approach can improve attack detection by approximately 10%, compared to similar strategies in the literature. Conclusion: In this work, we address the problem of detecting identity attacks at the physical layer of wireless local area networks. We validate the hypothesis of using the propagation phenomenon called shadowing as a principle to identify attacks. The proposed strategies efficiently identify attacks and can be adapted to other wireless network scenarios and technologies.

Keywords: Wireless Networks, WLANs, Identity Attack, Machine Learning, Mobile Radio Channel.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ARP	Protocolo de Resolução de Endereços (Address Resolution Protocol)
DR	Taxa de Detecção (Detection Rate)
RSSI	Indicador de Intensidade do Sinal Recebido (*Received Signal Strength Indication*)
RSS	Potência do Sinal Recebido (*Received Signal Strength*)
CSI	Informação de Estado do Canal (Channel State Information)
CDF	Distribuição Acumulada de Probabilidade
FPR	Taxa de Falso Positivo (False Positive Rate)
IoT	Internet das Coisas (Internet of Things)
K-means	Método de Clusterização K-means.
KNN	Método de Clusterização k-Nearest Neighbor
M2M	Máquina-para-Máquina (Machine-to-Machine)
MAC	Controle de Acesso ao Meio (Media Access Control)
MITM	Homem no Meio (Man In The Middle)
OFDM	Multiplexação por Divisão de Frequências Ortogonais (*Orthogonal Frequency Division Multiplexing*)
RSSF	Redes de Sensores sem Fio
SMS	Serviço de Mensagens Curtas (Short Message Service)
SNR	Relação Sinal-Ruído (*Signal-to-Noise Ratio*)
TCP	Transmission Control Protocol

LISTA DE SÍMBOLOS

d_0	Distância de Referência
d_{ij}	Distância entre os nós i e j
h_{ij}	Ganho do Canal
N	Potência do Ruído
P_r	Potência Média Recebida
S	Sombreamento
γ_{ij}	Relação Sinal-Ruído Instantânea
α	Expoente de Perda de Percurso
σ^2	Variância

SUMÁRIO

1 INTRODUÇÃO

Nos últimos tempos, as Redes de Sensores sem Fio (RSSF) tornaram-se um elemento essencial da Internet das Coisas (Internet of Things - IoT), assim como para o cotidiano da sociedade, cujas aplicações podem envolver implantação massivas de dispositivos e um grande volume agregado de dados. A interconectividade de equipamentos domésticos, comerciais e industriais, tem despertado constante preocupação para estudiosos de segurança da informação. No contexto atual, as redes de computadores são algo inerente ao nosso cotidiano, principalmente as redes locais sem fio WLANs. Temos uma crescente implantação de WLANs devido ao baixo custo, facilidade de instalação, capacidade de transmissão e considerada mobilidade para seus usuários. Uma outra característica relevante das WLANs é a flexibilidade para agregar novos serviços, interagindo assim com novas plataformas tecnológicas como por exemplo a Internet das Coisas (SHIHUAN, 2010).

Para tentar entender a abrangência da interação tecnológica, se faz necessário estudar um pouco dos dispositivos que constituem a ideia de IoT (KHAN et al., 2015). Estes dispositivos estão presentes nas mais diversas áreas, como por exemplo saúde, agricultura, automação residencial, Indústria 4.0 e cidades inteligentes. Em geral estes dispositivos possuem dimensões reduzidas e estão aptos a realizar medições, calcular e transmitir informações, atendendo assim as demandas diárias.

Com isso, a segurança é um fator que causa preocupação nos pesquisadores da área, pois a confidencialidade, integridade e disponibilidade dessas informações não podem ser ignoradas. Além disso, existe uma diversidade de equipamentos que permeiam a vida moderna, como por exemplo desktops, tablets, notebooks, smartphones, smart TVs e outros, conectados à internet por meio das WLANs. Tais redes estão suscetíveis a ataques maliciosos como os ataques *sybil* (DEMIRBAS; SONG, 2006) e *spoofing* (ARBAUGH et al., 2002).

Como forma de oferecer um ambiente mais seguro aos usuários, é importante identificar ataques do tipo *sybil* e *spoofing* em uma WLAN. Em um ataque *sybil* (SHENG et al., 2008), um nó tenta forjar múltiplas identidades. Os ataques *sybil* são particularmente fáceis de lançar em WLANs devido ao meio de comunicação sem fio que permite transmitir e interferir de forma direta ou indireta no desempenho da rede. No ataque *spoofing* o tráfego parece vir de um dispositivo legítimo da rede, o que faz a estrutura do ataque parecer inocente, para que não seja detectado (BERNASCHI; FERRERI; VALCAMONICI, 2008). Diante disso, os ataques *sybil* e *spoofing* se apresentam como grandes oportunidades para teste de soluções de segurança. O uso de informações de intensidade do sinal recebido (Received Signal Strength - RSS) podem auxiliar significativamente, pois através da análise

desses dados podemos identificar medidas preventivas ou corretivas para esses ataques.

Algoritmos usando informações de RSS juntamente com técnicas de aprendizagem de máquina (Machine Learning - ML) têm sido propostos na literatura (Chen et al., 2010), (YANG et al., 2012), (WU et al., 2020), (ZHENG; MASUDA; SHIBATA, 2021) como forma de auxiliar o posicionamento de dispositivos no ambiente, e assim, propor métodos que auxiliem na detecção de algum nó malicioso em redes sem fio. Com isso, através da análise desses dados, definir se existe ou não um ataque a rede em questão. Posteriormente, seja propostas medidas preventivas ou corretivas contra estes ataques.(HUAN; KIM; ZHANG, 2021), (Chen et al., 2010), (YANG et al., 2012), (PEI et al., 2014), (FUKUSHIMA et al., 2022), (TANG et al., 2020), (WANG et al., 2021), (XIE; LI; TAN, 2020).

É importante observar que WLANs com múltiplos APs, principalmente em ambientes urbanos, podem estar mais suscetíveis a ataques. Assim, as informações de RSS afetadas pelo efeito do sombreamento (shadowing) do canal de rádio, podem ser exploradas para análise do tráfego da WLAN, visando confirmar ou descartar a presença de possíveis ataques do tipo *sybil* e do tipo *spoofing*. Neste sentido é importante implementar novas soluções para detectar ataques, sem um aumento significativo de custos, do consumo de energia, do tráfego da rede, de processamento e de outros problemas inerentes a agregação de serviços em ambientes WLANs.

Nesta tese, se propõe uma estratégia para detecção de ataques *spoofing* com uso de medidas de RSS em WLANs, sem a necessidade de implantação de novos recursos físicos. Assim, nossa proposta é uma solução eficiente, confiável e escalável para detecção de dispositivos falsos que visam ataques de identidade.

1.1 Motivação

A crescente utilização das redes WiFi em ambientes industriais, comerciais e residenciais, trouxe a potencialização de ataques de identidades como *spoofing*, o qual segundo o ICP Brasil é considerado uma ameaça para o uso destas redes (RIBEIRO; JUNIOR; ARANHA, 2017). As WLANs são uma solução bem difundida e consolidada. Portanto, novas plataformas tecnológicas vêm utilizando sua infraestrutura, trazendo junto alguns problemas inerentes a qualquer crescimento não planejado. Dentre estes problemas, podemos citar a taxa de transmissão de dados insuficiente a novas demandas, passando por congestionamentos indesejados (perda de desempenho) e a questão da segurança, tanto física, quanto lógica. A segurança lógica em particular, precisa ser consistente o suficiente, sem sobrecarregar recursos como energia, memória e/ou processamento.

Além disso, no contexto atual, novos equipamentos são agregados a rede. Assim, serviços inteligentes através da IoT nas WLANs vão gerando mais demanda. Isso acontece, por exemplo, através de eletroeletrônicos domésticos e de serviços públicos ou privados

de monitoramento. Porém, interligar dispositivos como carros, geladeiras, microondas, TVs, PCs, notebooks e tablets a essas redes de forma segura, transparente, eficiente e sem custos adicionais, é algo desafiador.

A infraestrutura, o gerenciamento e a segurança das WLANs são aspectos cruciais para o bom funcionamento dessas redes. Diante disso, aplicações e serviços do cotidiano que utilizam WLAN, têm sofrido constantemente com as fragilidades dos atuais sistemas da segurança frente a novas demandas. Tais fragilidades se apresentam através de ataques maliciosos de várias naturezas, com destaque para os ataques *spoofing* e *sybil*. Um ataque de *spoofing* pode ser descrito de forma geral (HUAN; KIM; ZHANG, 2021) como sendo nós maliciosos, que tenta se passar por nós verdadeiros, enviando pacotes falsos na rede. A expansão das soluções para IoT e Indústria 4.0, (YAQIONG et al., 2018) são demandas que podem gerar um tráfego considerável de informações em uma WLAN. Portanto, se faz necessário estudos que apresentem trafegabilidade segura nessas redes, como forma de atender estes cenários, e que se utilizem de informações da camada física.

Com base na análise dessas informações é desejável a implementação, de estratégias de detecção de ataques *spoofing* em redes WiFi comuns, como forma de se evitar que serviços/sistemas sejam vítimas de ataques maliciosos. Tais ataques costumam se utilizar de informações, da camada física as quais estão disponíveis em qualquer WLAN. O propósito deste trabalho de pesquisa é aprofundar os estudos de segurança em WLANs, com foco em ataques *spoofing*, usando informações de camada física, as quais servirão de base para uma proposta, sem a necessidade de implementação de novos recursos físicos nas redes. Os ambientes para detecção podem ser redes em ambientes domésticos, empresariais ou universidades, onde existam múltiplos APs instalados.

1.2 Objetivos

O objetivo principal deste trabalho é a detecção de um nó falso em ambientes de rede WiFi. Assim, este trabalho é uma proposta de implementação de um método para tratamento de ataques *spoofing* em WLANs, usando medidas de RSS processadas usando técnicas de aprendizagem de máquina (*machine learning*), que sejam mais eficientes em relação aos modelos já implementados na literatura. A estratégia deve ser leve no tocante ao tráfego para coleta e processamento das informações. Também é desejável utilizar poucos recursos adicionais de hardware.

1.2.1 Objetivos Específicos

1. Estudar e selecionar modelos matemáticos para o fenômeno de sombreamento (*shadowing*) que ocorre no canal de rádio móvel durante o processo de transmissão de dados da rede sem fio;

2. Propor estratégias para detecção de ataques de spoofing em WLANs usando medidas de RSS e técnicas de aprendizagem de máquina;

3. Avaliar o algoritmo de detecção de nós maliciosos, considerando a distância, o cenário de rede e os parâmetros do método proposto.

4. Avaliar o desempenho do método proposto através de simulação do comportamento do canal de rádio;

5. Adaptar e avaliar o desempenho do método proposto usando dados reais coletados.

1.3 Justificativa

No contexto tecnológico em que vivemos, as redes WiFi exercem uma função muito importante, facilitando o acesso a Internet para o uso dos mais diversos serviços, seja em ambiente residencial ou empresarial. Contudo, o aspecto de segurança é uma questão crucial, seja em nível de aplicação, de rede ou mesmo de camada física. Estes ataques podem ser localizados ou distribuídos. Diante da possibilidade de tais ameaças é importante identificar sua origem e estudar sua forma de atuação. Posteriormente tratar ataques de forma eficiente, a ponto deles não influenciarem no bom funcionamento da rede.

Podemos encontrar vários tipos de ataques em rede de sensores sem fio, mas nosso trabalho visa aprofundar os estudos no ataque *spoofing* em ambiente de rede WiFi. Embora estes ambientes ofereçam mecanismos elaborados de segurança, como autenticação e criptografia, podem ainda assim estar suscetíveis a ataques de camada física.

Os problemas de segurança comuns de redes WiFi costumam ser potencializados, por lacunas deixadas em redes mal gerenciadas, associadas a usuários inexperientes e que ainda carecem de um aculturamento de segurança da informação, tanto em ambiente escolar quanto em empresarial. Assim, esta pesquisa visa tais lacunas tecnológicas, sem necessariamente precisar investir em novas tecnologias físicas, mas elaborar etapas de procedimentos para melhores práticas de uso de ambientes de redes WiFi. O estudo exploratório visa propor um modelo para reduzir o índice de ataques digitais, com vazamento de dados pessoais na internet através de redes WiFi. Assim, propor novas técnicas, como formas de evitar danos as redes WiFi comuns, com a utilização apenas do *RSS*.

1.4 Publicações

- E. M. d. L. Pinto, R. Lachowski, M. E. Pellenz, M. C. Penna and R. D. Souza, "A Machine Learning Approach for Detecting Spoofing Attacks in Wireless Sensor Networks," 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), 2018, pp. 752-758, doi: 10.1109/AINA.2018.00113.

1.5 Estrutura do Documento

Este documento está estruturado da seguinte forma:

- Capítulo 2 – Apresenta conceitos, características e requisitos associados às WiFi. Além disso, este capítulo aborda a relevância da segurança da informação nas WiFi, seja em ambiente reais comuns, seja na IoT, WLAN e RSSF.
- Capítulo 3 – Investiga as propostas já disponíveis na literatura que se propõe a identificar ataques de identidade, como os ataques *spoofing* com uso do RSS.
- Capítulo 4 – Descreve o método proposto. Apresentamos a arquitetura, explicamos o funcionamento e a estratégia utilizada para o mecanismo de detecção de ataques *spoofing*. Propomos e descrevemos uma estratégia para identificar a presença de um nó falso dentro de uma determinada rede, analisando apenas a variação dos valores de RSS.
- Capítulo 5 – Descreve as simulações nos cenários que foram realizadas para avaliar a eficiência e confiabilidade da proposta.
- Capítulo 6 – Discute as direções futuras deste trabalho e apresenta as conclusões finais.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo apresentamos os principais conceitos relacionados às redes sem fio, incluindo RSSFs e sua relevância para a IoT, e particularmente as redes WiFi. Também se apresentou os principais conceitos sobre os ataques nas redes sem fio (JAWANDHIYA et al., 2010). Assim, será aplicado técnicas de *machine learning* com o *K-means*, visando apresentar soluções que visam aprimorar a segurança em redes WiFi. Na oportunidade a modelagem do canal de rádio móvel, junto com conceitos básicos de WLANs, assim como os fundamentos do uso do RSS para detecção de invasores será objeto de estudos.

Assim, com os avanços tecnológicos de comunicações e redes WiFi agilizam de forma substancial o cotidiano do cidadão moderno, pois através, de dispositivos inteligentes, como laptops, smartphones, sistemas de monitoramento e dispositivos da Internet das Coisas (IoT) (ZANELLA et al., 2014) o mundo simplesmente está ao alcance das mãos do usuário. Com isso, um considerado número de informações sensíveis estão disponíveis na rede, o que pode servir de base para possíveis ataques.

Para isso, algumas técnicas baseadas em identificadores pré-definidos (IDs) ou endereços de controle de acesso ao meio (MAC) têm sido amplamente utilizadas para identificação de nós. Assim, recursos tecnológicos, podem se passar por recursos legítimos, para que invasores entrem numa rede, interceptem dados trocados e até lancem ataques. (ZOU et al., 2016), (XIAO et al., 2019). Dentre esses ataques, podemos citar o ataque *spoofing* (HUAN; KIM; ZHANG, 2021) (WANG; MA; BAI, 2020). Em comparação com as redes cabeadas, as redes sem fio são mais vulneráveis a ataques de segurança, devido à recursos limitados e a falta de uma autoridade confiável centralizada para melhor controlar as informações de seu meio.

2.1 Redes Sem Fio

2.1.1 Redes de Sensores Sem Fio

Para (Chen et al., 2010), a proporção que mais redes WiFi e de sensores são implementadas, estas se tornarão cada vez mais alvos tentadores para ataques maliciosos. Tal situação se dá devido à natureza compartilhada do meio sem fio. Os invasores podem coletar informações de identidade úteis durante o monitoramento passivo e utilizar as informações para lançar ataques baseados de identidade. Em particular, existem dois ataques mais prejudiciais e comuns: ataques de *Spoofing* e ataques de *Sybil*. Com tais ataques de falsificação baseados em identidade, um invasor pode forjar sua identidade para se passar por outro dispositivo ou até mesmo criar várias identidades ilegítimas nas redes.

Diante disso, é necessário que soluções para WiFi sejam escaláveis, que mantenham a eficiência e viabilidade em aplicações de larga-escala. Nesse caso, o termo *larga-escala* refere-se ao fato de que a área monitorada é extensa, e os nós sensores estão densamente implantados (DJEDOUBOUM et al., 2018). Visando tal realidade, várias aplicações demandam implantações em larga-escala como: agricultura inteligente, monitoramento ambiental, sistemas inteligentes de vigilância, cidades inteligentes dentre outros (JAGANNATH et al., 2019). Nestes cenários, um dos maiores desafios é a coleta de grandes volumes de dados (RANI et al., 2017; TAKAISHI et al., 2014) o que exige um cuidado extra na segurança dessa rede.

Outro grande desafio das redes WiFi, mais precisamente as redes de sensores sem fio, é o ambiente altamente dinâmico. A instabilidade dos enlaces sem fio causada por mobilidade, esgotamento de recursos energéticos, falhas e efeitos do canal sem fio torna a perda de pacotes um evento comum. Além disso, essas redes, podem ser implantadas em regiões que apresentam condições climáticas adversas e bem hostis ao seu funcionamento. Por estes motivos, a topologia da rede pode alterar-se rapidamente, consequentemente tornar as informações armazenadas nos nós ultrapassadas. Assim também como, os nós sensores utilizam estas informações para construir as rotas ascendentes e descendentes a fim de encaminhar dados e consultas. Portanto, a instabilidade do ambiente exige soluções robustas e mecanismos de confiabilidade, para manter a rede operacional.

Uma das características mais marcantes das RSSFs é o foco nas informações fornecidas pela rede. Ao contrário das redes de comunicação tradicionais, o endereço ou identidade dos dispositivos que fornecem a informação não são relevantes. As RSSFs são inerentemente centradas em dados (HOLGER; WILLIG., 2005). A importância está na informação, que pode ser o resultado de dados monitorados por vários dispositivos distintos.

As redes de sensores sem fio se enquadram na categoria de redes de baixa potência e com alta probabilidade de perda de pacotes. Assim, tais redes são constituídas por um grande número de dispositivos com recursos limitados, no tocante a energia, armazenamento e processamento, afetando assim, a comunicação confiável entre os pontos. Até porque, as redes de sensores sem fio, costumam ser uma extensões para ambientes residenciais, comerciais e industriais, que muitas vezes se interconectam as redes WiFi.

Diante disso, como todos os dispositivos produzem RSS, ou seja, nó verdadeiro, seja nó falso, um nó falso pode tentar se passar por um nó verdadeiro da rede e assim começar a capturar informações. Posteriormente usar informações comuns, tentar se passar por um nó verdadeiro, comprometer o funcionamento da rede. Em cenários assim, que se faz necessário o estudo aprofundado das redes WiFi comuns, as quais disponibilizam os RSSs, como métrica para detecção ou não, de um nó malicioso ou falso na rede em questão.

2.1.2 Internet das Coisas (IoT)

Para (RATASUK; MANGALVEDHE; GHOSH, 2015), o crescimento exponencial das tecnologias *IoT* trará ao nosso cotidiano um número crescente de novas aplicações práticas e acessíveis, entre as quais podemos encontrar desde aplicações em rastreamento de ativos, agricultura automatizada com medição inteligente, até cidades inteligentes e casas inteligentes (MEKKI et al., 2019). A Internet das Coisas (SHADEED; MOREB, 2021) permite que tanto pessoas, quanto objetos e até dispositivos, se comuniquem com qualquer ser humano, em qualquer lugar e a qualquer hora, usando caminhos e recursos de redes e serviços.

No entanto, para (HAMDAN et al., 2019) a maioria dos dispositivos de IoT apresentam considerada facilidade de invasão. Normalmente, os dispositivos IoT são limitados em capacidade de computação, armazenamento e rede, portanto, são mais vulnerável a ataques do que outros dispositivos de terminal, como smartphones, tablets ou computadores. Assim, dispositivos heterogêneos que trocam informações entre si, nesse tipo de cenário, o número de objetos inteligentes é cada vez mais alto, o que aumenta a possibilidade de ataques que comprometam a segurança, a privacidade e a confiança das informações (PÉREZ et al., 2016). Ao mesmo tempo que a IoT abre oportunidades para dispositivos portáteis, como eletrodomésticos inteligente conectados a internet, os dados compartilhados entre esses dispositivos, contêm uma grande quantidade de informações privadas, e preservar a segurança desses dados é uma questão relevante que não pode ser negligenciada (ZHANG et al., 2014).

Para (RAJAN; JITHISH; SANKARAN, 2017), as IoTs são vulneráveis a ataques *sybil*, onde o atacante fabrica identidades fictícias ou rouba identidades de nós legítimos. Já para situação de ataques spoofing na IoT, para (NAWIR et al., 2016) no início, o *spoofer* não transmite um sinal, mas apenas ouve o transmissor apropriado, quando o nó legítimo transmissor para de enviar um sinal ao receptor legítimo, o *spoofer* começa a enviar o sinal não confiável. Diante disso, se faz necessário considerada atenção no tocante a possíveis ataques com caráter *spoofing* e/ou *sybil* em ambientes que se utilizam de tecnologias IoT.

2.1.3 Redes Locais Sem Fio

No tocante as redes locais sem fio (Wireless Local Area Networks - WLANs) (FUKUSHIMA et al., 2022) a localização é fator fundamental para soluções internas de Internet das Coisas, sensores incorporados e maior eficiência no gerenciamento destas redes. Assim, uma WLAN é algo inerente ao cotidiano do mundo moderno, dando suporte a aplicações modernas usando smartphones, ou mesmo um sensor medidor de glicemia ou marcapasso no corpo humano.

Diante de tantos dispositivos, com tantas informações, um gerente de rede ao buscar

de mais eficiência do ambiente de WiFi, se faz imperativo contar com a segurança física e lógica da rede em questão. Até porque, o monitoramento periódico, das condições físicas e ambientais dessas redes, é uma necessidade frequente, e estas são variáveis de segurança que precisam de atenção frequente. No tocante as RSSF, o envio de comandos e consultas para os nós sensores, é uma necessidade inerente a realidade no cotidiano (IOVA et al., 2016).

Até porque, frequentemente as aplicações requerem que comandos e consultas possam ser encaminhados, para um conjunto de dispositivos em casos de monitoramento de ambientes. Para isso, os usuários devem ser capazes de se comunicar com a rede, assim também como, solicitar informações, enviar comandos para nós implantados em certa área ou que monitoram determinado parâmetro da rede(XIAO et al., 2019; BOUBICHE et al., 2018), que visa manter os serviços, destas redes de maneira segurança, eficiente e íntegra.

2.2 Modelagem do Canal de Rádio Móvel

Para melhor entendimento das redes WLANs no tocante a comunicação entre os nós, ou na segurança das informação que recebem maior atenção nesse trabalho, uma breve abordagem do modelo do canal proposto e do *RSS* é apresentado nas sessões posteriores.

2.2.1 Modelo do Canal

Na proposta desse trabalho, assumimos que um canal sem fio está sujeito ao desvanecimento quase estático. A potência instantânea (RSS - Received Signal Strength) medida por um AP (landmark) a partir de um pacote recebido de um nó i é dado por (GOLDSMITH, 2005)

$$\gamma_{ij} = h_{ij}^2 \; P_r \tag{2.1}$$

onde P_r é a potência média de recepção. Assim como, desvanecimento rápido do canal h_{ij} segue a distribuição Nakagami-m, enquanto h_{ij}^2 segue a distribuição gama. A função densidade de probabilidade de γ_{ij} é

$$f_{\gamma_{ij}}(x) = \frac{(m/\overline{\gamma}_{ij})^m \; x^{m-1}}{\Gamma(m) \; e^{mx/\overline{\gamma}_{ij}}} \tag{2.2}$$

onde $\Gamma(a) = \int_0^\infty y^{a-1} \; e^{-y} dy$ é a função gama para $\overline{h^2}$ unitário, $E[h^2{=}\overline{h^2} = 1$ e $\overline{\gamma}_{ij} = P_r$ é o RSS médio, a média de temos:

$$P_r(d_i) \; [\text{dBm}] = P_r(d_0) \; [\text{dBm}] - 10 \; \alpha_i \log_{10} \left(\frac{d_i}{d_0} \right) + S_i^t, \tag{2.3}$$

onde d_i é a distância entre o nó i e o ponto de referência, $Pr(d_0)$ representa a potência recebida a uma distância de referência d_0, o parâmetro α_i é o expoente de perda de percurso e S_i^t representa o sombreamento no instante t, que segue uma distribuição Gaussiana de média zero e variância σ_i^2 em dB (CHEN et al., 2010).

2.2.1.1 Efeito de Sombreamento (Shadowing)

Este modelo de propagação estatístico provê uma estimativa aproximada do que representa a perda de percurso como uma função da distância e outros parâmetros como a frequência e a altura da antena (STÜBER; STÈUBER, 1996). Desta forma é possível obter uma previsão da qualidade da intensidade do sinal quando aumenta-se a distância entre o transmissor e receptor. Vale ressaltar que, para se obter uma representação mais realista da perda do sinal, é necessária uma coleta significativa de dados empíricos. Uma possível explicação para o motivo pelo qual este modelo utiliza uma distribuição log-normal é que, para cada caminho há muitos fatores que contribuem com a perda do sinal, incluindo a combinação de perda no espaço livre, difrações, reflexões, interferências de equipamentos, dentre outros motivos.

O sombreamento tem uma correlação empírica ao longo da distância para diferentes ambientes (GOLDSMITH, 2005). A correlação espacial do sombreamento entre dois pontos separados por uma distância δ é assumido como (GUDMUNDSON, 1991)

$$\rho_s(\delta) = \sigma_i^2 \ e^{-\delta/X_t} \tag{2.4}$$

onde X_t é a distância de descorrelação, que é a distância em que a autocorrelação do sinal é igual a $1/e$ do seu valor máximo. Tipicamente, X_c varia de 50 a 100 m para ambientes externos (WEITZEN; LOWE, 2002). Além da correlação espacial, o processo de sombreamento também apresenta uma dependência temporal. Em particular, esse modelo assume uma correlação temporal para S_i^t por mais tempo τ é dado por

$$\rho_t(\tau) = E[S_i^t, S_i^{t+\tau}] = e^{-\tau/X_t} \tag{2.5}$$

onde X_t é o tempo de descorrelação. A potência recebida P_r (d_0) é definida como uma potência de referência que na prática deve ser medida a uma distância d_0 do transmissor. Ela depende da frequência de operação f (comprimento de onda λ), da potência de transmissão P_t, dos ganhos das antenas de transmissão e recepção, G_t e G_r e da distância d entre transmissor e receptor.

2.2.2 Aplicações da Métrica de RSS (Received Signal Strength)

Avanços na comunicação sem fio, com sensores de baixa potência e microcontroladores permitem a implantação de redes de sensores sem fio em larga escala, porém um problema fundamental nessas redes é a determinação das localizações geográficas dos nós sensores (DAN; HALDER; DASBIT, 2011). A localização é um problema desafiador, principalmente se a questão de baixo custo é algo relevante.

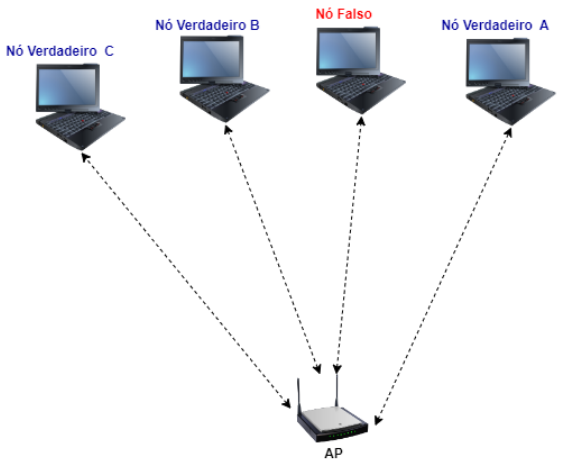
As informações do RSS podem ser utilizadas tanto para nós estáticos, quanto para nós em movimento, sem a necessidade de novos investimentos em hardware para a

localização centralizada na rede (AWAD; FRUNZKE; DRESSLER, 2007), reduzindo os custos. Pesquisas revelaram que a correlação entre distância e RSS é a chave das tecnologias de localização nas WLANs (XU et al., 2010) e (ADEWUMI; DJOUANI; KURIEN, 2013). Já (HEURTEFEUX; VALOIS, 2012) diz que o posicionamento absoluto nem sempre está disponível, assim, a localização baseado em RSS é algo bem popular, pelo fato de não exigir hardware extra ao contrário das soluções baseadas em infravermelho ou ultrassônico. Na teoria o RSS fornece a localização em função da distância. No entanto, alguns cuidados são necessários ao uso do RSS como métrica, pois à distância envolve erros nos valores medidos, como efeitos de perda de percurso, desvanecimento e sombreamento (HEURTEFEUX; VALOIS, 2012). Na proposta esplanada por esse trabalho, o RSS é uma variável chave, que visa realizar um estudo aprofundado, sobre a presença ou não de um nó falso numa rede WiFi comum residencial ou comercial.

2.3 Ataque Spoofing

O spoofing é um ataque no qual um atacante se passa por outro dispositivo ou usuário de uma rede sem fio no intuito de capturar dados, disseminar *malware* e até mesmo driblar os controles de acesso. Assim, o spoofing é um tipo de falsificação de identidade que visa enganar uma rede, um sistema ou uma pessoa, fazendo acreditar que a origem de uma informação é verdadeira, quando na realidade tal origem é falsa, caracterizando um ataque (AVAST, 2019). Esse tipo de ataque é baseado em falsificação de identidade, assim podem causar sérias ameaças à rede, pois representam uma forma de comprometimento de identidade e com isso, podem facilitar uma série de ataques de injeção de tráfego (LOURENÇO, 2013).

Figura 1 – Exemplo de Ataque Spoofing em uma WLAN.



Fonte: Autoria Própria

De maneira geral, esse tipo de ataque pode se apresentar de diferentes maneiras, como por exemplo: spoofing de IP, e-mail, DNS, ARP, SMS. No caso do spoofing de IP o atacante faz uma requisição se passando por outro endereço de IP, assim o sistema para o qual um usuário é direcionado não consegue identificar corretamente o remetente. O spoofing de e-mail é o caso mais comum, quando um atacante envia um e-mail se passando por uma empresa ou pessoa conhecida da vítima. No *spoofing* de DNS o atacante faz uma manipulação da rede se passando por um site verdadeiro. Já no spoofing de ARP é uma técnica na qual um computador na rede local pode adulterar a tabela ARP de outro host, fazendo com que ele envie pacotes para o destino errado.

No tocante a ataques spoofing em WLANs, a natureza compartilhada do meio facilita que os atacante possam coletar informações úteis de identidade, durante o monitoramento passivo, posteriormente utilizar as informações de identidade, para iniciar ataques baseados em falsificação em redes sem fio. A Figura 1 ilustra uma idéia básica de um ataque spoofing em uma WLAN. Nesse caso o dispositivo falso passa a receber o sinal enviado pelo dispositivo verdadeiro, que posteriormente reenvia o sinal ao AP com pacotes modificados. Por outro lado, além dos ataques mais comuns de inundação de pacotes, os atacante podem fazer uso de *spoofing* de identidade, ou seja, falsificação de identidade para executar ataques de inundação mais sofisticados em APs, como por exemplo solicitação de autenticação e inundação de solicitação com ataques associados (BERNASCHI; FERRERI; VALCAMONICI, 2008).

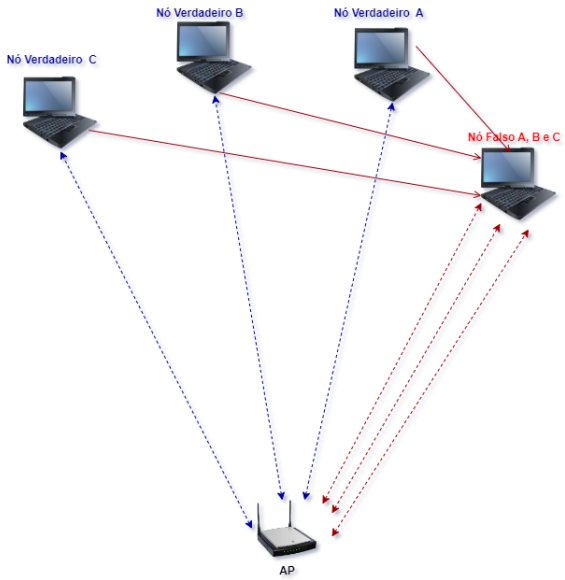
Um outro exemplo de *spoofing* em WLANs é quando um atacante pode lançar um ataque de autenticação depois que um cliente escolhe um AP para comunicação futura. Assim, o atacante deve autenticar-se ao AP antes que a sessão comece. Com isso, tanto o cliente quanto o AP são permitidos para solicitar de forma explícita a desautenticação. Dessa forma, um atacante pode falsificar esta mensagem de autenticação, seja em nome do cliente ou em nome do AP (BELLARDO; SAVAGE, 2003). Dessa maneira, o invasor pode persistentemente repetir este ataque e impedir completamente o cliente de transmitir ou receber dados. Outra possibilidade é quando um cliente entra na cobertura de um AP não autorizado (AP falso), onde a configuração de rede padrão fará com que o cliente se associe automaticamente com o AP não autorizado, que possui o sinal mais forte. Então, o invasor pode tomar ações para influenciar a comunicação (ARBAUGH et al., 2002).

Embora exista uma considerável variedade de ataques *spoofing*, nesse trabalho estaremos nos restrigindo aos ataques de camada física baseados nos endereços MAC dos dispositivos, tendo as métricas de RSS como base de informações para esse estudo.

2.4 Ataque Sybil

Para (YU et al., 2006), redes sem fio e mesmo outros sistemas descentralizados, são particularmente vulneráveis a ataques de sybil, pois um nó atacante sybil obtém várias identidades falsas e assim, finge ser vários nós verdadeiros no sistema. A partir disso, ao controlar um considerado número de nós do sistema, através de tarefas colaborativas o atacante ao se passar por nós verdadeiros, pode emitir informações errôneas ou maliciosas, influenciando assim no processamento das informações recebidas, desses nós atacantes. A Figura 2 ilustra um exemplo básico de ataque Sybil.

Figura 2 – Exemplo de Ataque Sybil



Fonte: Autoria Própria

Uma outra definição de ataque sybil é apresentada por (SUJATHA; ANITA, 2018). Ele apresenta o sybil como sendo nós com mais de uma identidade, que enganam o sistema e causam informações errôneas, imprudentes e podem até criar falsas demandas, como forma de afetar o funcionamento real. Estas réplicas falsificadas de identidades são conhecidas como nós Sybil e as atividades de tais identidades são chamadas de ataque Sybil. Na verdade, o nó compromete qualquer nó legítimo e executa um roubo de identidade, fazendo uso do sistema para várias ações desonestas (SUJATHA; ANITA, 2018).

Para (LIU; WU, 2019) pode-se definir o ataque sybil como um nó malicioso explorando identidades falsificadas. O ataque de Sybil pode reduzir significativamente a eficácia de esquemas tolerantes a falhas como armazenamento distribuído, dispersão e caminhos múltiplos, roteamento e manutenção de topologia. Réplicas, partições de armazenamento ou rotas que se acredita estarem usando nós disjuntos poderiam na

verdade estar usando um único adversário apresentando identidades múltiplas (KARLOF; WAGNER, 2003).

Segundo (SSU; WANG; CHANG, 2009) e (DOUCEUR, 2002), os ataques *sybil* são particularmente fáceis de lançar em uma rede sem fio, onde o meio de comunicação permite entrar e transmitir. Diante disso, à medida que a predominância das redes WLANs cresce, a necessidade de segurança de tornou-se uma preocupação latente, ainda mais quando se considera um ataque *sybil*.

Diante disso, tal ataque passou a ser mais estudado pela academia pela grande ocorrência no ambiente das WLANs, se tornando um problema grave que pode se apresentar em muitas áreas. Além disso, pode-se encontrar também, casos de disseminação de mensagens com múltiplas identidades, nesse caso, um nó da *sybil* prepara as decisões de votação na internet, assim também como interromper serviços de rede (DEMIRBAS; SONG, 2006). Segundo (LIU; WU, 2019), num ataque sybil, um atacante se utiliza de um dispositivo que se passa por uma única entidade maliciosa, reivindicando ilegalmente várias identidades para coletar mais pacotes, o que é extremamente danoso ao desempenho da rede.

2.5 Aprendizagem de Máquina

Para (LANTZ, 2013) uma definição formal de aprendizado de máquina é proposta pelo cientista da computação Tom M. Mitchell que afirma, *"que uma máquina aprende sempre que é capaz de utilizar sua experiência de modo que seu desempenho melhore em experiências semelhantes no futuro"*.

No contexto de aprendizagem de máquina, a clusterização (ANDERBERG, 1973); (DUBES; JAIN, 1979) e (KAUFMAN; ROUSSEEUW, 2009) é uma abordagem popular para implementar a operação de particionamento. Os métodos de clusterização dividem um conjunto de dados em n clusters, de maneira que os dados no mesmo cluster sejam mais similares uns aos outros (HUANG, 1998).

Dentro da ideia de aprendizado de máquina, podemos encontrar algoritmos não supervisionados como é o caso do *K-means* e *K-medoids*, e algoritmos supervisionados, como o caso do *K-NN*. Esse algoritmos costumam ser aplicados como forma de resolver problemas reais tendo acesso a base de dados que darão suporte ao processamento desses algoritmos. Para isso se aplicará métodos de aprendizado de máquina, um não supervisionado através do *k-means*.

2.5.1 Algoritmo K-Means

Um dos mais populares e eficientes métodos de agrupamento é o método *k-means* (HARTIGAN; WONG, 1979); (LLOYD, 1957); (MACQUEEN et al., 1967) que usa centróides para representar clusters otimizando a função de erro quadrático. O *k-means* é frequentemente usado para executar tarefas de aprendizado não supervisionadas (DING; HE, 2004). Assim, os métodos de análise de dados são essenciais para analisar a crescente quantidade de dados de alta dimensão.

Por um lado, a análise de agrupamentos (DUDA; HART; STORK, 2000); (FRIEDMAN; HASTIE; TIBSHIRANI, 2001); (JAIN; DUBES, 1988) tenta passar por dados rapidamente para obter conhecimento de primeira ordem, particionando pontos de dados em grupos separados, como pontos de dados pertencentes ao mesmo *cluster* são similares, enquanto pontos de dados pertencentes a diferentes *clusters* são diferentes (DING; HE, 2004).

Assim, o *K-means* é uma boa opção a ser aplicado num conjunto de objetos em bancos de dados com grupos ou clusters homogêneos (KLÖSGEN; ŻYTKOW, 1996), isso é uma operação fundamental na mineração de dados, tal operação pode ser útil em inúmeras tarefas, como classificação (não supervisionada), agregação e segmentação ou dissecção (CORMACK, 1971).

2.6 Considerações Finais

Num mundo cada vez mais interligado, o qual oferece inúmeras aplicações inteligentes baseadas na arquitetura WiFi, o controle de acesso seguro aos dados de usuários é uma variável, que precisa de considerada atenção. Como forma de se minimizar ataques de identidades nestas redes, se faz necessários, estudos aprofundados de ambientes comuns, os quais estão acessíveis e democratizados no cotidiano do usuário final. Estes visam planejar um método através de aprendizado de máquina. Assim permite que os ataques de identidade na rede, se utilize da métrica de RSS, que se constitui numa abordagem simples e interessante. Entretanto, é prudente considerar a importância da aplicação de métodos de aprendizado de máquina, nos processos para detecção de nós falso/maliciosos, onde tais métodos auxiliarão como forma, de identificar a presença ou não de um nó malicioso na rede WiFi.

3 TRABALHOS RELACIONADOS

As redes sem fio podem ser vulneráveis a ataques de identidade. Assim, algumas técnicas de ataques podem ser usadas com considerado sucesso, dentre eles encontramos ataques *spoofing* e *sybil*. Diante disso, no atual contexto, alguns trabalhos merecem citação no tocante a sua contribuição para oferecer um ambiente seguro em redes WLAN, reduzindo de forma significativa as vulnerabilidades dessas redes. Abordagens clássicas centradas no RSS, visam minimizar tentativas de ataques de identidades as redes, essas já apresentam estratégias interessantes e bem relevantes, a quais servem como referencial desta tese.

Assim, o foco desse trabalho está em localizar a presença de um nó atacante (malicioso), através da utilização do RSS em redes WiFi com aplicação de técnicas de *machine learning*. Para isso, nas próximas sessões apresentamos algumas propostas já encontradas na literatura, mas que apresentam dificuldades relacionadas à localização do nó falso/malicioso quando este está a uma mesma distância do(s) APs que o nó legítimo.

3.1 Detectando Ataques de Identidade em Redes Sem Fio

Na proposta apresentada por (CHEN et al., 2010), um método para detectar ataques de *spoofing e Sybil* em redes sem fio foi proposto. A estratégia é baseada na correlação espacial do RSS. O método emprega aprendizado de máquina sem supervisão, usando o algoritmo *k-means*, para calcular uma métrica de comparação e aplicar um teste estatístico para detectar ataques baseados em identidade. Nessa proposta os autores evitam usar técnicas de criptografia, pois argumentam que nem sempre é possível por exigir recursos de gerenciamento de chaves, sobrecarregando assim a infraestrutura em questão. Na proposta é apresentado um método para detectar ambos ataques, *spoofing* e *sybil* usando o mesmo conjunto de técnicas.

Num primeiro momento é proposto um modelo de detecção de ataques que utiliza a correlação espacial da intensidade do sinal recebido. Além disso, fornecem uma análise teórica da abordagem através das estatísticas de teste para a detecção de ataques baseados em identidade usando o algoritmo *K-means*. O detector de ataque se apresenta robusto, quando lida com situações de invasores, que usam diferentes níveis de potência para atacar o esquema de detecção. Ainda descrevem como ocorre a integração do detector com o ambiente de ataque num sistema de localização de tempo real, tal sistema também pode localizar as posições de atacantes, usando algoritmos de localização.

O ambiente de experimentação ocorre em dois escritórios reais usando uma rede IEEE802.11 (WiFi) e uma rede IEEE802.15.4 (ZigBee). Com isso, os resultados mostram

que é possível detectar ataques em redes sem fio, baseados em identidade com alto percentual de detecção e uma baixa taxa de falso-positivo, fornecendo assim forte evidência da eficácia do detector de ataque utilizando a técnica de correlação espacial de RSS. A análise teórica da correlação espacial do RSS considera um modelo de propagação que inclui apenas efeitos sombreamento e perda de percurso. Portanto, seu modelo de canal não considera os efeitos de desvanecimento rápido ou pequena escala. Os autores não exploram como as amostras instantâneas de RSS corrompido pelo desvanecimento de pequena escala afeta o processo de detecção. Além disso, a estratégia é sensível ao limiar de decisão e não funciona bem quando o nó legítimo e o nó de ataque estão na mesma distância, ou em uma distância muito próxima um do outro em relação ao *landmark*.

3.2 Mútiplos Ataques *Spoofing*

Por outro lado, o problema de detectar vários ataques de *spoofing* foi investigado em (YANG et al., 2012). A estratégia também emprega o correlação do RSS para detectar ataques de *spoofing*, associada a cada nó, o que segundo o modelo, torna difícil de falsificar e assim, não depender de criptografia, como base para: 1) detectar ataques de falsificação; 2) determinar o número de invasores quando vários adversários se disfarçam como a mesma identidade de nó; e 3) localizar múltiplos adversários. Posteriormente, formulam-se o problema de determinar o número de invasores como um problema de detecção de *multiclasse*. Mecanismos baseados em *cluster* são desenvolvidos para determinar o número de invasores. Quando os dados de treinamento estão disponíveis, explora-se o uso do método SVM (*Support Vector Machines*) para melhorar ainda mais a precisão da determinação do número de invasores. Além disso, empregam um sistema integrado de detecção e localização que pode localizar as posições de vários invasores. Segundo os autores, os resultados experimentais mostram que os métodos propostos podem alcançar mais de 90% de taxa de acerto e precisão ao determinar o número de invasores.

3.3 Autenticação Baseada em Canal

A estratégia proposta em (PEI et al., 2014) adota uma abordagem alternativa onde o receptor pode identificar e autenticar os remetentes através de vetores de canais estimados. O processo de autenticação é formulado como uma sequência de problemas de teste de hipótese. A fim de melhorar a probabilidade de detecção e reduzir a probabilidade de falso alarme, dois esquemas são propostos com base em diferentes algoritmos de classificação em aprendizado de máquina.

Especificamente, os esquemas de autenticação baseados no SVM e o esquema de autenticação baseado na análise discriminante de *Fisher* (Linear Fischer Discriminant Analysis - LFDA) são propostos, explorando três características de canal, incluindo

tempo de chegada, intensidade do sinal recebido e características cíclicas dos canais. Nos esquemas baseados em SVM, os SVMs lineares e não-lineares são usados para gerar classificadores para resolver os problemas do teste de hipótese. No esquema baseado em LFDA, uma combinação linear desses três recursos de canal é usada como a estatística de teste, que é comparada com um limite para executar a autenticação. Os resultados da simulação demonstram que os esquemas propostos apresentam um desempenho melhor em termos de probabilidade de erro de detecção e probabilidade de alarme falso do que vários esquemas típicos de autenticação baseados em canal existentes. Além disso, a complexidade dos esquemas propostos são analisados, e o esquema baseado no LFDA tem o melhor desempenho.

A principal desvantagem desta estratégia é que requer a estimativa da informação do estado do canal (CSI), que é uma tarefa mais complexa a ser realizada pela rede e requer acesso a camada física especifica como parâmetros do rádio, que geralmente não são acessíveis em módulos de rádio comerciais.

3.4 Detecção e Classificação de Ataques com Variações Recíprocas de RSS em Redes Móveis Sem Fio

Para (TANG et al., 2020), os ataques baseados em identidade (IBAs) são ameaças mais sérias às redes sem fio. Diante disso que, há um interesse crescente em usar o RSS para detectar IBAs em redes sem fio. No entanto, os esquemas atuais tendem a gerar falsos alarmes excessivos no cenário móvel. Considerando isso que, os autores propõem um esquema de identificação e classificação baseado em variação recíproca de canal ou RCVIC (Reciprocal Channel Variation-based Identification and Classification), que explora a reciprocidade do canal de desvanecimento sem fio e as variações de RSS naturalmente incorridas pela mobilidade, visando melhorar o desempenho da detecção. Diferente dos esquemas atuais que apenas detectam IBAs, o esquema RCVIC conduz processos de detecção em vários estágios. Segundo os autores, se os IBAs forem detectados, o esquema RCVIC divide os quadros recebidos em duas classes. Os quadros da mesma classe devem ser enviados dos mesmos remetentes, o que pode beneficiar as análises posteriores, como forense de rede, localização de invasores e análise de trajetória.

A viabilidade do RCVIC é avaliada numericamente por meio de análises teóricas e simulações. Ele é validado ainda mais por meio de experimentos usando dispositivos 802.11 prontos para uso sob diferentes padrões de ataque em cenários móveis internos e externos reais. O método proposto avalia sua abordagem de forma simulada e não aplica as técnicas em redes WiFi comuns para a detecção do nó falso no ambiente de análise. O foco da abordagem é em ambientes de redes sem fio veiculares com auxílio do CSI.

3.5 Estudo de Viabilidade da Estimativa Prática de AoA Usando CSI em Dispositivos WLAN Comerciais

Uma outra abordagem mais recente que trabalha a questão da segurança de dados em redes WiFi foi proposta em (FUKUSHIMA et al., 2022). O foco são as aplicações internas avançadas de IoT e sensores agregados aos serviços. Esta proposta aborda o problema da localização interna baseada em WLAN usando (CSI) além das informações existentes fornecidas pelo RSS. Nesse trabalho os autores propuseram um método prático para estimar o AoA para resolver quatro problemas: 1) CSI comprimido, que não pode ser usado para estimar o AoA diretamente, 2) o cabo de antena, em que a fase muda dependendo do comprimento da linha fixa, 3) o espaçamento das antenas, em que a distância entre as antenas restringe a estimativa de AoA, e 4) a individualidade da antena, em que as antenas usadas na comunicação MIMO real têm características diferentes. Segundo os autores, os resultados indicam que o método proposto pode estimar AoA com um erro médio de 9,1° e reduzir o erro de estimativa em 85,4% em comparação com uma abordagem direta. Mas vale ressaltar que o método não utiliza *machine learning*, nem a detecção de um possível nó falso na rede WiFi.

3.6 NISA: Node Identification and Spoofing Attack Detection Based on Clock Features and Radio Information for WSNs

O trabalho proposto por (HUAN; KIM; ZHANG, 2021) apresenta uma abordagem para identificação de nós com base em recursos exclusivos de hardware, como desvios de clock. Tem sido considerada uma técnica eficiente em redes de sensores sem fio WSNs. Os ataques de falsificação que imitam recursos exclusivos de hardware, no entanto, podem prejudicar significativamente ou quebrar a identificação de nó convencional baseada em distorção de relógio devido à exposição de informações de relógio por meio de transmissão. Para se defender contra ataques de *spoofing*, os autores propõem um novo esquema de identificação de nós chamado de identificação de nós contra ataques de *spoofing* (NISA). Ele utiliza a estrutura de sincronização de tempo reverso, onde as distorções do relógio dos nós sensores são estimadas nas camadas superiores numa RSSF e as informações de link de rádio espacialmente correlacionadas para obter a identificação simultânea do nó e a detecção de ataques. Além disso, a proposta mostra o NISA centralizado e distribuído para cobrir cenários de salto único e vários saltos, o primeiro dos quais emprega uma rede neural convolucional de entrada única e saída múltipla. Com um *testbed* de RSSF real consistindo de nós sensores TelosB rodando TinyOS, nessa proposta foi investigado a identificabilidade de distorções de *clock* sob variações de temperatura e tensão e avaliado o desempenho de NISA centralizado e distribuído. Para (HUAN; KIM; ZHANG, 2021) os

resultados experimentais demonstram que tanto o NISA centralizado quanto o distribuído podem fornecer identificação precisa de nós e detecção de ataque de *spoofing*. Embora o trabalho explore as informações de RSS, a abordagem não aplica técnicas de *machine learning* para a identificação de nós falsos na rede, em ambientes reais de redes WiFi.

3.7 Algoritmo de Detecção do Ataque de Mimetismo baseado no Auto-Encoder Variacional

O trabalho de (WANG et al., 2021) apresenta um proposta que explora o ataque de mimetismo, o qual é um ataque de fraude no qual um atacante pode forjar um usuário legítimo imitando o endereço de controle de acesso à mídia de um usuário legítimo ou outras credenciais de identidade. Esta pesquisa apresenta um modelo não supervisionado, denominado DAMA, para detectar ataques de mimetismo usando a série temporal do RSS para detectar a mudança de localização do dispositivo. O RSS de um dispositivo está relacionado à sua localização, por isso, segundo os autores, é difícil forjá-lo. Segundo os autores da proposta, os resultados apresentados nos experimentos, demonstram que tal método retorna considerada melhora em relação, as referências que serviram de base para a proposta.

3.8 Autenticação de Camada Física em Comunicações Sem Fio

No caso da proposta apresentada por (XIE; LI; TAN, 2020), a autenticação é uma questão importante nas comunicações sem fio, pois a natureza aberta do meio sem fio fornece mais vulnerabilidades de segurança. Assim, a Autenticação de Camada Física (Physical Layer Authentication - PLA) atrai muitos interesses de pesquisa porque fornece segurança da teoria da informação e baixa complexidade. O autor traz um levantamento detalhado das funcionalidades e técnicas que podem ser utilizadas no PLA. Eles categorizam os esquemas de PLA existentes em duas categorias: esquemas passivos e ativos. Nos esquemas passivos, um receptor autentica o transmissor com base nas características da camada física dos sinais recebidos. Dividimos ainda os esquemas passivos em duas subcategorias: recursos baseados em dispositivo e recursos baseados em canal. Nos esquemas ativos, um transmissor gera uma tag com base em uma chave secreta e a incorpora em uma mensagem de origem. Por conseguinte, um receptor autentica o transmissor com base na tag existente no sinal recebido. Além disso, se divide os esquemas ativos em duas subcategorias: esquemas com abrangência e esquemas sem abrangências.

3.9 Uma nova fronteira para a segurança de IoT emergindo de três décadas de geração de chaves que dependem de canais sem fio

O trabalho proposto em (ZHANG et al., 2020) faz uma abordagem de *IoT*, que está revolucionando nossa vida cotidiana ao conectar tudo e todos. O grande número de dispositivos são preferencialmente conectados a rede de maneira sem fio, devido à facilidade de instalação e implantação flexível. No entanto, a natureza de transmissão do meio sem fio torna as informações acessíveis a todos, incluindo usuários mal-intencionados, e devem ser protegidos por criptografia. Infelizmente, o fornecimento seguro e eficiente de chaves criptográficas para dispositivos IoT de baixo custo é um desafio. Chaves fracas resultaram em graves violações de segurança, como evidenciado por vários ataques cibernéticos notórios. Diante disso, o autor deste fornece um levantamento abrangente de soluções de segurança leves concebidas para IoT, contando com geração de chaves a partir de canais sem fio. Primeiro, apresenta os fundamentos e protocolos de geração de chaves. Em seguida, examina como aplicar essa técnica emergente para proteger a IoT e demonstra que a geração de chaves que depende da aleatoriedade dos canais sem fio é adequada para a IoT. Com isso, o trabalho faz uma revisão dos extensos esforços de pesquisa nas áreas de modelagem teórica, validação baseada em simulação e exploração experimental. Finalmente discute os obstáculos e desafios que a geração de chaves está enfrentando. Também sugere trabalhos futuros para tornar a geração de chaves uma solução confiável e segura para proteger a IoT, assim como a validação baseada em simulação e exploração experimental.

3.10 Um estudo teórico de informações mais gerais do sistema de verificação de localização sem fio

O estudo apresentado em (GHOLAMI; HODTANI, 2020) trata de uma abordagem generalista e teórica, o qual cita que nos últimos anos, a questão do sistema de verificação de localização (LVS) tem recebido significativa atenção em redes de comunicação sem fio, principalmente em Sistemas Inteligentes de Transporte (ITS) e tecnologia veicular, no quais as informações de localização são importantes para a segurança e proteção dos usuários. Com isso, nesta proposta, foi considerada uma rede veicular e analisado um estudo mais geral da teoria da informação de LVS, especificamente, visando encontrar um limite de decisão ótimo para detectar um local falsificado e aumentar a capacidade do sistema em detectar corretamente usuários maliciosos. Para isso, foi proposto diferentes medidas teóricas de informação (divergência de *Renyi*, informação mútua de Renyi, divergência de *Kullback-Leibler*, informação mútua de *Kullback-Leibler* e divergência de *Jensen-Shannon*) para identificação de usuários maliciosos.

3.11 Principais Características das Soluções

A Tabela 1 apresenta algumas características dos principais trabalhos relacionados discutidos nas seções anteriores.

Proposta	**RSS**	**Spoofing**	**CSV**	**CSI**	**MAC**	**IoT**
3.1	X	X		X	X	X
3.2	X	X	X			
3.3	X		X		X	
3.4	X			X		
3.5	X			X	X	
3.6	X	X				
3.7	X	X			X	
3.8	X	X			X	
3.9	X					X
3.10	X					

Tabela 1 – Resumo de Características

3.12 Considerações Finais

Neste capítulo foram analisadas as características, vantagens e desvantagens de soluções propostas na literatura que mais se aproximam da proposta deste trabalho. Vale ressaltar que os trabalhos que serviram como base para essa propostas, foram citados pela proposta inicial de (Chen et al., 2010) e/ou as propostas publicadas por um de seus autores que procuram realizar uma abordagem muito próxima, como é o caso de (YANG et al., 2012), que explora uma abordagem com aprendizado de máquinas, mas considerando ambiente simulado e não ambiente real. Vale ressaltar que a proposta desta tese visa contribuir para detecção de ataques *spoofing* em ambientes de redes WiFi comuns.

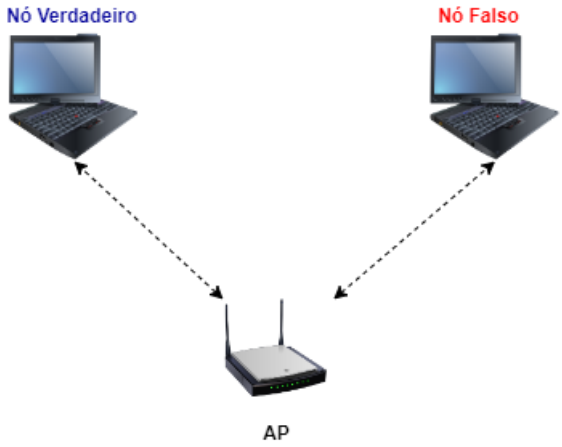
4 MODELO PROPOSTO

Este capítulo apresenta inicialmente a estratégia proposta para análise e detecção da presença de um nó atacante, usando medidas de RSS obtidas pelo AP e técnicas de aprendizado de máquina. Este modelo é avaliado em ambiente simulado, usando o software Matlab, para três cenários distintos, considerando apenas um AP, um nó verdadeiro e um nó falso. No primeiro cenário os valores de RSS são analisados considerado-se apenas a presença do nó verdadeiro. No segundo e terceiro cenários consideramos a presença do nó verdadeiro e do nó falso.

4.1 Descrição da Proposta

Como forma de apresentar a ideia inicial da proposta, consideramos um cenário de WLAN num ambiente *indoor* com três elementos básicos, sendo um nó verdadeiro, um nó falso (atacante) e um ponto de acesso (AP). Como a estratégia de detecção se baseia nos dados coletados pelo AP, ele também é denominado de ponto de referência ou *landmark*. Este cenário é ilustrado na Figura 3. Embora na WLAN podemos ter vários clientes conectados ao AP, para prova de conceito assumimos que apenas um dos clientes (nó verdadeiro/legítimo) sofrerá o ataque de *spoofing*.

Figura 3 – Cenário Conceitual de Ataque



Autoria Própria

Como o método proposto se baseia em informações da camada física, mais especificamente de medidas de RSS, implementamos no software *Matlab* um cenário de comunicação assumindo uma modelagem de canal sem fio mais realista. Embora diversas estratégias de detecção encontradas na literatura utilizem medidas de RSS, o método

proposto se baseia no princípio de explorar o efeito do sombreamento ou *shadowing* causado pelo canal de rádio. Este efeito pode ajudar a identificar a presença de um nó atacante, mesmo que ele esteja na mesma distância do AP que está o nó legítimo.

A partir da coleta de um conjunto de medidas de RSS, provenientes do nó legítimo e do nó atacante, a ideia inicial é utilizar uma estratégia de aprendizagem de máquina não supervisionada para identificar padrões de comportamento no conjunto de medidas, analisadas em janelas de amostragem, que possam ser usadas para determinar a presença ou não de um nó atacante. Uma primeira abordagem seria o uso de técnicas de clusterização, como por exemplo o algoritmo *k-means*, para identificar possíveis *clusters* de valores de RSS que caracterizem a presença ou não do atacante. O algoritmo *k-means* é um método usado para particionar automaticamente um conjunto de dados em k *clusters* ou grupos. No cenário inicial que descreve os princípios da proposta, assumimos um único *landmark* e um ambiente homogêneo, com o mesmo expoente de perda de percurso e distribuições de sombreamento, tanto para o nó legítimo quanto para o nó atacante. Antes de descrever em detalhes a estratégia proposta, apresentamos na Seção 4.1.1 uma descrição do problema a partir de uma estratégia importante proposta na literatura (CHEN et al., 2010) para identificar ataques usando *RSS* e que serviu como base principal para nosso estudo.

4.1.1 Descrição do Problema

Em (CHEN et al., 2010) foi proposta uma estratégia de detecção usando o *k-means* para agrupar as amostras de potência recebida e usar a distância dos centroides como uma métrica para identificar ataques. Na prática, as amostras instantâneas de *RSS* estão sujeitas aos efeitos de desvanecimento rápido e sombreamento, além da perda de percurso. Em (CHEN et al., 2010) os autores modelam matematicamente este problema de detecção. Considere como d_1 a distância entre o nó verdadeiro e o *landmark* e como d_2 a distância entre o nó atacante e o *landmark*. A diferença entre as potências recebidas é modelada como uma variável aleatória ΔP,

$$\Delta P = Pr(d_1)[dBm] - Pr(d_2)[dBm] \tag{4.1}$$

que segue uma distribuição normal com média

$$\mu = -10\ \alpha_1\ \log_{10}(d_1/d_0) + 10\ \alpha_2\ \log_{10}(d_2/d_0), \tag{4.2}$$

e variância

$$\sigma^2 = \sigma_1^2 + \sigma_2^2. \tag{4.3}$$

Os parâmetros α_1 e α_2 são os expoentes de perda de percurso como percebidos no ponto de referência (landmark) para o nó legítimo e atacante, e σ_1^2 e σ_2^2 são as variâncias do sombreamento.

Além disso, os efeitos de desvanecimento lento causados pelo sombreamento dependem de modificações no ambiente que circundam o transmissor e o receptor. Este efeito é geralmente modelado por uma variável aleatória com distribuição *log-normal* com média zero, cujo desvio padrão depende do ambiente de propagação. Em ambientes urbanos, o desvio padrão é tipicamente de 4 dB a 10 dB. Na prática as amostras de potência *(RSS)* também estão sujeitas aos efeitos do desvanecimento rápido, que não é incluído na modelagem de canal de (CHEN et al., 2010).

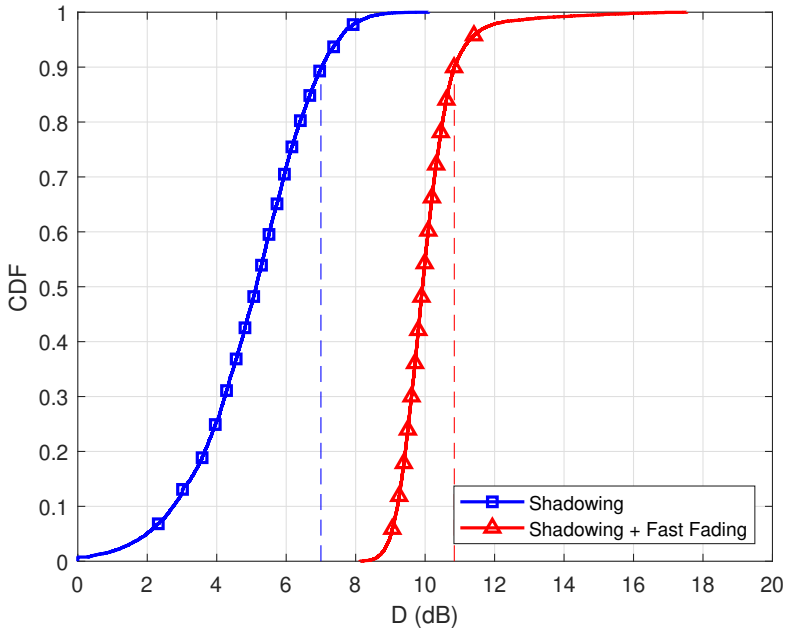
Quando modelamos o enlace de rádio, uma variável aleatória com distribuição normal pode ser gerada de tempo em tempo para representar as variações do sombreamento no enlace de rádio, causado por alterações do ambiente. A forma adequada de simulação é gerar amostras da variável aleatória de forma independente, para cada janela diferente de tempo, e para cada enlace de rádio diferente. Na prática, o sombreamento apresenta uma correlação espacial e temporal, que não foi levada em conta na modelagem de (CHEN et al., 2010). Particularmente, a correlação temporal do fenômeno de sombreamento pode impor um intervalo de tempo onde o nível de atenuação imposto pelo sombreamento pode ser assumido constante (GOLDSMITH, 2005). De intervalos em intervalos de tempo a atenuação causado pelo sombreamento pode alterar devido a modificações no ambiente. Isso implica que a precisão do processo de decisão empregado em (CHEN et al., 2010) pode ser afetado pela correlação do sombreamento, pois o limiar ótimo de decisão deveria ser modificado. De maneira a ilustrar este comportamento, avaliamos como a estratégia original proposta em (CHEN et al., 2010) é afetada pelo modelo do canal. Apresentamos nas Figuras 4 e 5 como a função distribuição acumulada de probabilidade (CDF) para a distância (D) entre os centroides é afetada pelo modelo do canal. O cenário de simulação considera tanto o nó legítimo quanto o atacante localizados na mesma distância $d_1 = d_2 = 50m$ do *landmark* de referência e um tamanho da janela de análise com 50 amostras de RSS. Assumimos duas situações de variância de sombreamento, $\sigma = \sigma_1^2 = \sigma_2^2 = 4$ dB e 8 dB. Também consideramos para fins de análise um expoente de perda de percurso $\alpha = \alpha_1 = \alpha_2 = 2.5$ e parâmetro de desvanecimento de Nakagami-m, com $m = 1$.

Assumimos um modelo de canal com desvanecimento de bloco (*block fading*), onde o receptor estima o valor de RSS a cada pacote recebido. Tipicamente, o desvanecimento (atenuação) de sinal causado pelo sombreamento apresenta uma variação lenta ao longo do tempo quando comparado ao *block fading*. Assumimos que a atenuação causada pelo sombreamento se altera em média a cada 25 amostras de RSS, sendo um total de 50 amostras. Quando o canal é modelado apenas com a atenuação do sombreamento, a distribuição da distância entre os centroides possui uma média menor do que no caso onde o efeito do *block fading* está também incluído na modelagem do canal.

A linha tracejada nas Figuras 4 e 5 indica o valor do limiar (threshold) para uma probabilidade de 90% na curva da CDF. A implicação prática disso é que se utilizamos

um limiar menor que o necessário devido ao modelo simplificado do canal, podemos gerar um número excessivo de falsos positivos. Em contraste, se utilizamos um valor de limiar muito alto, não seremos capazes de detectar muitos ataques de *spoofing* quando o nó legítimo e o nó atacante estiverem a uma distância muito similar com relação ao *landmark*, gerando um grande número de falsos negativos. Mas, para melhor entendimento é necessário compreender a necessidade da existência do D como sendo: A medidas de distância entre dois *clusters* para uso em algoritmos hierárquicos. (a) *single link*, onde a distância entre os *clusters* é dada pela distância entre seus pontos mais próximos. (b) *average link*, onde a distância entre os *clusters* é dada pela distância entre seus centróides e (c) complete link, onde a distância entre os clusters é dada pela distância entre seus pontos mais distantes (LINDEN, 2009).

Figura 4 – CDF da distância (D) entre os Centroides ($\sigma = \sigma_1^2 = \sigma_2^2 = 4$ dB).



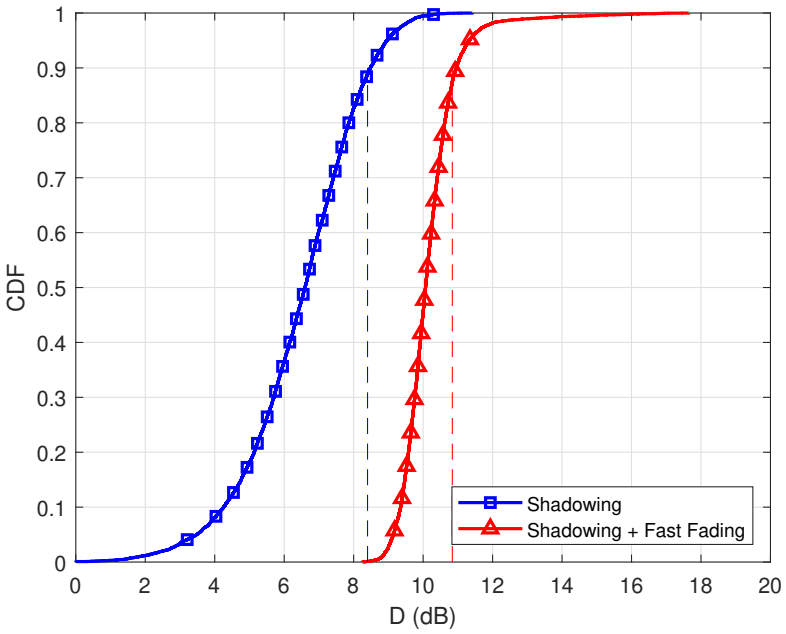
Autoria Própria

Em (CHEN et al., 2010), os autores obtiveram o limiar ótimo para um modelo de canal teórico baseado apenas em sombreamento, que é dado por

$$\tau = 5\ \alpha\ \log_{10}\left(\frac{d_2}{d_1}\right) \tag{4.4}$$

De acordo com os resultados apresentados em (CHEN et al., 2010), a precisão é maior que 90% quando $d_1/d_2 = 0{,}5$, onde d_1 é a distância do ponto de referência ao nó mais próximo. No entanto, quando o dois nós estão exatamente à mesma distância do ponto de referência ($d_1 = d_2$), a precisão é de 50% e o limiar de decisão dada pela equação (4.4) é $\tau = 0$. Na prática, o limiar de decisão deve ser ajustado para minimizar a taxa de falsos positivos (False Positive Rate - FPR) e maximizar a taxa de detecção (Detection Rate - DR).

Figura 5 – CDF da distância (D) entre os Centroides ($\sigma = \sigma_1^2 = \sigma_2^2 = 8$ dB).



Autoria Própria

Portanto, o principal desafio deste trabalho é a detecção de ataques quando os dois nós estão em distâncias muito semelhantes em relação ao ponto de referência (landmark). Durante um ataque de *spoofing*, um nó potencialmente posicionado em qualquer lugar da região considerada, reivindica a mesma identidade de um nó legítimo. Portanto, as medidas de RSS coletadas pelo *landmark* em uma janela de análise, contêm valores provenientes tanto das transmissões do nó legítimo, quanto dos nós atacantes.

Para melhor entendimento da proposta de (CHEN et al., 2010) e dos efeitos do canal, definimos e avaliamos três cenários *spoofing*. No Cenário 1 (sem ataque), ilustrado na Figura 6, temos apenas a presença do nó legítimo, que está transmitindo a uma distância d_1 do *landmark*. No Cenário 2 (com ataque) apresentado na Figura 7 temos o nó legítimo e o nó atacante transmitindo, e eles estão localizados na mesma distância do *landmark*, $d_1 = d_2$. No Cenário 3 (com ataque), mostrado na Figura 8, temos o nós legítimo e atacante, também localizados na mesma distância do *landmark*, $d_1 = d_2$, porém muito próximos entre si.

Na Figura 8 quando os nós legítimos e atacantes estão próximos um do outro (Cenário 3), nos deparamos com duas situações distintas. As vezes o atacante está transmitindo e deve ser detectado pelo o *landmark*. No entanto, em outros momentos apenas o nó legítimo está transmitindo. Se o sistema de detecção baseado no limiar da distância entre os centroides estão habilitados, as conclusões podem ser contraditórias, dependendo do valor do limiar de decisão (τ). Por exemplo, considere o Cenário 1 com d_1= 50m e duas

Figura 6 – Cenário 1 - $d_1 = 50m$

Autoria Própria

Figura 7 – Cenário 2 - $d_1 = d_2 = 50m$

Autoria Própria

configurações do Cenário 3, onde os nós legítimos e atacantes estão em distâncias $d_1 =$ 50m, $d_2 =$ 60m e $d_1 =$ 50m, $d_2=$ 70m.

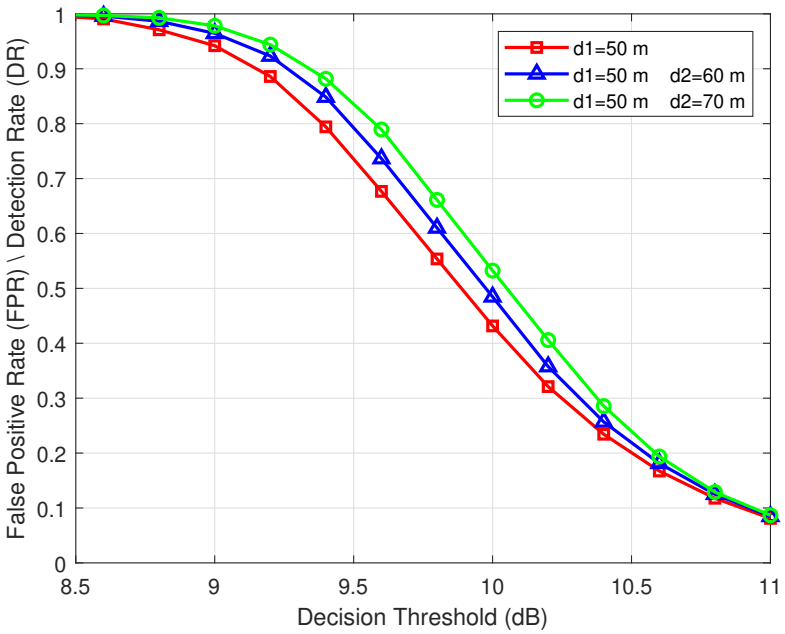
Aplicamos a estratégia de detecção proposta em (CHEN et al., 2010), usando os parâmetros do canal definidos anteriormente. As Figuras 9 e 10 apresentam uma análise de como a taxa de *falsos positivos* (FPR) e de detecção (DR) se relacionam com o limiar de decisão (τ), para uma variância do sombreamento $\sigma_1^2 = \sigma_2^2 = 4$ dB e $\sigma_1^2 = \sigma_2^2 = 8$ dB, respectivamente. Quando apenas o nó legítimo está transmitindo, um limiar de decisão ($\tau \leq 9$ dB) resulta em uma taxa de falsos positivos acima de 90%. Em contraste. Quando

Figura 8 – Cenário 3 - $d_1 = d_2 = 50m$

Autoria Própria

ambos os nós, legítimo e atacante estão transmitindo, um limiar de decisão mais alto ($\tau \leq 11$ dB) resulta em uma taxa de detecção muito baixa, com mais de 90% de falsos negativos.

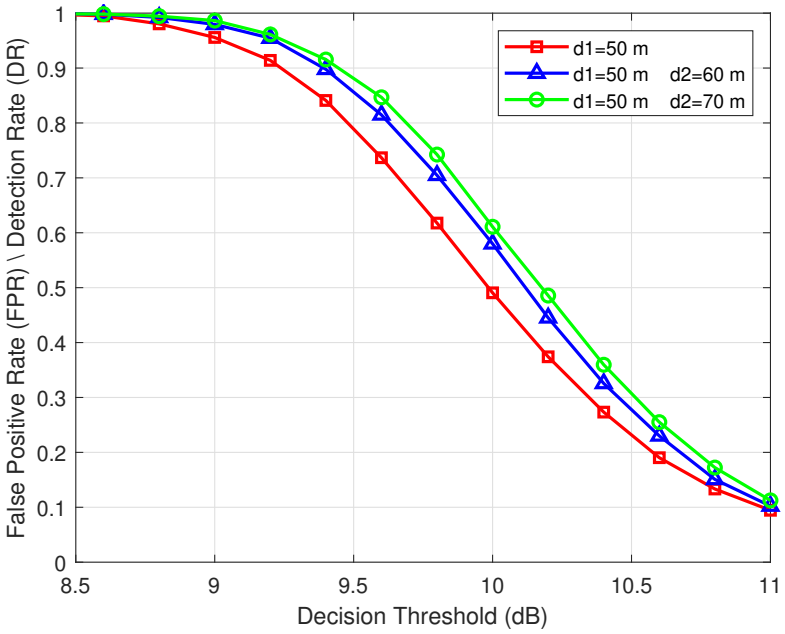
Figura 9 – Análise dos Cenários - $\sigma_1^2 = \sigma_2^2 = 4$ dB



Fonte: Autoria Própria

Com isso, a detecção de ataques de *spoofing* usando o método de (CHEN et al., 2010) não é efetivo quando os dois nós estão a uma distância similar do *landmark*, o

Figura 10 – Análise dos Cenários - - $\sigma_1^2 = \sigma_2^2 = 8$ dB



Fonte: Autoria Própria

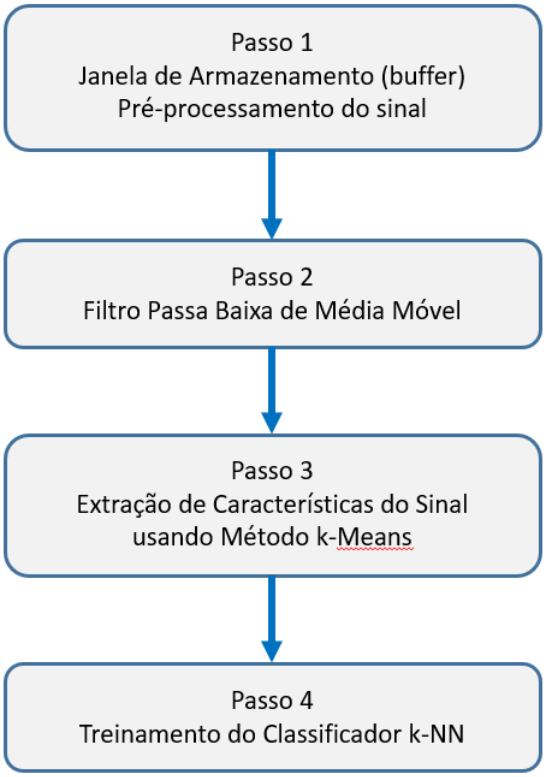
que indica que outras métricas e características de comportamento do canal devem ser investigadas, a fim de melhorar o sistema de detecção. Assim, a intuição por trás dessa proposta é que a atenuação imposta pelo sombreamento muda ao longo do tempo devido as mudanças físicas no ambiente de propagação do sinal. Com base em evidências de medidas reais de RSS, podemos assumir que a atenuação de sombreamento é quase constante (altamente correlacionada no tempo) durante certos períodos. Estes períodos podem se modificar de tempos em tempos devido à alterações topológicas no ambiente. Em ambientes internos (indoor), essas mudanças podem ser causadas, por exemplo, pela abertura ou fechamento de portas e janelas, reposicionamento de objetos e movimentação de pessoas. Em ambientes externos (outdoor), a mobilidade de veículos em estacionamento e ruas, são exemplos de fontes que causam mudanças na atenuação de sombreamento.

4.2 Método Proposto

A proposta inicial do sistema de detecção possui quatro fases, apresentadas na Figura 11. Na Etapa 1, as amostras de RSS coletadas no ponto de referência são armazenadas em um *buffer* e agrupadas em l sub-janelas não sobrepostas, $w = \{w_0, w_1, \ldots, w_{l-1}\}$, com n amostras cada, onde $w_j = \{r_0, r_1, \ldots, r_{n-1}\}$. Estas amostras são filtradas na Etapa 2 usando um filtro de média móvel para reduzir a variabilidade dos valores de RSS, causadas pelo desvanecimento rápido do canal. Na Etapa 3 empregamos o algoritmo de agrupamento k-means (HASTIE et al., 2005) sobre o conjunto w_j e usamos a distância euclidiana

$D_j = \|c_1 - c_2\|$ como medida de dissimilaridade para determinar a distância entre dois centróides (c_1 e c_2), seguindo a abordagem empregada em (CHEN et al., 2010). Assumimos a partição dos n pontos de amostra em w_j em dois subconjuntos com centróides c_1 e c_2. Este procedimento é aplicado em b conjuntos consecutivos de amostras $w_j, w_{j+1}, \ldots, w_{j+b}$, o que resulta em um conjunto de distâncias $\mathcal{D}_c = \{D_j, D_{j+1}, \ldots, D_{j+b-1}\}$ com cardinalidade $|\mathcal{D}_c| = b$. A média e variância de $\mathcal{D}_c$ são denotados por μ_c and σ_c, respectivamente. Essas estatísticas são as características empregadas na Etapa 4 do método proposto, onde aplicamos o classificador supervisionado k-NN (BOEHMKE; GREENWELL, 2019) e (LANTZ, 2013) para detectar o ataque. Essas quatro etapas são executadas continuamente no ponto de referência usando a janela atual de amostras de RSS.

Figura 11 – Arquitetura Proposta



Fonte: Autoria Própria

4.2.1 Estratégia de Detecção de Ataque

Nessa etapa do sistema de detecção é incluído um classificador k-NN supervisionado. Baseado em uma análise preliminar para determinar as características mais significativas a serem usadas pelo algoritmo k-NN, selecionou-se o valor mínimo e o valor máximo de μ_c. Para justificar estas escolhas, considere como exemplo o Cenário 1 (Figura 6) com $d_1 =$ 50 m e o Cenário 2 (Figura 7) com $d_1 = d_2 = 50$ m. Considere w_j com tamanho $n = 50$,

conjuntos de $b = 6$ amostras consecutivas e os mesmos parâmetros de canal. Também foi empregado a correlação espacial e temporal do sombreamento, conforme definido na Seção 2.2.1.

Obviamente, devido à natureza probabilística do canal, podemos observar na Figura 12 que todas as métricas apresentam desvios significativos de uma janela de análise para outra. É importante salientar que os valores de μ_c podem mudar significativamente. Isso é causado principalmente pela modificação de sombreamento dentro da janela de análise, que pode ocorrer para um ou ambos os nós (legítimo/atacante). A variação do sombreamento pode acentuar a diferença de distância entre clusters, dando uma melhor indicação da presença de um nó malicioso que está gerando um ataque de **spoofing**. Um comportamento semelhante foi observado para o Cenário 3, figura 8 com $d_1 = 50$ m e $d_2 = 60$ m, cujos resultados foram omitidos aqui por uma questão de brevidade.

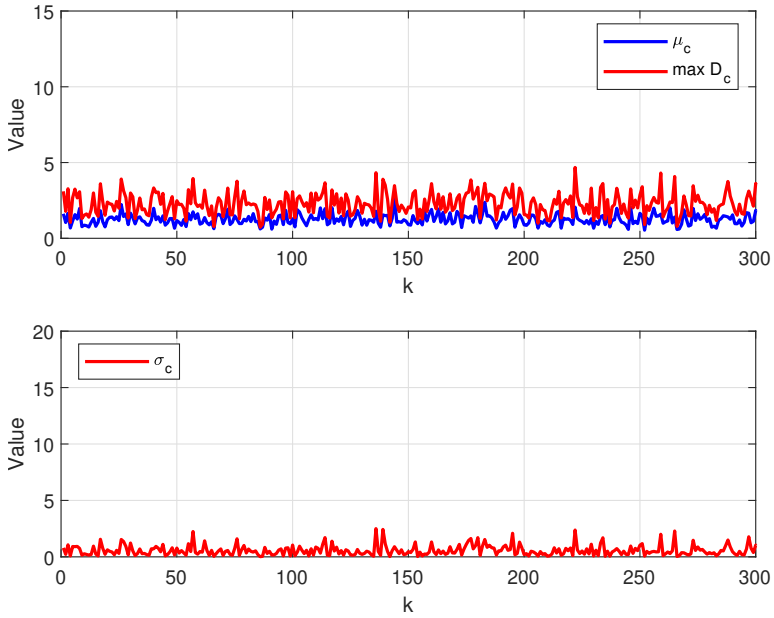
O classificador k-NN pode empregar qualquer conjunto de características nas estatísticas anteriores. Alguns resultados preliminares da simulação na seleção de características nos mostrou que o mínimo e o máximo valores de μ_c são boas escolhas para o classificador k-NN porque estes limites inferiores e superiores podem indicar melhor presença ou não de um atacante. Eles são denotados como mínimo de μ_c (min μ_c) e máximo de μ_c (max μ_c), respectivamente. As Figuras 12 e 13 apresentam exemplos destas métricas nos cenários sem atacante (Cenário 1) e no cenário com atacante (Cenário 2). As Figuras 14, 15 e 16 ilustram o comportamento das métricas do k-NN considerando apenas o nó verdadeiro e também os cenários com o nó atacante para distâncias $d_1 = d_2$=50m e $d_1 = d_2$=60m. Como exemplo, são apresentados os casos $min\ \mu_c \times \mu_c$, $max\ \mu_c \times \mu_c$ e $max\ \mu_c \times min\ \mu_c$. A partir de simulações, identificados que o cenário da Figura 16 apresentou um desempenho melhor e foi adotado para aplicação do método de identificação de ataques.

4.2.2 Resultados Parciais

O cenário de simulação assume um modelo de canal com parâmetros $\alpha = 2.5$, $\sigma^2 = 4$ dB e dois valores para o parâmetro Nakagami-m, $m = 1$ e $m = 4$. Estes valores representam as condições de desvanecimento severo e médio, respectivamente. A frequência de operação é $f = 2.4$GHz, potência de transmissão $P_t = 30$ dBm e ganhos de antena $G_t = G_r = 0$ dBi. A potência de referência recebida $P_r(d_0)$ foi calculada usando o modelo de perda de percurso no modelo *log-distance* (GOLDSMITH, 2005) a uma distância de referência $d_0 = 1$m. Assumimos uma distância de correlação $X_c = 50$ m e um tempo de correlação $X_t = 30$s. Inicialmente investigamos o desempenho da estratégia proposta nos três cenários definidos anteriormente, assumindo um modelo de canal com $m = 4$. Testamos o desempenho do classificador na identificação exata do tipo de cenário. Os resultados são apresentados na Tabela 2.

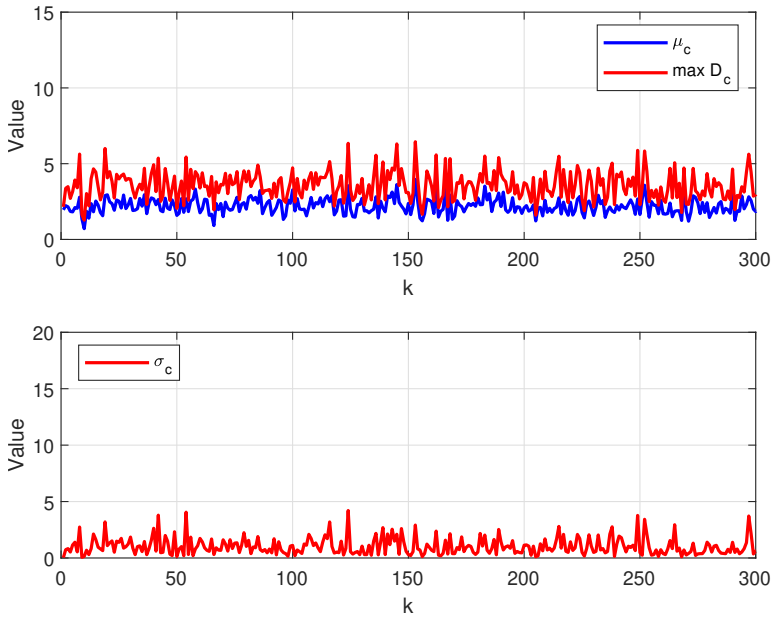
O Cenário 1 considera apenas presença do nó legítimo. Avaliamos a probabilidade

Figura 12 – Estatística do Cenário 1 (Sem Atacante)



Fonte: Autoria Própria

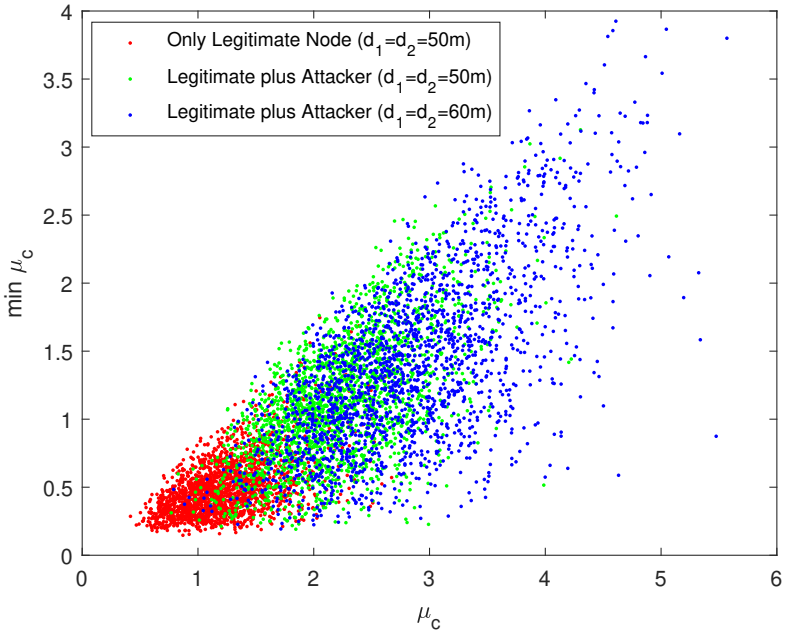
Figura 13 – Estatísticas do Cenário 2 (Com Atacante)



Fonte: Autoria Própria

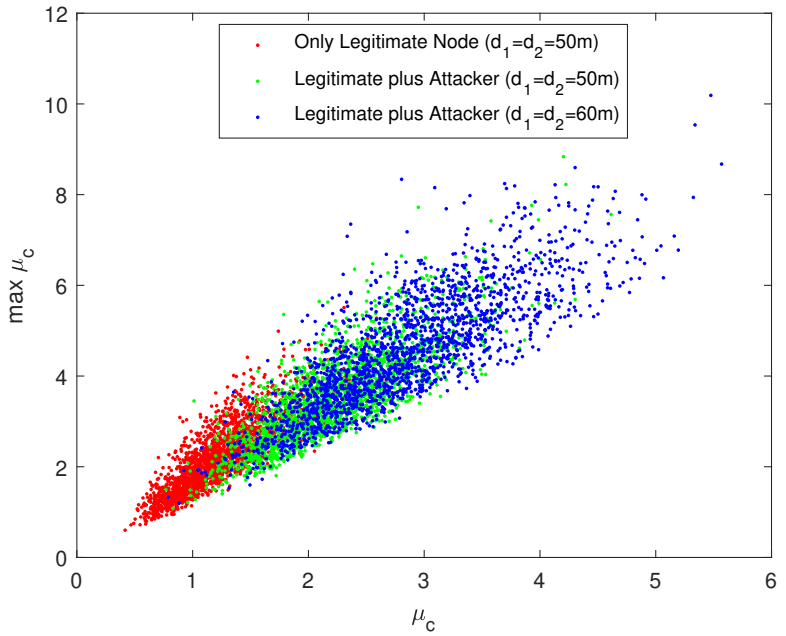
de detecção positiva verdadeira, que neste caso é a probabilidade de reconhecer a presença apenas do nó legítimo, que foi de 87,21%. O algoritmo pode detectar um ataque independentemente se o atacante está no Cenário 2 ou no Cenário 3, com probabilidade 87,06%. O algoritmo proposto foi comparado com a estratégia proposta em (CHEN et al.,

Figura 14 – Comportamento das Métricas para o k-NN



Fonte:Autoria Própria

Figura 15 – Comportamento das Métricas para o k-NN

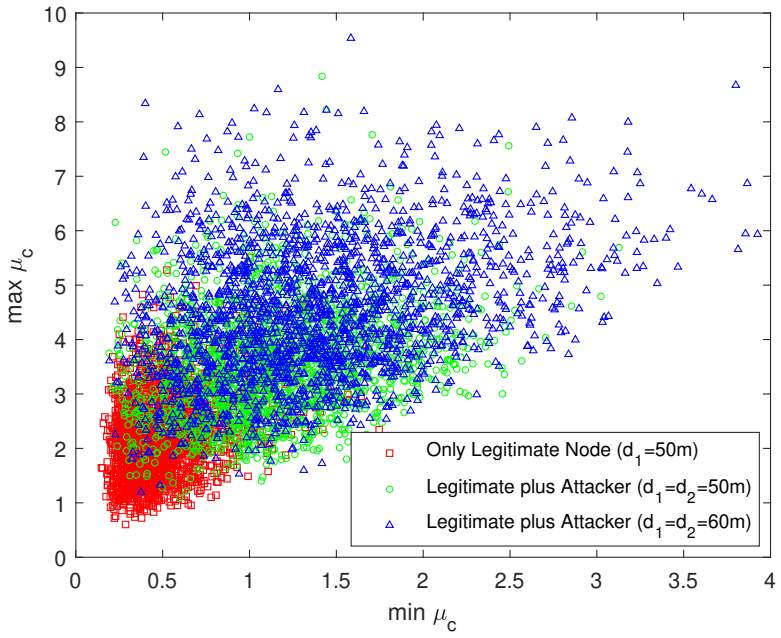


Fonte:Autoria Própria

2010). Dado que esse algoritmo depende do limiar de decisão, primeiro investigamos como o limiar afeta a precisão do algoritmo nos cenários 1, 2 e 3.

As Figuras 17 e 18 apresentam os resultados para os parâmetros de desvanecimento

Figura 16 – Comportamento das Métricas para o k-NN



Fonte:Autoria Própria

Tipo de Teste	**Probabilidade**
Identificação do Cenário 1	87,21%
Identificação com Atacante (Cenários 2 e 3)	87,06%

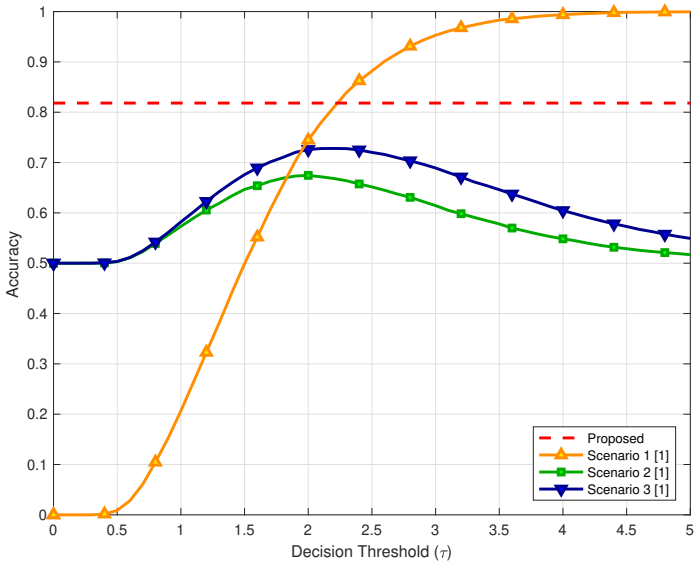
Tabela 2 – Resultados - Taxa de Detecção

$m = 1$ e $m = 4$, respectivamente. Considerando a Figura 17, um limiar de decisão em torno de $\tau = 1,5$ é uma boa escolha para maximizar o precisão nos cenários 2 e 3. Nesse caso, a precisão será de cerca de 70% no Cenário 1, 72% no Cenário 2 e 76% no Cenário 3. A partir das figuras podemos observar que um limiar mais baixo reduz a precisão em todos os cenários e um limiar mais alto melhora a precisão no Cenário 1, enquanto reduz a precisão nos Cenários 2 e 3. A linha vermelha tracejada representa a precisão da estratégia proposta. Ele desempenha bem em todos os três cenários e melhora a precisão de detecção em pelo menos 10% quando comparado com (CHEN et al., 2010).

Para uma condição de canal mais severa ($m = 1$), como apresentado na Figura 17, os resultados são semelhantes, no entanto, o desempenho de ambos os algoritmos diminui. O método proposto ainda apresenta bom desempenho em cenários com um invasor. A precisão melhora pelo menos 8% em relação à estratégia de (CHEN et al., 2010).

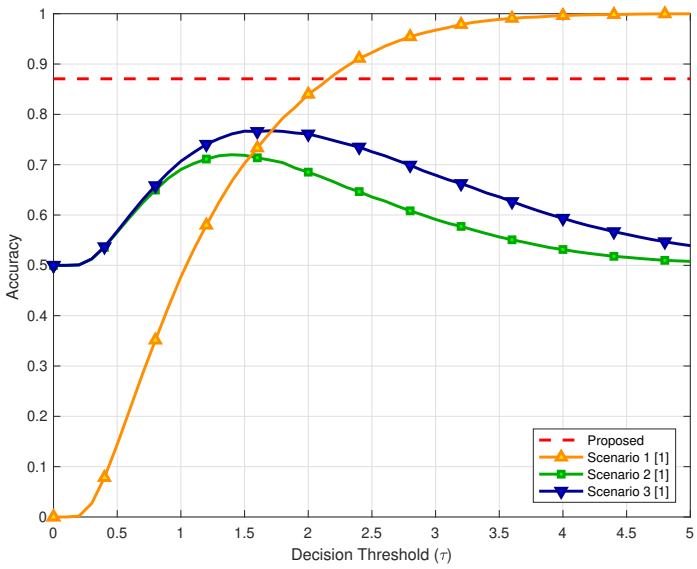
A estratégia proposta foi otimizada para cenários em que o nó legítimo e o nó atacante estão na mesma distância ou a uma distância muito próxima com relação ao *landmark*, o que é uma situação de pior caso. Consideramos uma modelagem mais precisa do canal de rádio, o que inclui efeitos de desvanecimento rápido e sombreamento com

Figura 17 – Cenário 1 2 3 para m=1



Fonte:Autoria Própria

Figura 18 – Cenário 1 2 3 para m=4



Fonte:Autoria Própria

correlação espacial e temporal. Os resultados mostram que a estratégia pode melhorar em cerca de 10% a detecção de ataque em relação à abordagem apresentada em (CHEN et al., 2010).

4.3 Modelo para Múltiplos APs em Ambientes Reais

A Figura 11 apresenta o método proposto inicial, que foi avaliado assumindo-se um cenário de simulação usando modelos de canal específicos. O método foi aplicado apenas para ambientes com características de ataques *spoofing* e com apenas um *landmark*. Nesta seção investigamos ambientes físicos reais com múltiplos *landmarks*, visando detectar a incidência de ataques *spoofing*.

O objetivo é adaptar e testar a aplicação da estratégia proposta em ambientes domésticos/empresariais reais de WLANs. Também pretendemos avaliar se a premissa de variação do sombreamento com as movimentações do ambiente, como por exemplo a movimentação de pessoas/objetos, abertura e fechamento de portas e janelas, entre outros fatores, pode gerar diferenciações significativas e que sejam úteis para o processo de detecção. A Figura 19 apresenta uma ideia inicial do cenário de avaliação real proposto. Nesse modelo apresenta-se um ambiente de rede, que se passa por tentativa de ataques *spoofing*, com quatro *landmarks*, os quais emitem sinais de RSS, que estarão sendo catalogados e armazenados, tanto pelo nó verdadeiro, quanto pelo nó falso.

Posteriormente, os conjuntos de RSS serão utilizados pelo analisador, o qual retorna aos *landmarks* com relatórios de possíveis nós atacante e legítimos. Na análise dos dados, a probabilidade sendo menor que 85% de falso-positivo, os possíveis alertas são descartados, acima de 85% de falso-positivos, o alerta é ativado para a presença de um possível invasor ou nó malicioso.

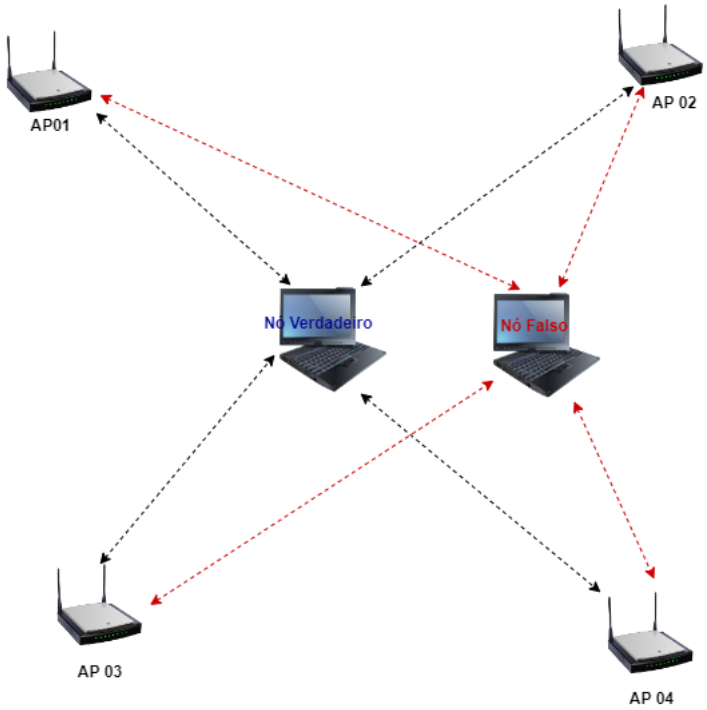
Como forma de aproveitar os avanços alcançados no modelo proposto na Figura 11 e na Tabela 2, se aplicará as técnicas já comprovadas em (CHEN et al., 2010), visando alcançar maior eficiência com novo modelo, porém, nessa proposta, se considera ambientes reais de redes WiFi, os quais se diferem e demandam muito mais da proposta, com maior demanda, que visam detectar técnicas de *spoofing* de endereço MAC.

Para isso utilizaremos de métricas da camada física como *shadowing*, $fading$ e RSS do canal numa rede sem fio, com múltiplos *landmarks* (APs). O compromisso é trazer uma breve ideia da proposta, a qual identificará uma variabilidades do RSS em cenários, desde ambientes comerciais, residenciais, acadêmicos, educacionais e residenciais com suas variações, com quatro *landmarks*. Para isso é necessário delimitar o ambiente estruturado, para a capitação dos pacotes de RSS, por cada *landmark* em cenários e suas peculiaridades.

Os *landmarks* emitem um sinal RSS, que é captado por um nó, seja ele verdadeiro ou falso. Assim, se estará num ambiente com múltiplos APs, nesse caso, quatro APs, com captação de RSS num intervalo de tempo específico, pré-definido, pra decidir se existe ou não a presença de um nó falso.

Com isso, buscar maior eficiência e evitar a sobrecarga de informações é primordial. A proposta prevê a aplicação num trabalho futuro, das informações do CSI. Vale ressaltar

Figura 19 – Modelo com 4 Landmarks (APs).



Fonte:Autoria Própria

que se recomenda aplicar aprendizado de máquina, que pode não se limitar aos cenários, já utilizados no modelo inicial, mas procurar testar outros algoritmos que possam se mostrar mais eficientes. Caso se apresente mais eficiente nos testes de detecção poderão ser agregado, substituídos ou descartados.

Assim, para a proposta desta tese serão propostos a captação de RSS, com quatro *landmarks*, em seis cenários bem distintos do mundo real. Posteriormente, se analisará o comportamento das métricas do canal com suas variações, com análise dos cenários sem atacante e com atacante no ambiente de teste. Diante disto verificar o comportamento e as possíveis variações do modelo, o dinamismo dos RSS de cada nó e a eficiência de cada AP. Nesta etapa se testará o comportamento do modelo proposto para nós próximos, a uma distância de cinquenta centímetros e um metro, entre os nós verdadeiro e falso.

4.3.1 Descrição da Metodologia

A seguir são apresentadas as principais etapas e procedimentos para a implementação do método prático, conforme etapas indicadas pelo fluxograma da Figura 20.

1. Definir a topologia física da rede, considerando os elementos físicos que estarão sendo usados no cenário de avaliação: APs, nó verdadeiro (laptop), nós falso (laptop),

distância entre os nós e configuração do software Netspot Mongiovi et al. (2013) para coleta das devidas métricas.

2. Definir quais redes (APs/landmarks) passarão por captação do RSS e trabablhar com pelo menos quatro APs.

3. Definir o intervalo de tempo de escaneamento das redes (landmarks): 5, 10 ou 30 segundos.

4. Coletar e armazenar os arquivos de RSS a serem analisados pelo modelo (CSV/XLSX).

5. Realizar um pré-processamento dos valores de RSS, eliminando medidas inconsistentes.

6. Sequenciamento (entrelaçamento temporal) das amostras de RSS coletadas do nó verdadeiro e falso, com relação a cada um dos *landmarks*.

7. Definição da janela temporal de amostras de RSS que serão analisadas, sendo de 25 ou 50 amotras.

8. Aplicar algoritmo de clusterização k-means para cada janela (i-ésima janela), determinando a distância entre os centróides, $D_c^{i,j}$, com relação a todos os *landmark* (j-ésimo *landmark*).

9. Repetir o passo anterior para todos os *landmarks* do cenário.

10. Cálculo da soma das distâncias entre os centróides, para cada janela, e com relação a cada *landmarks*. Esta nova métrica é denotada de $D_c^{i,sum}$

11. Definição automática do *threshold* de decisão com base nos dados iniciais de coleta, onde assume-se apenas a presença do nó verdadeiro (período inicial após a conexão do nó verdadeiro à rede).

12. A partir do *threshold*, a decisão de ataque é feita não com base em uma decisão instantânea, mas em função de um número de sequências positivas de detecções, denotado por *nds* (número de detecções positivas em sequência).

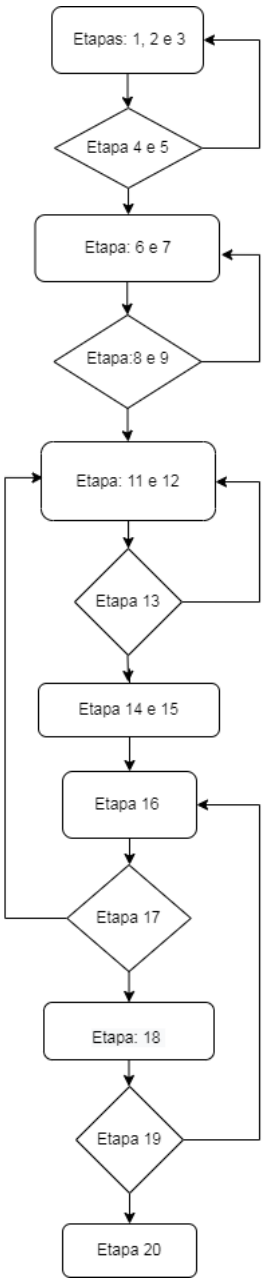
13. A decisão final estabelece um *threshold* para comparar o *nds*, visando alcançar uma maior probabilidade de detecção do nó falso no ambiente analisado.

14. Tomada de decisão para a presença ou não de nó malicioso na rede.

15. Respeitar o número mínimo de pelo menos quatro APs, tal recomendação é relevante, pela necessidade da realização da somas dos centroides dos conjuntos de cluster de RSS.

16. Identificar a variação do comportamento dos centroides, considerando os *RSS* de nós verdadeiros e falso juntos.

Figura 20 – Diagrama Fluxo do Modelo



Fonte:Autoria Própria

4.3.2 Ambientes de Teste

Para esse trabalho, a captação de RSS foi realizada em seis locais com características bem distintas de redes WiFi. Os ambientes serão detalhados no Capítulo 5, onde são apresentados os resultados nos ambientes diversos, como podemos encontrar na Tabela 3.

Cenários	Descrição
Ambiente 1	Centro multi-serviços públicos em Tubarão
Ambiente 2	Escola Pública Braço do Norte
Ambiente 3	Apartamento Condomínio em Tubarão
Ambiente 4	Residência Urbana Braço do Norte
Ambiente 5	Universidade Tubarão
Ambiente 6	Residência de Veraneio Jaguaruna

Tabela 3 – Ambientes de Testes - Análises - Distâncias

4.3.3 Disposição e Organização dos Componentes Físicos

Nesta etapa do trabalho procura-se abordar os equipamentos físicos básicos para a composição do cenário mínimo necessário para os experimentos. Posteriormente começar a captação de sinais dos *RSS* vale ressaltar que aplicaremos aprendizado de máquina, a proposta adotada no modelo inicial em (PINTO et al., 2018), porém procurando implementar novas etapas no algoritmo, que possam se mostrar mais eficientes, exequível e implementável. No decorrer destas implementações, surgindo novas situações que apresente resultados significativos de detecção de nós maliciosos/atacantes, poderão ser agregadas ao modelo proposto. No entanto, se faz necessário respeitar as colunas recomendadas na Tabela 3.

Para isso, se aplica a mesma estratégia adotada em (PINTO et al., 2018), porém não apenas com um *LandMark* mas com múltiplos *LandMark*. Posteriormente, se analisará o comportamento das métricas do canal com suas variações, com ênfase em cada cenário sem atacante e com atacante no ambiente de teste. Para isso, se verificará o comportamento do modelo diante da dinâmica dos *RSS*, em de cada nó e a eficiência de cada AP. Nesta etapa se testará o comportamento do modelo em distâncias pequenas, para casos de cinquenta centímetros e um metro, tanto tanto para nós legítimos, quanto para nós maliciosos/atacantes.

Os APs que estarão compondo o cenário, são APs comuns, acessíveis em redes WiFi para ambientes residencial, comercial, governamental, acadêmico ou educacional. Assim como os notebooks usados para captação dos RSSs, foram dois notebooks populares, sem a necessidade de configurações mais robustas ou específica, para implementar o método proposto.

4.4 Metodologia de Campo

Nesse momento serão utilizados cenários distintos, com característica bem particulares, para a execução do modelo. Nos cenários serão respeitados o limite de quatro APs para a captação de RSS das redes dos ambientes.

4.4.1 Etapa 1 - Especificação dos Ambientes de Medição

A Etapa 1 engloba os seguintes procedimentos:

- Levantamento de cenário o qual será objeto de análise no projeto de pesquisa.
- Catalogação dos equipamentos básicos e espaços físicos que estarão como objeto de estudos.
- Estudar a estrutura da rede WiFi dos seis ambientes;
- Realizar o levantamento de equipamentos e requisitos básicos das redes WiFi do ambiente pesquisado;
- Identificar as topologias físicas e lógicas dos Locais de execução do projeto.

4.4.2 Etapa 2 - Coleta de Dados

Essa etapa deve respeitar ocorrência semanal, em dias com maior atividade educacional na escola. Os dados de RSS serão coletados, através de dois notebooks com configuração equivalentes e especificações técnicas similares, sistemas operacionais Windows ou linux, com aplicativos/plataforma básica usada pela escola, visando assim, maior proximidade com um cenário da realidade com influência de nós falso, se passando por um nó verdadeiro. A periodicidade diária da coleta ocorrerá a cada trinta minutos, assim também como a cada a duas horas no dia da catalogação, com visualizações de cinco, trinta e sessenta minutos. Na oportunidade pretende-se ter em mãos dados estatísticos de acesso, concorrência de usuários ao sinal e o comportamento da rede. Enfocando nas oscilações e instabilidades dos sinais de cada ponto de acesso. Esse momento é considerado relevante, pois o ambiente de redes WiFi num futuro próximo, estará inserido em novos servidos oferecidos pelas tecnologias IoT. Posteriormente adotar/executar o *Netspot* na redes, para captação e variação dos RSS. Ao final dessa etapa, teremos em mãos os dados coletados na rede, os quais serão formatados a ponto de serem absorvidos pelo modelo inicial, o qual servirá de base para implementação da metodologia do projeto.

4.4.3 Etapa 3 - Análise de Dados

Com os dados em mãos levantados nos ambientes de teste, analisados, compilados, com o uso de técnicas de *Machine Learning* para o conjunto de RSS captados, tanto pelo nó verdadeiro, quanto pelo nó falso. Com isso intruso estará usando técnica de ataque *spoofing*, o que reduz os danos causados no ambiente WiFi da pesquisa. Com os dados pré-coletados, realizar comparativos de eficiência e eficacia, em relação a outros métodos já propostos de detecção de *spoofing* ataques. Nesse momento, se medirá o percentual da presença de falso-positivos e falso-negativo, pelo método proposto, e assim ajustar o

melhor faixa de limiar, que esteja bem caracterizado a presença de um nó falso/invasor. Assim, ao final dessa etapa se terá em mãos os dados iniciais, os quais servirão de subsídios para a construção do modelo proposto.

4.4.4 Etapa 4 - Construção de Modelo Proposto

A segurança da informação não tem recebido a devida atenção por parte dos usuários de ambiente de redes WiFi e que num futuro próximo estarão inseridos em novos servidos oferecidos pela filosofia tecnológica de IoT. Tais práticas visam encontrar soluções à problemática específica do projeto. Mas, apenas propondo novos métodos e práticas com a tecnologia existente. Nesse momento o projeto estará aplicando os dados coletados, organizados e formatados na ferramenta Matlab, no qual os códigos serão implementados através de rotinas de programação, com auxílio de técnicas do machine learning. Diante disso, essa etapa possibilitará a viabilidade de teste e análise dos dados que servirão como base para o modelo proposto.

4.4.5 Etapa 5 - Testes do Modelo Proposto

O modelo proposto passará por um período de teste em ambiente real. Nesse período será verificado se a proposta retrata as variações de um ambiente real, comparar com modelos já propostos que são retratados em artigos científicos, analisando as consistências, realizando os ajustes necessário, visando alcançar um modelo estável e mais próximo da realidade assim também como os desafios para novos estudos e pesquisas científicas. Ao final dessa etapa, se faz necessário analisar as virtudes e limitações do modelo.

4.5 Considerações Finais

Este capítulo apresentou a ideia do modelo proposto por este trabalho, onde foi considerado a proposta inicial, seus cenários, a descrição do problema, modelo simulado, estratégia de detecção de ataque, resultados parciais, modelos com múltiplos *APs* em ambientes reais. Na oportunidade foi realizado a descrição do método para ambientes reais, os componentes físicos, que precisa ser considerado ao se implementar a proposta num ambiente de WiFi.

5 RESULTADOS

Como forma de se avaliar a eficiência do método proposto, realizamos experimentos em diferentes cenários. Se procurou analisar o comportamento do RSS do nó verdadeiro e do nó falso, em duas etapas. No primeiro momento, para a distância de cinquenta centímetros e na segunda etapa para a distância de um metro. Ainda na mesma linha de raciocínio, a captação de sinais de RSS seguiram um intervalo de captação de cinco segundos. Posteriormente foram analisados os pacotes de RSS coletados em intervalos de trinta minutos, através de janelas de análise 25 e 50 amostras de RSS, com e sem a presença do nó falso na rede. O experimento foi aplicado em seis cenários distintos, que serão mais detalhados nas próximas seções:

- Facilita Tubarão
- Escola de Ensino Básico Dom Joaquim
- Biblioteca Ânima em Tubarão
- Residência em Braço do Norte
- Condomínio no Centro de Tubarão
- Residência de Veraneio Jaguaruna

Porém, para se realizar as medições nos ambientes pré-definidos, se faz necessário no primeiro momento analisar o ambiente que receberá o experimento, identificando o posicionamento e a quantidade de roteadores existentes na rede. Neste trabalho, procurou-se enfatizar também a presença do nó falso, mesmo considerando variações de ambientes, com suas peculiaridades mais sensíveis, como o abrir e fechar de portas e janelas, ruídos e interferências que poderiam comprometer a sinal do roteador e a oscilação das potências transmitidas e recebidas no cenário. Nessa fase do trabalho, executamos os experimentos com o auxílio de dois notebooks, tendo como base o sistema operacional Windows e o uso do software Netspot (MONGIOVI et al., 2013), (BREITENBAUCH, 2015), com a seguinte configuração para a captação de RSS nas redes WiFi:

- Intervalo de Scan: 5 segundos;
- Tempo total de coleta de RSS de 30 minutos para cada cenário.

Com isso, obtivemos um total de 360 RSS de cada nó, referente aos quatro APs em questão, por um intervalo de tempo de pelo menos cinco horas de captação de RSS.

É importante ressaltar que o procedimento foi realizado tanto para distâncias de 50 cm, quanto para distâncias de um metro. Posteriormente, as medidas de RSS armazenadas em formato de planilha Excel. Aplicamos o algoritmo proposto, com janelas de análise de 25 e 50 amostras de RSS, para cada distância em questão (50 cm e 1 m).

Definimos o *threshold* entre 14 e 18, para se detectar com bom percentual de acerto a presença do nó atacante nos cenários. É relevante observar que foi possível detectar a presença do nó atacante nos cenários de teste de forma estatística, apenas com a utilização das informações de RSS.

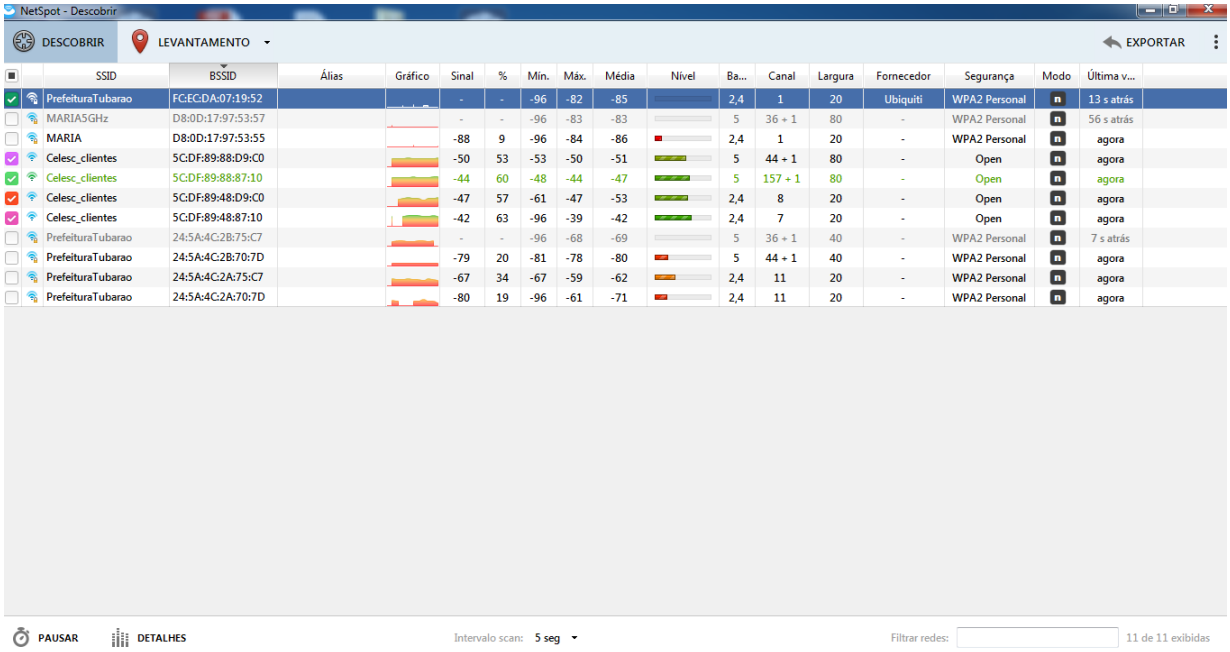
5.1 Facilita Tubarão

Realizamos a captação de RSSs em seis locais com características bem distintas de redes WiFi. O primeiro foi o ambiente do Facilita Tubarão, local de múltiplos serviços públicos e com grande movimentação de pessoas e dispositivos móveis. Foram coletadas aproximadamente 3500 amostras de RSSs dos APs da rede, com intervalos de escanamento de 5 segundos, usando o software Netspot. A Figura 21 mostra o ambiente externo do Facilita Tubarão.

Figura 21 – Ambiente 01 - Facilita Tubarão - Centro Público de Multi-Serviços

A Figure 22 ilustra a tela de coleta de dados da ferramenta NetSpot. Analisamos o cenário para distâncias de 50 centímetros e um metro no Facilita Tubarão, nas dependências do Sine, para janelas com 25 e 50 amostras. Observou-se que o *threshold* de decisão mais adequado seria de valor 17 para a distância de 50 cm e 1m, para um conjunto de 25 e 50 amostras.

Figura 22 – Netspot Facilita



Fonte: Própria

5.1.1 Distância de 50 cm

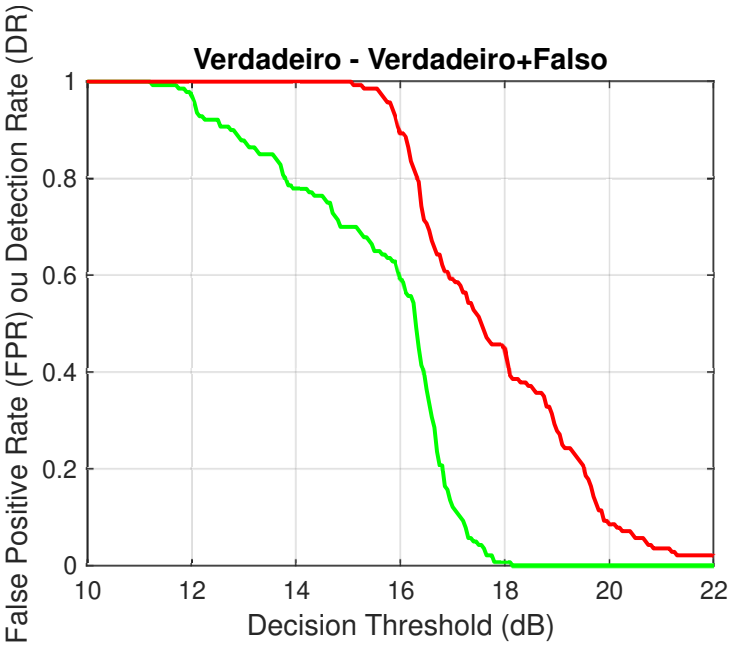
As Figuras 23 e 24 mostram os gráficos dos resultados, onde as linhas verdes consideram apenas o nó verdadeiro e as linhas vermelhas, consideram a presença do nó verdadeiro e falso na rede, para a distância de 50 cm.

5.1.2 Distância de 1 metro

Por outro lado, as Figuras 25 e 26 mostram o *threshold* variando entre 12 e pouco mais de 21. Os gráficos são resultados para a distância de um metro no ambiente do centro de atendimento ao cidadão Facilita Tubarão.

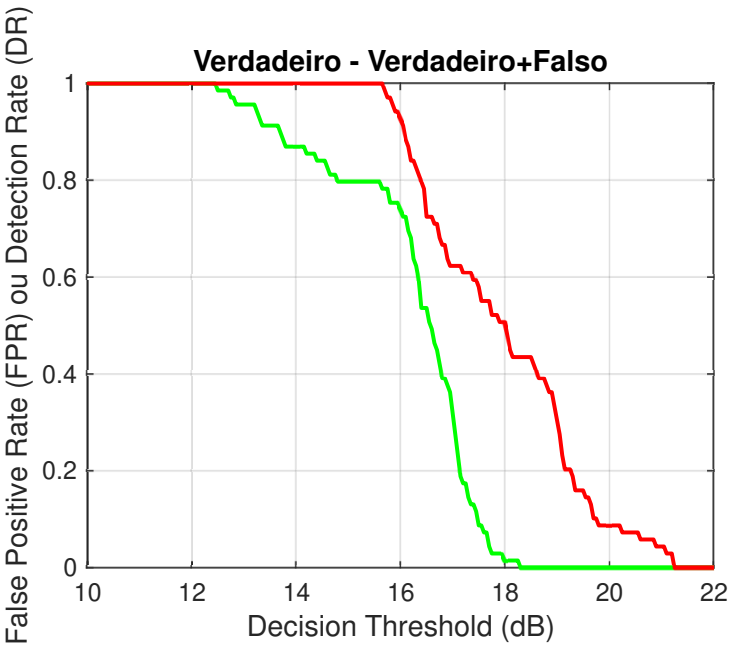
Quando a distância entre os nós verdeiro e falso passa para um metro, o melhor valor de *threshold* seria o 17 e com 63% de detecção aproximadamente. Isso também ocorre para uma distancia de 50 cm e janelas de 25 amostras de RSS. Já para a distância de 50 cm, 50 amostras, temos um retorno de 63% de detecção mas, com um *threshold* de 17,5. Por outro lado, quando ocorre a variação da distância para um metro, num conjunto de RSS de 25 amostras, chegamos a um *threshold* de 17,5 com 67% de percentual de detecção do nó falso. Tal situação varia um pouco quando temos um conjunto de 50 amostras de RSS, chegamos ao um *threshold* de 17,4 temos uma taxa de detecção de 76%, como mostrado na Tabela 4. Contudo, é interessante frisar que para as duas distâncias implementadas no método proposto, consegue-se identificar a presença do nó falso na rede. Assim, esses foram

Figura 23 – Distância de 50 cm com 25 amostras - Facilita



Fonte: Própria

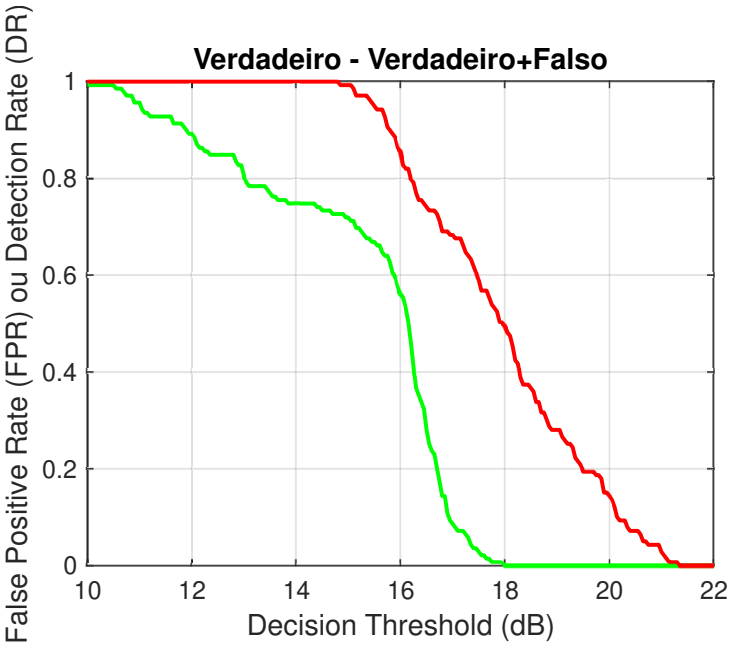
Figura 24 – Distância de 50 cm com 50 amostras - Facilita



Fonte: Própria

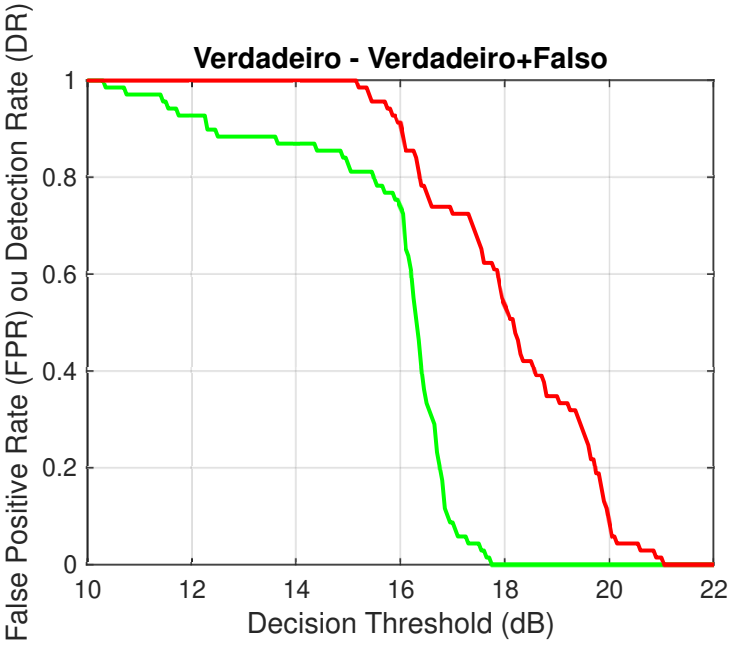
os resultados obtidos, com a aplicação do método proposto, nas distâncias de cinquenta

Figura 25 – Distância de 1 m com 25 amostras - Facilita



Fonte: Própria

Figura 26 – Distância de 1 m com 50 amostras - Facilita



Fonte: Própria

centímetro e um metro, nas dependências do Facilita Tubarão.

Janela	**Threshold**	Detecção	FP	Dist. 50 cm	Dist. 1 m
25 RSS	17	60%	20%	X	
50 RSS	17,7	60%	4%	X	
25 RSS	16,6	76%	19%		X
50 RSS	17,5	71%	5%		X

Tabela 4 – Análise dos Resultados do Facilita Tubarão

5.2 Escola Dom Joaquim - Braço do Norte

O segundo ambiente foi o cenário da Escola Pública Braço do Norte, escola centenária, que atualmente atende uma comunidade de aproximadamente 1100 alunos para toda a região do Vale do Braço do Norte. Para isso, foi inicialmente realizado o levantamento de roteadores do ambiente, as redes que circundam o cenário, para então iniciar a captação de RSS. A Figura 27 retrata o cenário externo do ambiente escolar público em questão. Observou que o threshold de decisão mais adequado seria 15, para um conjunto de 25 e 50 amostras, variando um pouco, mas 15 se apresenta como o mais adequado para a distância de meio-metro. No entanto, para a distância de um metro, pelo retorno das figuras, o mais adequado seria o threshold de 14, considerando-se que se apresenta como o mais adequado para que se alcance uma alta taxa de detecção do nó falso, com variação entre 11 e 22. Contudo, se faz interessante notar nos experimentos que, para as duas distâncias de 50 cm e um metro, o método conseguiu identificar o nó falso após a análise do método.

Figura 27 – Ambiente 02 - Escola Pública em Braço do Norte

Fonte: Google Maps

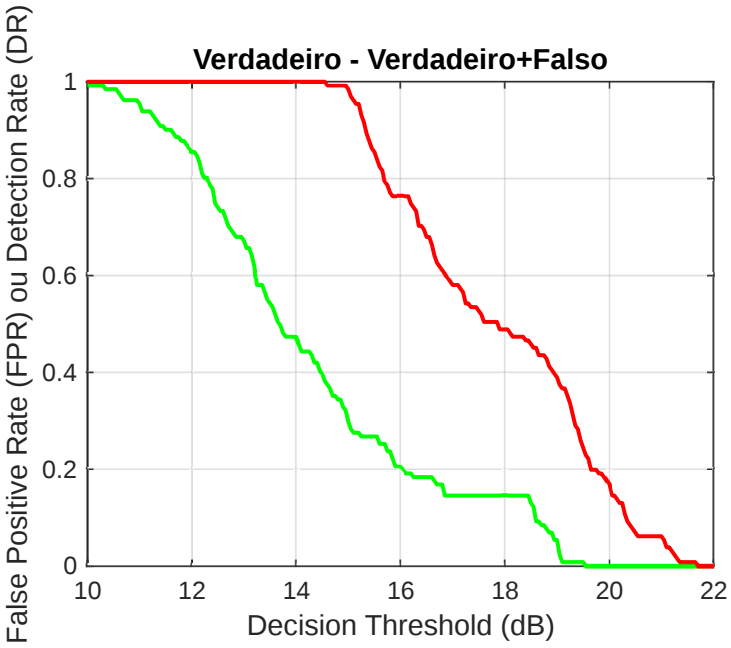
5.2.1 Distância 50 cm

As figuras 28 e 29 apresentam o comportamento entre os nós verdadeiros e falso no ambiente da escola pública numa distância de 50 cm.

5.2.2 Distância 1 metro

As figuras 30 e 31 mostram o resultado de detecção considerando o nó verdadeiro e o falso para distância de um metro.

Figura 28 – Distância de 50 cm com 25 amostras

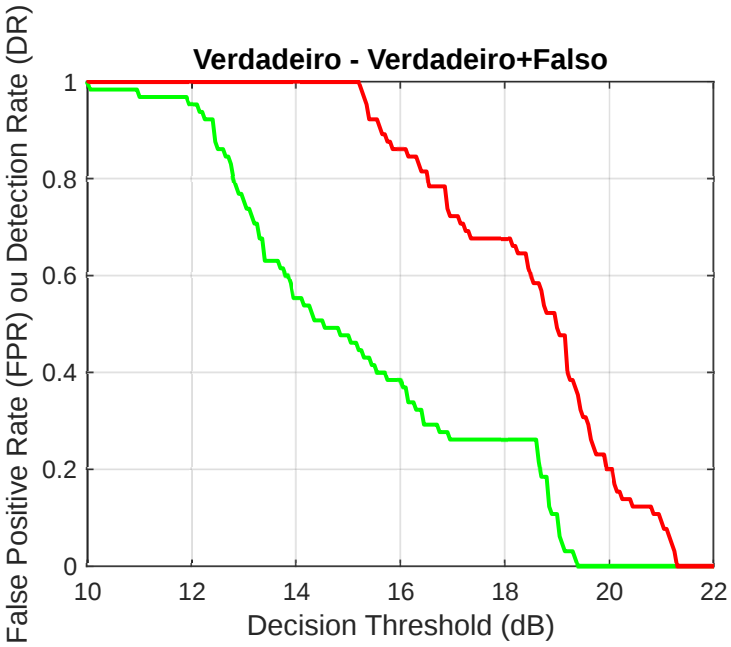


Fonte: Própria

Diante dos gráficos das Figuras 30 e 31 compreende-se que o melhor threshold com a presença do nó verdadeiro (linha verde) e nó falso (linha vermelha), é o 16. Esse se apresenta como o mais apropriado visando se alcançar uma alta taxa de detecção do nó falso. No entanto, quando a distância entre os nós verdeiro e falso passa para um metro, nas Figuras 28 e 29, o melhor threshold número já não seria o 16, o que retorna uma baixa taxa de erro, na distância de um metro para este cenário que a distância de cinquenta centímetro.

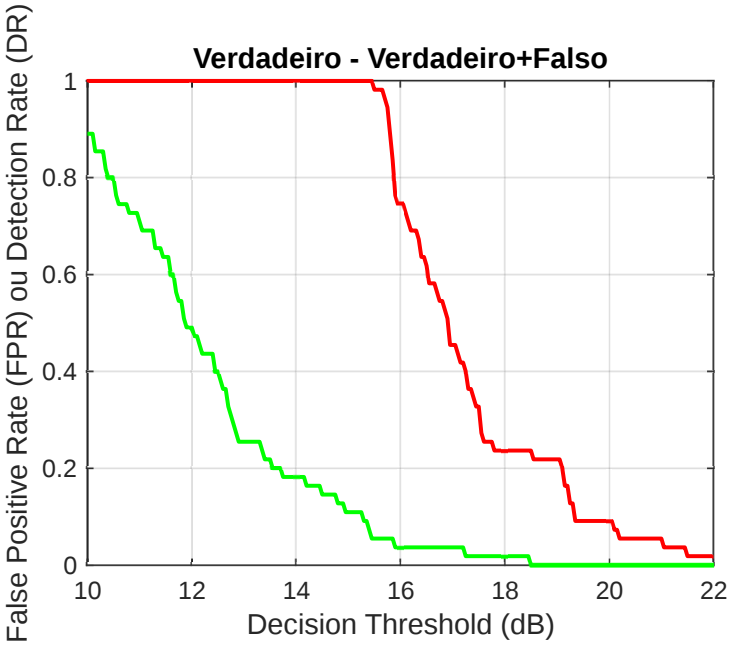
Porém, é relevante considerar a necessidade de respeitar as peculiaridades, do cenário da Escola Estadual Dom Joaquim, como forma de se chegar a estes resultados, através das variáveis envolvidas como: o número total de RSS catalogados, os conjuntos de janelas que foram utilizadas e as distâncias que foram adotadas (cinquenta centímetro e um metro), que retornaram tais resultados na Tabela 5.

Figura 29 – Distância de 50 cm com 50 amostras



Fonte: Própria

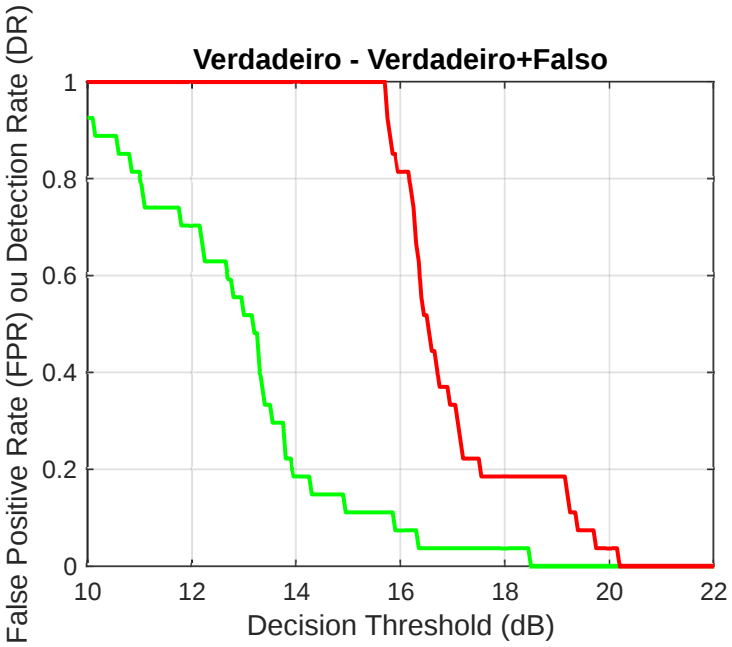
Figura 30 – Distância de 1m com 25 amostras



Fonte: Própria

Contudo, é interessante frisar que para nas duas distâncias implementadas no

Figura 31 – Distância de 1m com 50 amostras



Fonte: Própria

Janela	**Threshold**	Detecção	FP	Dist. 50 cm	Dist. 1 m
25 RSS	14,85	99%	30%	X	
50 RSS	16	90%	46%	X	
25 RSS	15,85	80%	11%		X
50 RSS	15,75	100%	11%		X

Tabela 5 – Análise dos Resultados Dom Joaquim

método proposto, consegue-se identificar a presença do nó falso na rede. Assim, esses foram os resultados obtidos, com a aplicação do método proposto, nas distâncias de cinquenta centímetro e um metro, nas dependências da Escola Dom Joaquim em Braço do Norte - Santa Catarina. Neste caso, a distância de um metro, com o threshold de 15,75%, se encontra um percentual baixo de falso positivo de 11%.

5.3 Condomínio Centro de Tubarão

O Ambiente 03 foi definido como sendo um condomínio residencial no centro da cidade de Tubarão-SC com 77 apartamentos, conforme mostrado na Figura 32. Este cenário possui a presença muito forte de roteadores atuando na redondeza do residencial. Neste caso observou que o threshold de decisão mais adequado seria 14, para um conjunto de 25 e 50 amostras, variando um pouco, mas 14 se apresenta como o mais adequado para esse cenário apenas para a distância de meio metro, num cenário onde o threshold variou de 10

a pouco mais de 18.

Assim, para a distâncias de um metro, o threshold de 14 também se apresenta a melhor decisão. Se percebe um comportamento bem diferente dos resultados, de outros cenários, para a distância de um metro. Mesmo assim, recomenda-se especificamente mais estudos, para esse experimento e nessa distância, porém, sua variabilidade ocorreu num cenário onde o threshold variou de 10 a pouco mais de 16. Contudo, mesmo nesse caso, a presença do nó falso é realçada pela linha vermelha nas figuras, ou seja, o método consegue identificar a presença de um nó falso no cenário.

Figura 32 – Ambiente 03 - Condomínio Residencial de Tubarão - SC

Fonte: Google Maps.

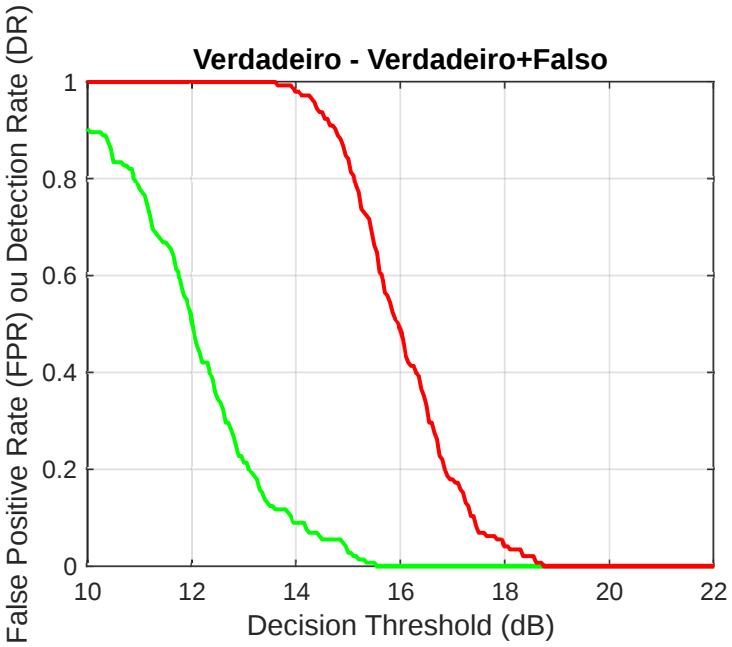
5.3.1 Distância 50 cm

As Figuras 33 e 34 mostram os gráficos dos resultados para a distância de 50 centímetros.

5.3.2 Distância 1 metro

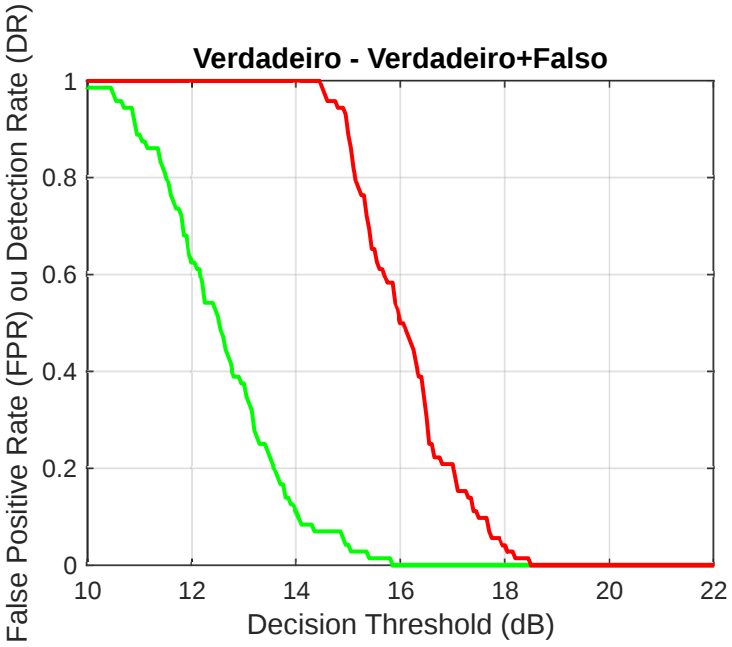
Realizou-se a captação de RSS para a distância de um metro, com 25 amostras num primeiro momento e 50 amostras no segundo momento. As Figuras 35 e 36 mostram os gráficos dos resultados para a distância de um metro.

Figura 33 – Distância de 50 cm com 25 amostras



Fonte: Própria

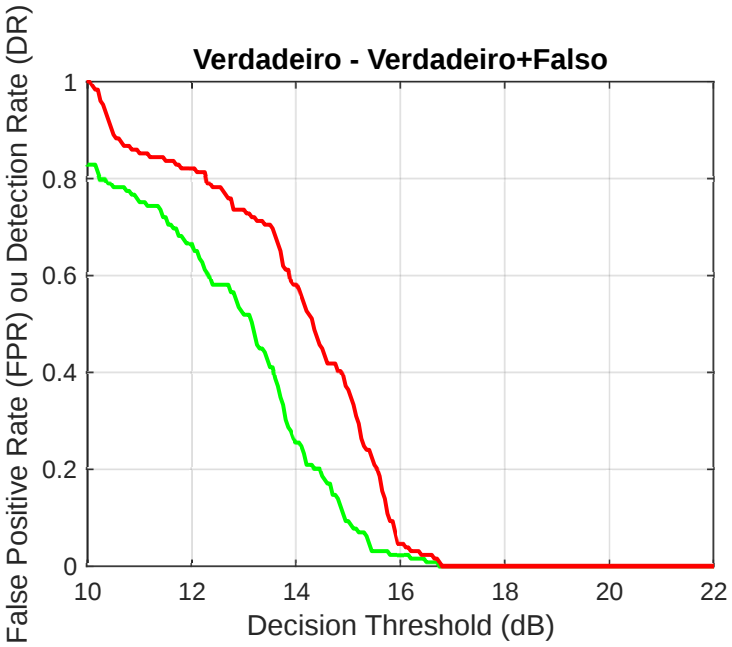
Figura 34 – Distância de 50 cm com 50 amostras



Fonte: Própria

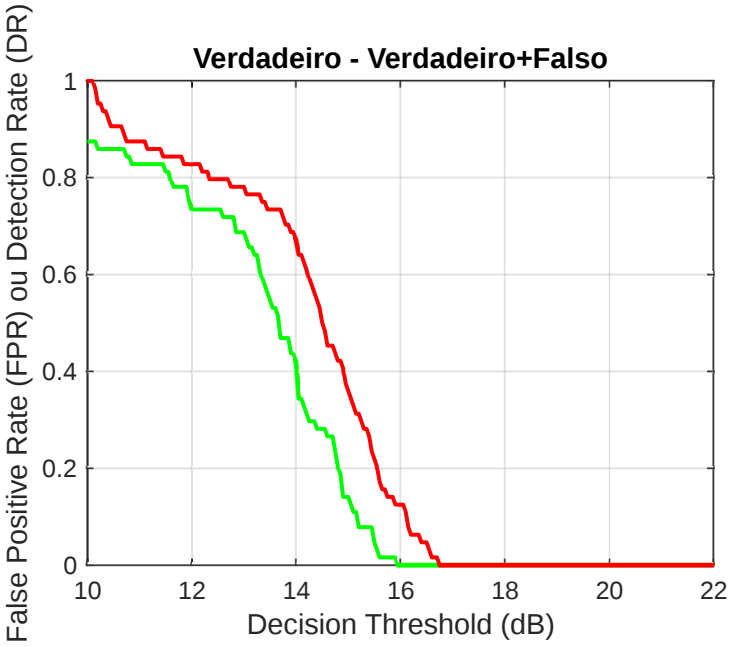
Diante dos gráficos das Figuras 33 e 34 compreende-se que o melhor threshold com

Figura 35 – Distância de 1m com 25 amostras



Fonte: Própria

Figura 36 – Distância de 1m com 50 amostras



Fonte: Própria

a presença do nó verdadeiro (linha verde) e nó falso (linha vermelha), é o 18. No entanto,

quando a distância entre os nós verdeiro e falso passa para um metro, nas Figuras 35 e 36 o melhor threshold seria 14, o que retorna uma taxa equilibrada, na distância de um metro para o mesmo cenário que a distância de cinquenta centímetros.

Porém, é relevante considerar a necessidade de respeitar as peculiaridades do cenário do Residencial Murano, local altamente ruidoso e muito desafiador. Contudo, é interessante frisar que para nas duas distâncias implementadas no método proposto, consegue-se identificar a presença do nó falso na rede. Assim, esses foram os resultados obtidos, com a aplicação do método proposto, nas distâncias de cinquenta centímetros e um metro, nas dependências do Residendial Murano em Tubarão, como mostra a Tabela 6.

Janela	**Threshold**	Detecção	FP	Dist. 50 cm	Dist. 1 m
25 RSS	14	97%	9%	X	
50 RSS	14	98%	13%	X	
25 RSS	14	60%	28%		X
50 RSS	14	60%	31%		X

Tabela 6 – Análise dos Resultados do Residencial Murano

Assim, esses foram os resultados dos experimentos realizados de um apartamento, dentro de um condomínio no centro de Tubarão. Vale ressaltar que os horários de captação de RSS na rede WiFi foram variados, passando por momentos de horários de picos de uso das redes no condomínio, como entre meio-dia e as quatorze horas, assim também como das dezoito horas até as vinte horas. Vale ressaltar, que o *threshold* apresenta uma variação referente aos ambientes anteriores, devido a topologia física do ambiente de teste, a presença de equipamento eletro-eletrônicos, os quais emitem sinais eletromagnéticos e a poluição dos canais/espectro, que são diferentes.

5.4 Residência em Braço do Norte

O Ambiente 04 foi definido como sendo a Residência Popular na cidade de Braço do Norte em Santa Catarina. A Figura 37 apresenta o cenário externo do ambiente residencial popular, para a captação de sinal dos APs.

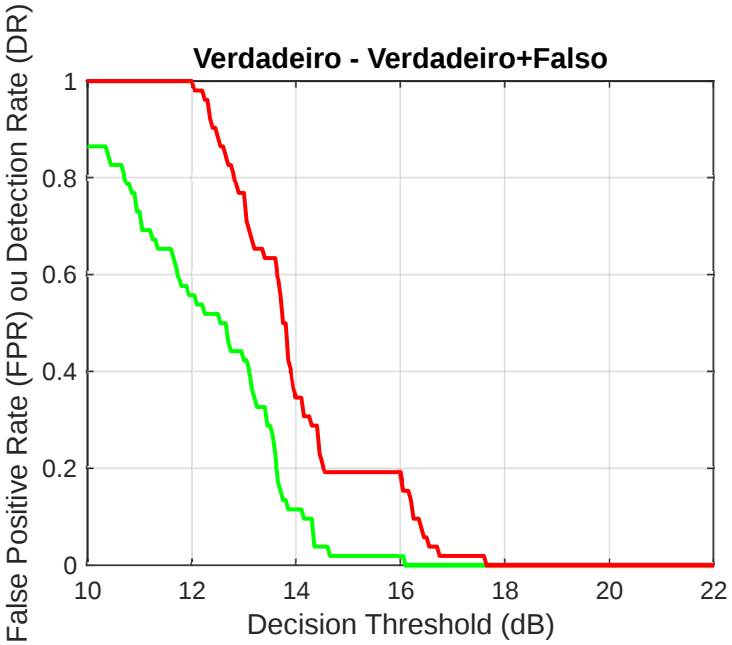
5.4.1 Distância 50 cm

As Figuras 38 e 39 retratam o comportamento do RSS, do nó falso em relação ao nó verdadeiro para as distâncias de 50 centímetros e um metro, num ambiente residencial comum, com uso de redes WiFi, com janelas de análise de 25 e 50 amostras.

Figura 37 – Ambiente 04 - Residência Braço do Norte

Fonte: Própria

Figura 38 – Distância de 50 cm com 25 amostras

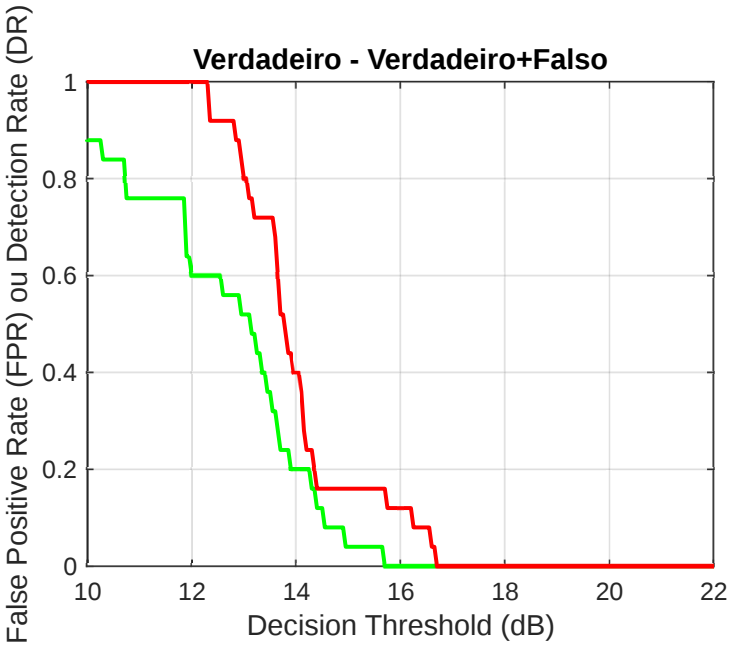


Fonte: Própria

5.4.2 Distância 1 Metro

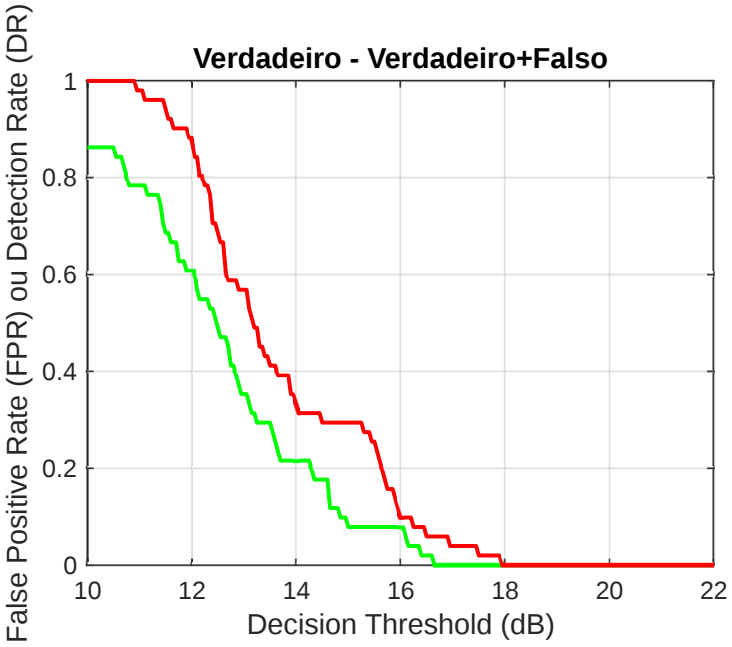
As Figuras 40 e 41 retratam o comportamento do RSS do nó falso em relação ao nó verdadeiro para as distâncias de 50 centímetros e um metro.

Figura 39 – Distância de 50 cm com 50 amostras



Fonte: Própria

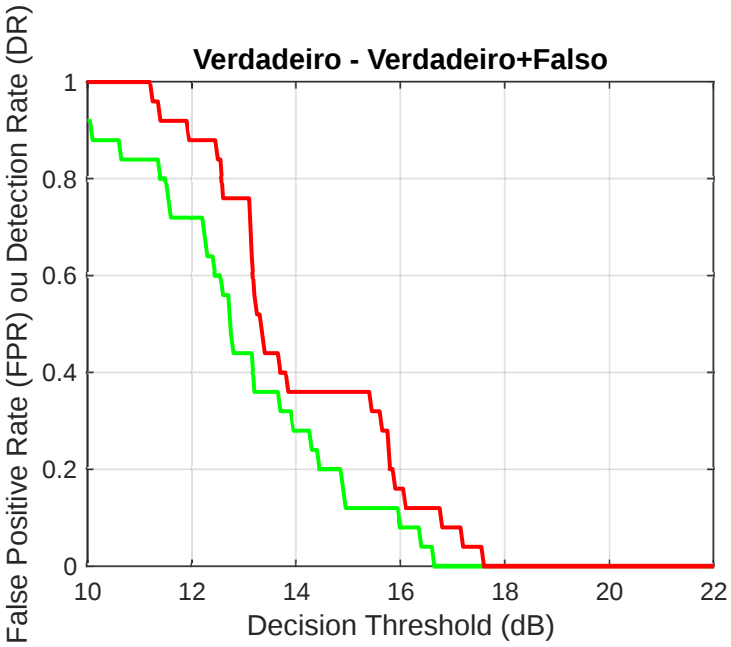
Figura 40 – Distância de 1m com 25 amostras



Fonte: Própria

Dessa forma, vemos um exemplo da dinâmica de detecção nas Figuras 40 e 41 na

Figura 41 – Distância de 1m com 50 amostras



Fonte: Própria

utilização de redes WiFi comuns numa residência popular na cidade de Braço do Norte em Santa Catarina. Nesse experimento, foram utilizados pelo menos 1500 amostras de RSS, para as distâncias de cinquenta centímetros e um metro, sendo as janelas de análise de 25 e 50 amostras de RSS.

Nesse caso, se observou que o threshold de decisão mais adequado não ficou muito claro, entre o intervalo de 10 a 18, para um conjunto de 25 e 50 amostras, precisando assim mais estudo, ajustes e maior adequação do ambiente para definir o threshold mais adequado para esse cenário. A partir dos gráficos das Figuras 38 e 39 compreende-se que o melhor threshold com a presença do nó verdadeiro (linha verde) e nó falso (linha vermelha), é o 13. Com esse threshold, pode-se observar um percentual mais equilibrado da influência dos RSS do nó verdadeiro, e por consequência, um mais alto percentual de detecção na presença de um nó falso em relação ao nó verdadeiro, por consequência threshold mais equilibrado para esse cenário numa distância de 50 cm.

No entanto, quando a distância entre os nós verdeiro e falso passa para um metro, nas Figuras 40 e 41 observa-se que o melhor threshold não seria o 13, mas sim o 15, que retorna uma melhor taxa de erro para o cenário de um metro, tanto para o conjunto de 25 janelas, quanto para 50 janelas de RSS. Os resultados estão sumarizados na Tabela 7.

Contudo, a presença do nó falso pode ser identificado através da linha vermelha nas figuras, ou seja, mesmo nesse caso, o método consegue identificar o nós falso no ambiente.

Janelas	**Threshold**	Detecção	FP	Dist. 50 cm	Dist. 1 m
25 RSS	13	80%	38%	X	
50 RSS	13	80%	38%	X	
25 RSS	15	31%	9%		X
50 RSS	15	32%	12%		X

Tabela 7 – Análise de Resultados da Residência em Braço do Norte

5.5 Ambiente da Biblioteca Ânima em Tubarão

O Ambiente 05 foi definido como sendo a Biblioteca Universidade Ânima no município de Tubarão em Santa Catarina, conforme mostrado na Figura 42. A Figura 43 ilustra a tela de coleta do software Netspot.

Figura 42 – Ambiente 05 - Biblioteca Universidade Ânima

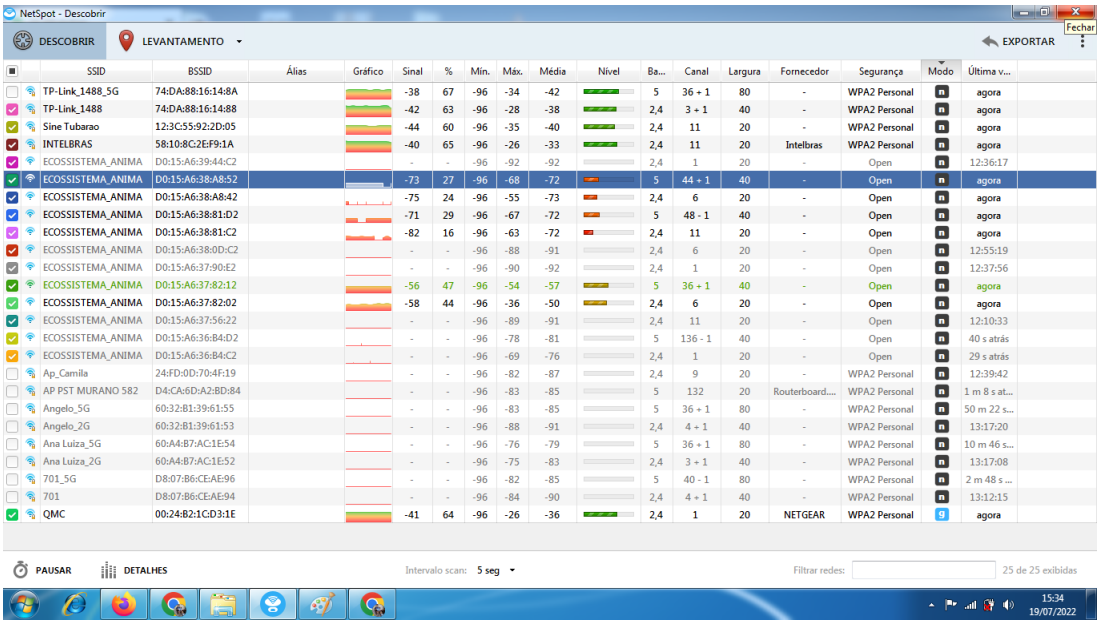
Fonte: Própria

Observou-se que o threshold de decisão mais adequado seria 15, para um conjunto de 25 e 50 amostras, variando um pouco, mas 15 se apresenta como o mais adequado para esse cenário, para as duas distâncias analisadas. Diante disso, é prudente considerar que mesmo, na distância de meio metro, ocorrendo variação, no caso do cenário com o conjunto de 50 amostras de RSS, ainda podemos considerar o 15 pode ser ainda a melhor escolha de decisão do threshold.

5.5.1 Distância 50 cm

As Figuras 44 e 45 apresentam os resultados do método para a distância de 50 cm. Observa-se que o melhor threshold com a presença do nó verdadeiro (linha verde) e nó falso (linha vermelha), é de 18. Com esse threshold, pode-se observar um baixíssimo

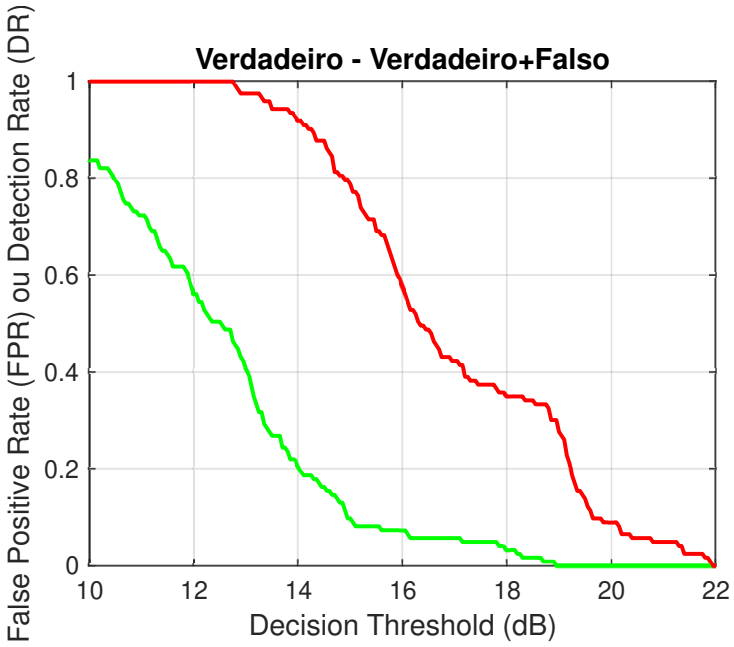
Figura 43 – Software Netspot - Universidade Ânima



Fonte: Própria

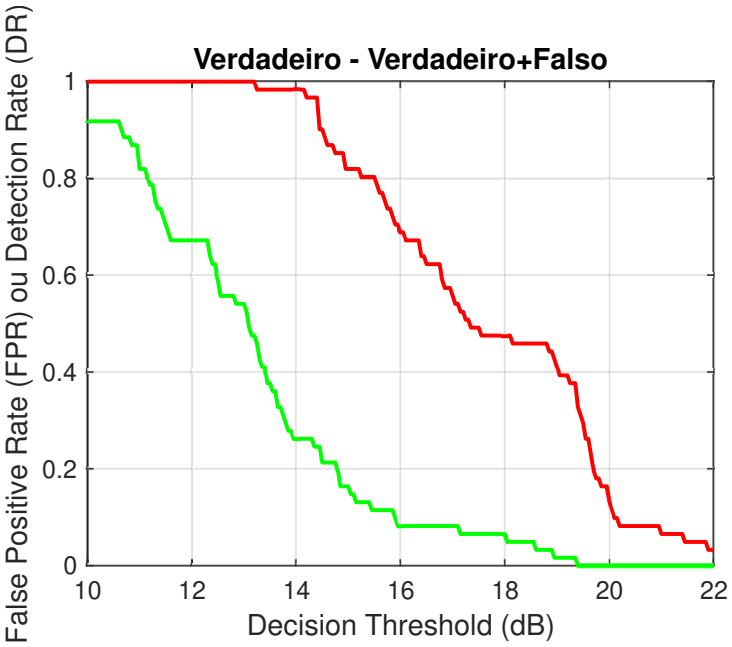
percentual da influência dos RSS do nó verdadeiro, e por consequência, um alto percentual de detecção na presença de um nó falso.

Figura 44 – Distância de 50cm com 25 amostras



Fonte: Própria

Figura 45 – Distância de 50cm com 50 amostras



Fonte: Própria

5.5.2 Distância de 1 metro

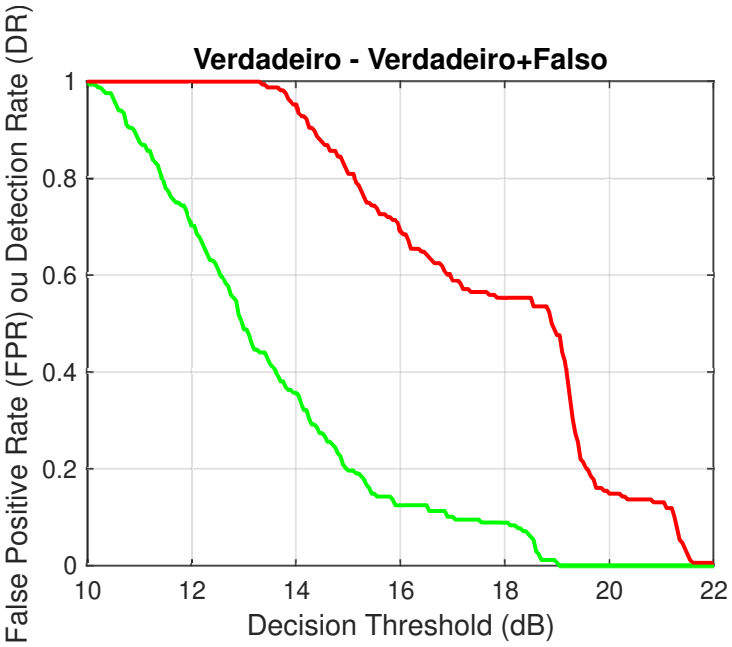
As Figuras 46 e 47 apresentam os resultados para a distância de 1 metro. O melhor threshold seria em torno de 18, pois é o melhor valor para um bom equilíbrio entre a presença do nó verdadeiro e nó falso, para ambas as janelas de 25 e 50 amostras.

Contudo, é interessante frisar que para nas duas distâncias implementadas no método proposto, consegue-se identificar a presença do nó falso na rede. Assim, esses foram os resultados obtidos, com a aplicação do método proposto, nas distâncias de cinquenta centímetro e um metro, nas dependências da Universidade Ânima em Tubarão, como mostra a Tabela 8.

Janela	**Threshold**	Detecção	FP	Dist. 50 cm	Dist. 1 m
25 RSS	14	90%	22%	X	
50 RSS	14	96%	24%	X	
25 RSS	18	97%	32%		X
50 RSS	18	92%	32%		X

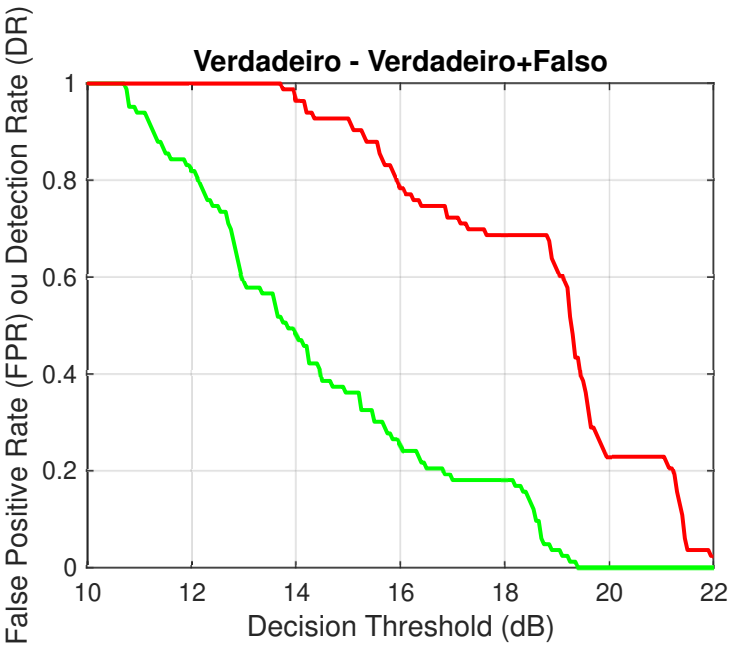
Tabela 8 – Análise de Resultados da Biblioteca Ânima

Figura 46 – Distância de 1m com 25 amostras



Fonte: Própria

Figura 47 – Distância de 1m com 50 amostras



Fonte: Própria

5.6 Residência de Veraneio em Jaguaruna

O Ambiente 06 foi definido como sendo a Residência de Veraneio no município de Jaguaruna, Santa Catarina visando analisar o comportamento do RSS numa rede WiFi, num ambiente com baixa incidência de interferência no espectro. A figura 48 mostra ambiente externo que serviu para a captação de RSS. Observou-se que o threshold de decisão mais adequado foi de 14, para um conjunto de 25 e 50 amostras, variando um pouco, mas 14 se apresenta como o mais adequado para esse cenário, para as duas distâncias analisadas. Diante disso, é prudente considerar que mesmo, nas duas distâncias ocorreram poucas variações.

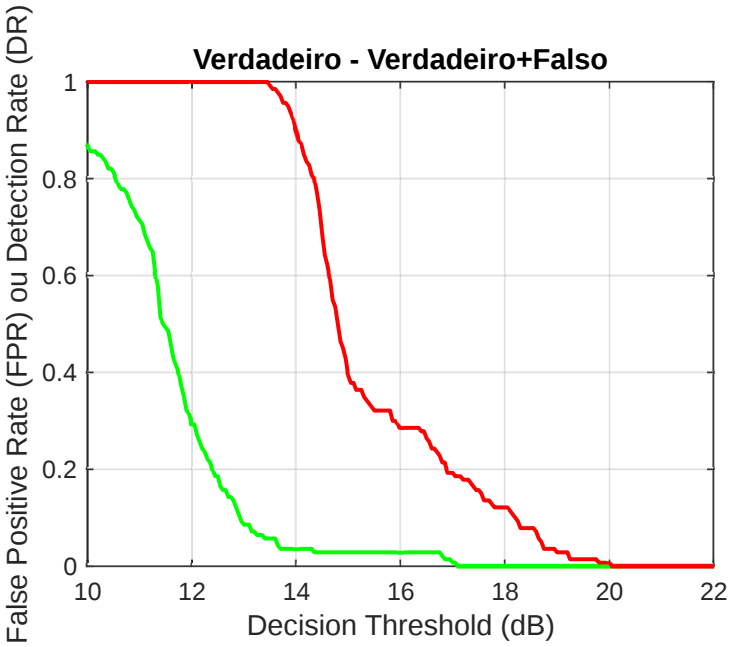
Figura 48 – Ambiente 06 - Ambiente Residencial Jaguaruna

Fonte: Autoria Própria

5.6.1 Análise das Distâncias

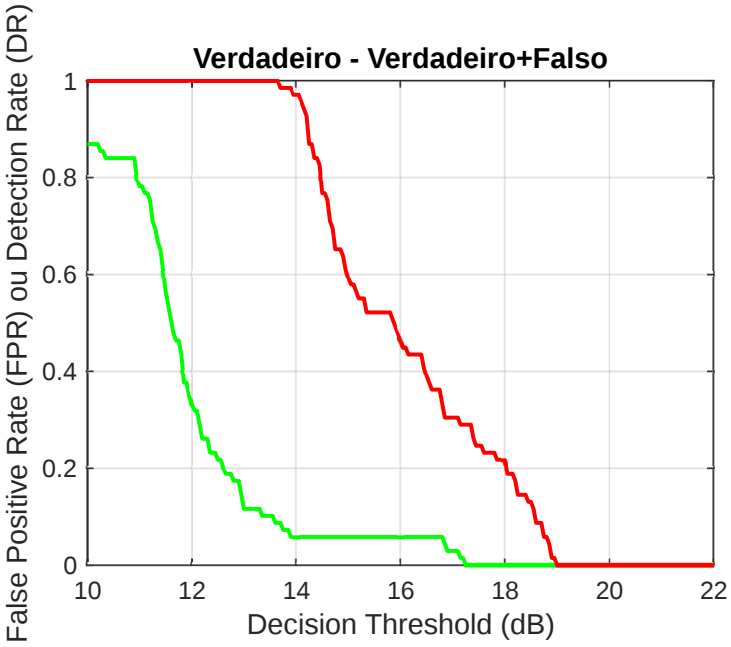
As Figuras 49 e 50 apresentam os resultados do método para uma distância de 50 cm e as Figuras 52 e 53 apresentam os resultados para a distância de 1 metro. Diante dos gráficos das Figuras 52 e 53 compreende-se que o melhor *threshold* com a presença do nó verdadeiro (linha verde) e nó falso (linha vermelha), é o 14, para um cenário de variação entre 10 e 19. Com esse *threshold*, pode-se observar um baixíssimo percentual do falso-positivo e um alto percentual de detecção, para as distâncias de 50 cm e um metro.

Figura 49 – Distância de 50 cm com 25 amostras



Fonte: Própria

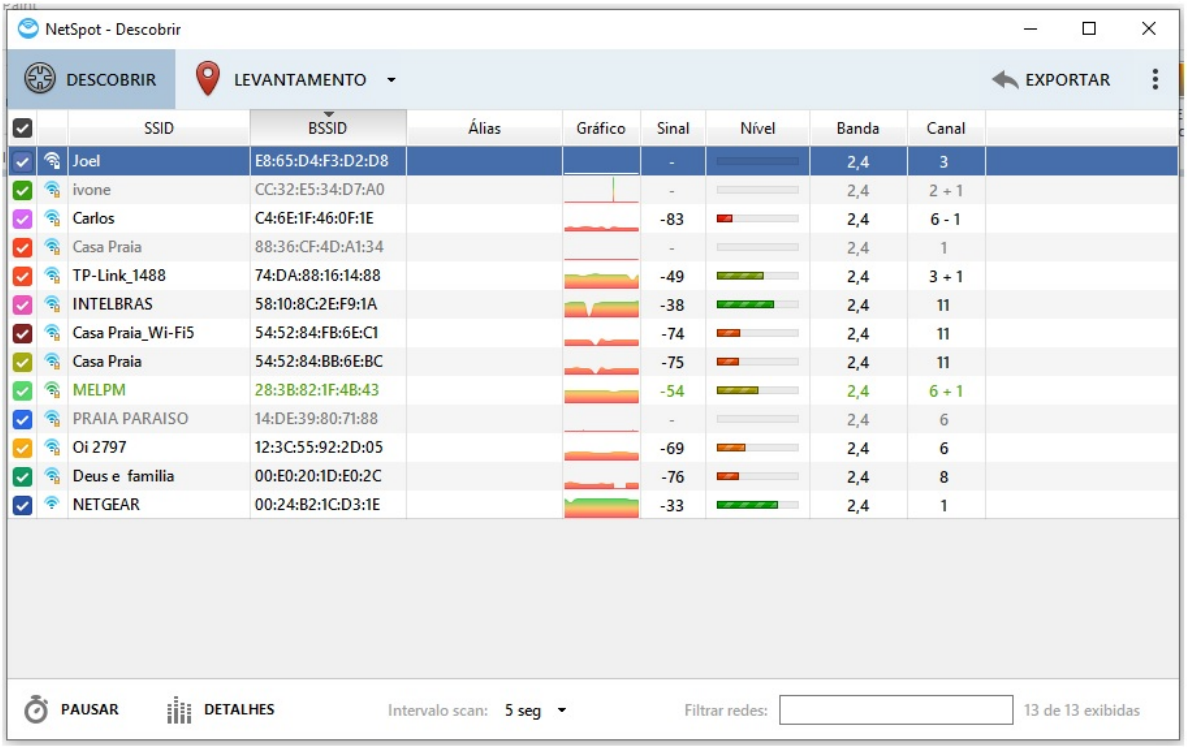
Figura 50 – Distância de 50 cm com 50 amostras



Fonte: Própria

Assim, é relevante citar que, foram captados aproximadamente 3500 RSS das redes

Figura 51 – Ambiente da Casa de Veraneio



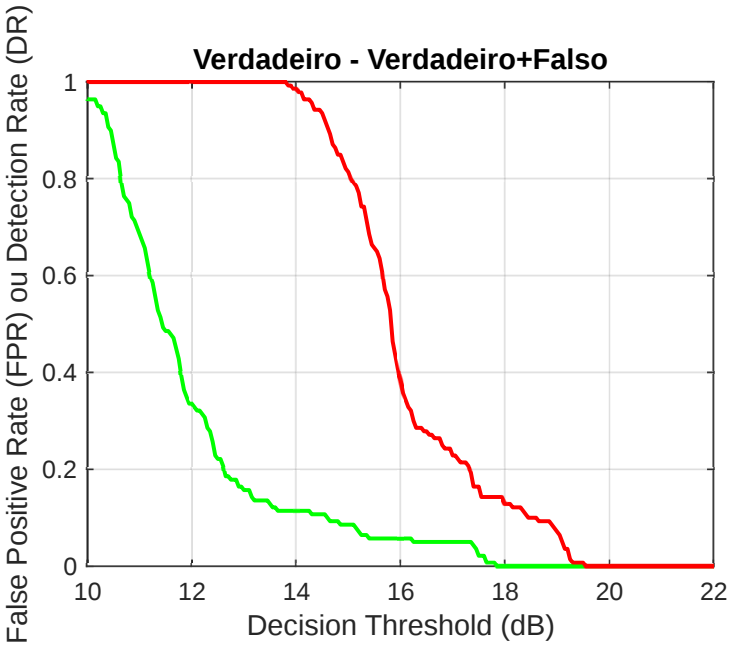
Fonte: Própria

Janela	**Threshold**	Detecção	FP	Dist. 50 cm	Dist. 1 m
25 RSS	14	88%	3%	X	
50 RSS	14	94%	7%	X	
25 RSS	14	98%	11%		X
50 RSS	14	98%	18%		X

Tabela 9 – Análise de Resultados do Ambiente 06

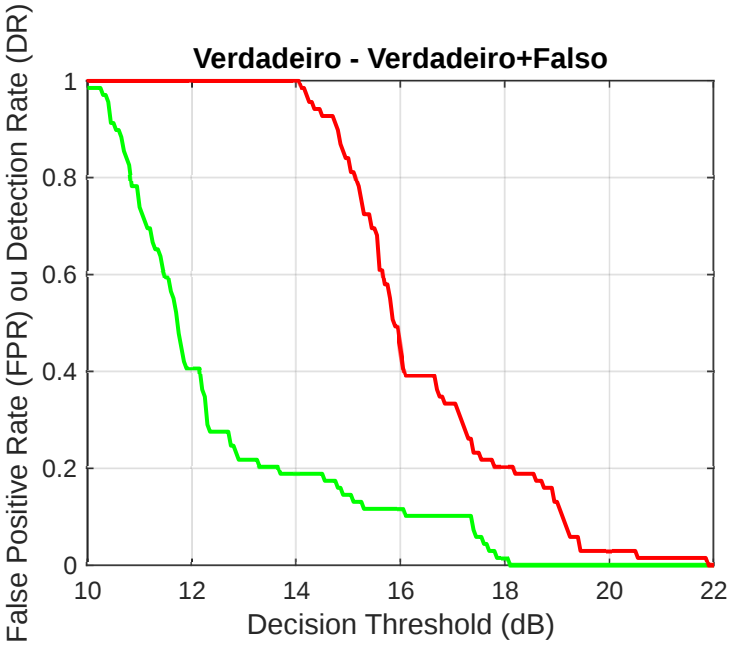
WiFi que emitiam sinal na região de localização da residência. Procurou-se definir o escaneamento a cada cinco segundos, com o uso do Netspot. Dentre os ambientes de estudo que apresentaram melhores dados para aplicação do método proposto por esta tese, foi o cenário da Residência de Veraneio, já o pior cenário foi a distância de um metro para esse cenário. Vale ressaltar, que o *threshold* apresenta uma variação referente aos ambientes anteriores, por se tratar da topologia física do ambiente de teste, a baixa presença de equipamento eletro-eletrônicos, os quais emitem sinais eletromagnéticos. Neste caso, a baixa poluição do espectro de radiofrequência influenciou positivamente no resultado final das captações de RSS.

Figura 52 – Distância de 1m com 25 amostras



Fonte: Própria

Figura 53 – Distância de 1m com 50 amostras



Fonte: Própria

5.7 Considerações Finais

Este capítulo apresentou os resultados do método proposto a partir da captação do RSS envolvendo diferentes tipos de ambientes de rede WiFi. Estes ambientes incluiram uma residência comum, um ambiente acadêmico, um centro de serviços públicos, duas residências populares, uma residência num condomínio de apartamentos e uma universidade pública. Nesses experimentos, para cada cenário, foram captados aproximadamente 3500 amostras de RSS das redes WiFi que tiveram o sinal captado pelo aplicativo Netspot. O intervalo de escaneamento foi de aproximadamente cinco segundos.

Conforme os resultados experimentais, o cenário mais favorável para aplicação do método foi o ambiente da casa de praia, devido a menor incidência de poluição do espectro. O pior cenário para análise dos dados captados nos demais ambientes foi o do residencial Murano no centro de Tubarão, para uma distância de um metro. A razão disso se deve principalmente a maior poluição do espectro.

A Tabela 10 apresenta um resumo do comportamento dos *thresholds* mais adequados para cada ambiente, considerando as duas distância analisadas na proposta, de 50 cm e 1 metro. Considerando a escolha ótima dos *thresholds*, os resultados do método proposto foram classificados como *Bom*, *Médio* e *Ruim* entre os experimentos dos cenários analisados na adoção do método proposto por esta tese. Esta classificação leva em conta a relação entre falso positivo (quando so temos a presença do nó verdadeiro) e falso negativo (quando temos a presença do nó verdadeiro e falso).

Tipo de Ambiente	**Bom**	**Médio**	**Ruim**
Ambiente 01		X	
Ambiente 02	X		
Ambiente 03		X	
Ambiente 04			X
Ambiente 05	X		
Ambiente 06	X		

Tabela 10 – Resumo dos Resultados para os Diferentes Ambientes

Contudo, é relevante citar que foram captados aproximadamente 3500 amostras de RSS das redes WiFi analisadas. Procurou-se definir o escaneamento a cada cinco segundos, com o uso do Netspot. Dentre os ambientes de estudo que apresentaram melhores dados para aplicação do método proposto, estão os Ambientes 02, 05 e 06. Considerado como resultados medianos ou médios para a aplicação do experimento: Ambiente 1 e Ambiente 3. Por último, o pior cenário: Ambiente 4. Assim, esse trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 01.

6 CONCLUSÃO E TRABALHOS FUTUROS

Com grande potencial de impactar setores econômicos e sociais, a Internet das Coisas é considerada uma revolução. Dispositivos e equipamentos tradicionais do cotidiano poderão ser interligados a Internet, gerando mais dados e informações, os quais servirão de base para novos serviços e aplicações. Este contexto de IoT associado também as aplicações de RSSF, deve aumentar significativamente o uso das redes WiFi.

Neste contexto de crescente uso das WLANs, focamos nesta pesquisa no problema de ataques de camada física, mas especificamente os ataques de *spoofing*. Propomos um modelo de detecção de nó malicioso em ambientes de rede WiFi, com o auxílio de técnicas de *machine learning* utilizando medidas de RSS, em diferentes ambientes de redes WiFi. O cenário de teste incluiu o uso de quatro APs como *landmarks*, um nó verdadeiro e a presença de um nó falso aplicando ataque de *spoofing*.

Identificamos que dependendo do tipo de ambiente, o *threshold* mais adequado para minimizar os falsos positivos e também os falsos negativos pode variar. Isso se deve as características físicas e de mobilidade de cada ambiente. Tipicamente, os valores de *threshold* variaram entre 15 e 18 para os cenários investigados. Consideramos cenários de pior caso, quando os nós verdadeiro e falso então muito próximos um do outro e a praticamente a mesma distância dos *landmarks*. Também investigamos o efeito da janela de análise das amostras de RSS utilizada pelo algoritmo proposto. Em alguns casos isso afetou a definição do valor de *threshold*.

Nos experimentos realizados nos seis locais de captação de RSS, as presenças dos nós verdadeiro e falso, se apresentou muito bem definida. Contudo em dois dos seis cenários, que foram a biblioteca de uma universidade e o ambiente de uma residência popular na cidade de Braço do Norte, foi mais difícil diferenciar a presença do nó falso. Estes casos necessitam de mais estudos para analisar porque os resultados se apresentaram tão próximos um do outro. Especificamente nesses casos, identificamos que o *threshold* necessita receber um ajuste mais refinado ou mesmo adaptativo, para facilitar a identificação do ataque.

6.1 Resumo das Contribuições

A principal contribuição desta tese foi propor e desenvolver um método simples para identificar a presença de um nó falso/malicioso realizando ataque de *spoofing* em uma rede WiFi. O método utiliza informações de RSS capturadas por APs (landmarks) que compõem a infraestrutura da rede. A estratégia proposta utiliza técnicas de aprendizagem

de máquina e foi avaliada de forma experimental em diferentes tipos de ambientes reais de rede WiFi. O método proposto apresentou bons resultados na maioria dos ambientes investigados. Assim, nossas análise indicaram que a proposta apresentada nesta tese, consegue identificar com um percentual muito grande a presença do nó falso/malicioso numa rede WiFi, sem a necessidade de implementação de tecnologias e/ou equipamentos adicionais específicos. Apenas com a captação do RSS da rede e suas variações é possível identificar a presença ou não de um nó falso, e por consequência, tomar as devidas medidas de segurança no ambiente em questão.

6.2 Trabalhos Futuros

Considerando a estratégia proposta e as análises experimentais realizadas, podemos elencar alguns pontos que podem ser explorados em trabalhos futuros:

1. Analisar o desempenho do método considerando um número variável de *landmarks*.
2. Criar uma estratégia para ponderar as medidas de RSS dos diferentes *landmarks* em função da qualidade e variabilidade das medidas obtidas.
3. Investigar o efeito da mobilidade do nó verdadeiro e do nó falso no desempenho do método.
4. Adaptar a estratégia para detectar também ataques do tipo Sybil.

REFERÊNCIAS

ADEWUMI, Omotayo G; DJOUANI, Karim; KURIEN, Anish M. Rssi based indoor and outdoor distance estimation for localization in wsn. In: IEEE. *2013 IEEE international conference on Industrial technology (ICIT)*. [S.l.], 2013. p. 1534–1539. Citado na página 25.

ANDERBERG, MR. Cluster analysis for applications accademic press. *New York and London*, 1973. Citado na página 28.

ARBAUGH, William A; SHANKAR, Narendar; WAN, YC Justin; ZHANG, Kan. Your 80211 wireless network has no clothes. *IEEE Wireless Communications*, IEEE, v. 9, n. 6, p. 44–51, 2002. Citado 2 vezes nas páginas 15 e 26.

AVAST. Spoofing. Avast, https://www.avast.com/pt-br/c-spoofing, volume, setembro 2019. Anotação. Citado na página 25.

AWAD, Abdalkarim; FRUNZKE, Thorsten; DRESSLER, Falko. Adaptive distance estimation and localization in wsn using rssi measures. In: IEEE. *10th Euromicro Conference on Digital System Design Architectures, Methods and Tools (DSD 2007)*. [S.l.], 2007. p. 471–478. Citado na página 25.

BELLARDO, John; SAVAGE, Stefan. 802.11 denial-of-service attacks: Real vulnerabilities and practical solutions. In: WASHINGTON DC. *USENIX security symposium*. [S.l.], 2003. v. 12, p. 2–2. Citado na página 26.

BERNASCHI, Massimo; FERRERI, Francesco; VALCAMONICI, Leonardo. Access points vulnerabilities to dos attacks in 802.11 networks. *Wireless Networks*, Springer-Verlag New York, Inc., v. 14, n. 2, p. 159–169, 2008. Citado 2 vezes nas páginas 15 e 26.

BOEHMKE, Brad; GREENWELL, Brandon. *Hands-on machine learning with R*. [S.l.]: Chapman and Hall/CRC, 2019. Citado na página 45.

BOUBICHE, S.; BOUBICHE, D. E.; BILAMI, A.; TORAL-CRUZ, H. Big data challenges and data aggregation strategies in wireless sensor networks. *IEEE Access*, v. 6, p. 20558–20571, 2018. Citado na página 23.

BREITENBAUCH, Henrik. Defence planning. *Academic foresights*, v. 13, 2015. Citado na página 58.

Chen, Y.; Yang, J.; Trappe, W.; Martin, R. P. Detecting and localizing identity-based attacks in wireless and sensor networks. *IEEE Transactions on Vehicular Technology*, v. 59, n. 5, p. 2418–2434, 2010. Citado 3 vezes nas páginas 16, 20 e 36.

CHEN, Yingying; YANG, Jie; TRAPPE, Wade; MARTIN, Richard P. Detecting and localizing identity-based attacks in wireless and sensor networks. *IEEE Transactions on Vehicular Technology*, IEEE, v. 59, n. 5, p. 2418–2434, 2010. Citado 13 vezes nas páginas 23, 30, 38, 39, 40, 41, 42, 43, 45, 48, 49, 50 e 51.

CORMACK, Richard M. A review of classification. *Journal of the Royal Statistical Society: Series A (General)*, Wiley Online Library, v. 134, n. 3, p. 321–353, 1971. Citado na página 29.

DAN, Avishek; HALDER, Subir; DASBIT, Sipra. Localization with enhanced location accuracy using rssi in wsn. In: IEEE. *2011 Fifth IEEE International Conference on Advanced Telecommunication Systems and Networks (ANTS)*. [S.l.], 2011. p. 1–6. Citado na página 24.

DEMIRBAS, Murat; SONG, Youngwhan. An rssi-based scheme for sybil attack detection in wireless sensor networks. In: IEEE. *2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'06)*. [S.l.], 2006. p. 5–pp. Citado 2 vezes nas páginas 15 e 28.

DING, Chris; HE, Xiaofeng. K-means clustering via principal component analysis. In: ACM. *Proceedings of the twenty-first international conference on Machine learning*. [S.l.], 2004. p. 29. Citado na página 29.

DJEDOUBOUM, Asside Christian; ARI, Ado Adamou Abba; GUEROUI, Abdelhak Mourad; MOHAMADOU, Alidou; ALIOUAT, Zibouda. Big data collection in large-scale wireless sensor networks. *Sensors*, v. 18, n. 12, 2018. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/18/12/4474>. Citado na página 21.

DOUCEUR, John R. The sybil attack. In: SPRINGER. *International workshop on peer-to-peer systems*. [S.l.], 2002. p. 251–260. Citado na página 28.

DUBES, Richard; JAIN, Anil K. Validity studies in clustering methodologies. *Pattern recognition*, Elsevier, v. 11, n. 4, p. 235–254, 1979. Citado na página 28.

DUDA, RO; HART, PE; STORK, DG. Pattern classification. 2nd edn wiley. *New York*, v. 153, 2000. Citado na página 29.

FRIEDMAN, Jerome; HASTIE, Trevor; TIBSHIRANI, Robert. *The elements of statistical learning*. [S.l.]: Springer series in statistics New York, 2001. v. 1. Citado na página 29.

FUKUSHIMA, Takeru; MURAKAMI, Tomoki; ABEYSEKERA, Hirantha; FUJIHASHI, Takuya; WATANABE, Takashi; SARUWATARI, Shunsuke. Feasibility study of practical aoa estimation using compressed csi on commercial wlan devices. *IEEE Access*, IEEE, v. 10, p. 49128–49141, 2022. Citado 3 vezes nas páginas 16, 22 e 33.

GHOLAMI, Rozita; HODTANI, Ghosheh Abed. A more general information theoretic study of wireless location verification system. *IEEE Transactions on Vehicular Technology*, IEEE, v. 69, n. 9, p. 9938–9950, 2020. Citado na página 35.

GOLDSMITH, Andrea. *Wireless communications*. [S.l.]: Cambridge university press, 2005. Citado 4 vezes nas páginas 23, 24, 39 e 46.

GUDMUNDSON, Mikael. Correlation model for shadow fading in mobile radio systems. *Electronics letters*, IET, v. 27, n. 23, p. 2145–2146, 1991. Citado na página 24.

HAMDAN, Omar; SHANABLEH, Hassan; ZAKI, Inas; AL-ALI, AR; SHANABLEH, Tamer. Iot-based interactive dual mode smart home automation. In: IEEE. *2019 IEEE International Conference on Consumer Electronics (ICCE)*. [S.l.], 2019. p. 1–2. Citado na página 22.

HARTIGAN, John A; WONG, Manchek A. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, JSTOR, v. 28, n. 1, p. 100–108, 1979. Citado na página 29.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome; FRANKLIN, James. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, Springer, v. 27, n. 2, p. 83–85, 2005. Citado na página 44.

HEURTEFEUX, Karel; VALOIS, Fabrice. Is rssi a good choice for localization in wireless sensor network? In: IEEE. *2012 IEEE 26th international conference on advanced information networking and applications*. [S.l.], 2012. p. 732–739. Citado na página 25.

HOLGER, Karl; WILLIG., Andreas. *Protocols and architectures for wireless sensors networks*. [S.l.]: John Wiley & Sons, 2005. Citado na página 21.

HUAN, Xintao; KIM, Kyeong Soo; ZHANG, Junqing. Nisa: Node identification and spoofing attack detection based on clock features and radio information for wireless sensor networks. *IEEE Transactions on Communications*, IEEE, v. 69, n. 7, p. 4691–4703, 2021. Citado 4 vezes nas páginas 16, 17, 20 e 33.

HUANG, Zhexue. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery*, Springer, v. 2, n. 3, p. 283–304, 1998. Citado na página 28.

IOVA, O.; PICCO, P.; ISTOMIN, T.; KIRALY, C. Rpl: The routing standard for the internet of things... or is it? *IEEE Communications Magazine*, v. 54, n. 12, p. 16–22, December 2016. ISSN 0163-6804. Citado na página 23.

JAGANNATH, Jithin; POLOSKY, Nicholas; JAGANNATH, Anu; RESTUCCIA, Francesco; MELODIA, Tommaso. Machine learning for wireless communications in the internet of things: A comprehensive survey. *Ad Hoc Networks*, Elsevier BV, v. 93, p. 101913, Oct 2019. ISSN 1570-8705. Citado na página 21.

JAIN, Anil K; DUBES, Richard C. Algorithms for clustering data. *Englewood Cliffs: Prentice Hall, 1988*, 1988. Citado na página 29.

JAWANDHIYA, Pradip M; GHONGE, Mangesh M; ALI, MS; DESHPANDE, JS. A survey of mobile ad hoc network attacks. *International Journal of Engineering Science and Technology*, v. 2, n. 9, p. 4063–4071, 2010. Citado na página 20.

KARLOF, Chris; WAGNER, David. Secure routing in wireless sensor networks: Attacks and countermeasures. In: IEEE. *Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications, 2003.* [S.l.], 2003. p. 113–127. Citado na página 28.

KAUFMAN, Leonard; ROUSSEEUW, Peter J. *Finding groups in data: an introduction to cluster analysis*. [S.l.]: John Wiley & Sons, 2009. v. 344. Citado na página 28.

KHAN, Imran; BELQASMI, Fatna; GLITHO, Roch; CRESPI, Noel; MORROW, Monique; POLAKOS, Paul. Wireless sensor network virtualization: early architecture and research perspectives. *IEEE Network*, IEEE, v. 29, n. 3, p. 104–112, 2015. Citado na página 15.

KLÖSGEN, Willi; ŻYTKOW, Jan M. Knowledge discovery in databases terminology. In: AMERICAN ASSOCIATION FOR ARTIFICIAL INTELLIGENCE. *Advances in knowledge discovery and data mining*. [S.l.], 1996. p. 573–592. Citado na página 29.

LANTZ, Brett. *Machine learning with R*. [S.l.]: Packt Publishing Ltd, 2013. Citado 2 vezes nas páginas 28 e 45.

LINDEN, Ricardo. Técnicas de agrupamento. *Revista de Sistemas de Informação da FSMA*, n, v. 4, n. 4, p. 18–36, 2009. Citado na página 40.

LIU, Yinghong; WU, Yuanming. An enhanced rssi-based detection scheme for sybil attack in wireless sensor networks. In: SPRINGER. *Future of Information and Communication Conference*. [S.l.], 2019. p. 87–102. Citado 2 vezes nas páginas 27 e 28.

LLOYD, Sarah. Least squares quantization in pcm's: Bell telephone laboratories paper. *Murray Hill*, 1957. Citado na página 29.

LOURENÇO, André Figueira. Webspy: uma aplicação de monitoramento web em tempo real. 2013. Citado na página 25.

MACQUEEN, James et al. Some methods for classification and analysis of multivariate observations. In: OAKLAND, CA, USA. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. [S.l.], 1967. v. 1, n. 14, p. 281–297. Citado na página 29.

MEKKI, Kais; BAJIC, Eddy; CHAXEL, Frederic; MEYER, Fernand. A comparative study of lpwan technologies for large-scale iot deployment. *ICT express*, Elsevier, v. 5, n. 1, p. 1–7, 2019. Citado na página 22.

MONGIOVI, Misael; BOGDANOV, Petko; RANCA, Razvan; PAPALEXAKIS, Evangelos E; FALOUTSOS, Christos; SINGH, Ambuj K. Netspot: Spotting significant anomalous regions on dynamic networks. In: SIAM. *Proceedings of the 2013 Siam international conference on data mining*. [S.l.], 2013. p. 28–36. Citado 2 vezes nas páginas 53 e 58.

NAWIR, Mukrimah; AMIR, Amiza; YAAKOB, Naimah; LYNN, Ong Bi. Internet of things (iot): Taxonomy of security attacks. In: IEEE. *2016 3rd International Conference on Electronic Design (ICED)*. [S.l.], 2016. p. 321–326. Citado na página 22.

PEI, Chengcheng; ZHANG, Ning; SHEN, Xuemin Sherman; MARK, Jon W. Channel-based physical layer authentication. In: IEEE. *2014 IEEE Global Communications Conference*. [S.l.], 2014. p. 4114–4119. Citado 2 vezes nas páginas 16 e 31.

PÉREZ, Salvador; MARTÍNEZ, Juan A; SKARMETA, Antonio F; MATEUS, Márcio; ALMEIDA, Bruno; MALÓ, Pedro. Armour: Large-scale experiments for iot security & trust. In: IEEE. *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*. [S.l.], 2016. p. 553–558. Citado na página 22.

PINTO, Eliel Marlon de Lima; LACHOWSKI, Rosana; PELLENZ, Marcelo Eduardo; PENNA, Manoel Camillo; SOUZA, Richard Demo. A machine learning approach for detecting spoofing attacks in wireless sensor networks. In: IEEE. *2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)*. [S.l.], 2018. p. 752–758. Citado na página 55.

RAJAN, Anjana; JITHISH, J; SANKARAN, Sriram. Sybil attack in iot: Modelling and defenses. In: IEEE. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. [S.l.], 2017. p. 2323–2327. Citado na página 22.

RANI, S.; AHMED, S. H.; TALWAR, R.; MALHOTRA, J. Can sensors collect big data? an energy-efficient big data gathering algorithm for a wsn. *IEEE Transactions on Industrial Informatics*, v. 13, n. 4, p. 1961–1968, Aug 2017. ISSN 1551-3203. Citado na página 21.

RATASUK, Rapeepat; MANGALVEDHE, Nitin; GHOSH, Amitava. Overview of lte enhancements for cellular iot. In: IEEE. *2015 IEEE 26th annual international symposium on personal, indoor, and mobile radio communications (PIMRC)*. [S.l.], 2015. p. 2293–2297. Citado na página 22.

RIBEIRO, Bruno C Dias; JUNIOR, Edson Floriano S; ARANHA, Diego F. Análise de segurança da distribuição de raízes na icp-brasil. In: SBC. *Anais do XVII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais*. [S.l.], 2017. p. 549–556. Citado na página 16.

SHADEED, Mohammad; MOREB, Mohammed. Lightweight encryption for multimedia in the internet of thing (iot). In: IEEE. *2021 International Conference on Information Technology (ICIT)*. [S.l.], 2021. p. 27–32. Citado na página 22.

SHENG, Yong; TAN, Keren; CHEN, Guanling; KOTZ, David; CAMPBELL, Andrew. Detecting 802.11 mac layer spoofing using received signal strength. In: IEEE. *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*. [S.l.], 2008. p. 1768–1776. Citado na página 15.

SHIHUAN, WANG. Applications analysis of iot based on rfid and wlan technologies [j]. *Information and Electronic Engineering*, v. 5, p. 604–606, 2010. Citado na página 15.

SSU, Kuo-Feng; WANG, Wei-Tong; CHANG, Wen-Chung. Detecting sybil attacks in wireless sensor networks using neighboring information. *Computer Networks*, Elsevier, v. 53, n. 18, p. 3042–3056, 2009. Citado na página 28.

STÜBER, Gordon L; STÈUBER, Gordon L. *Principles of mobile communication*. [S.l.]: Springer, 1996. v. 2. Citado na página 24.

SUJATHA, V; ANITA, EA Mary. An efficient trust based method for sybil node detection in mobile wireless sensor network. In: AIP PUBLISHING. *AIP Conference Proceedings*. [S.l.], 2018. v. 2016, n. 1, p. 020138. Citado na página 27.

TAKAISHI, D.; NISHIYAMA, H.; KATO, N.; MIURA, R. Toward energy efficient big data gathering in densely distributed sensor networks. *IEEE Transactions on Emerging Topics in Computing*, v. 2, n. 3, p. 388–397, Sept 2014. ISSN 2168-6750. Citado na página 21.

TANG, Jie; JIAO, Long; ZENG, Kai; WEN, Hong; GOVINDAN, Kannan; WU, Daniel; MOHAPATRA, Prasant. Identity-based attack detection and classification utilizing reciprocal rss variations in mobile wireless networks. *IEEE Transactions on Mobile Computing*, IEEE, 2020. Citado 2 vezes nas páginas 16 e 32.

WANG, Hui; MA, Li; BAI, Hong-ying. A three-tier scheme for sybil attack detection in wireless sensor networks. In: IEEE. *2020 5th International Conference on Computer and Communication Systems (ICCCS)*. [S.l.], 2020. p. 752–756. Citado na página 20.

WANG, Qunke; FANG, Lanting; ZHU, Zhenchao; HUANG, Jie. Detection algorithm of the mimicry attack based on variational auto-encoder. In: IEEE. *2021 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*. [S.l.], 2021. p. 114–120. Citado 2 vezes nas páginas 16 e 34.

WEITZEN, Jay; LOWE, Terri J. Measurement of angular and distance correlation properties of log-normal shadowing at 1900 mhz and its application to design of pcs systems. *IEEE transactions on vehicular technology*, IEEE, v. 51, n. 2, p. 265–273, 2002. Citado na página 24.

WU, Yu-Chun; CHOW, Chi-Wai; LIU, Yang; LIN, Yun-Shen; HONG, Chong-You; LIN, Dong-Chang; SONG, Shao-Hua; YEH, Chien-Hung. Received-signal-strength (rss) based 3d visible-light-positioning (vlp) system using kernel ridge regression machine learning algorithm with sigmoid function data preprocessing method. *IEEE Access*, IEEE, v. 8, p. 214269–214281, 2020. Citado na página 16.

XIAO, Yingyuan; JIAO, Xu; WANG, Hongya; HSU, Ching-Hsien; LIU, Li; ZHENG, Wenguang. Efficient continuous skyline query processing in wireless sensor networks. *Sensors*, v. 19, p. 2902, 06 2019. Citado 2 vezes nas páginas 20 e 23.

XIE, Ning; LI, Zhuoyuan; TAN, Haijun. A survey of physical-layer authentication in wireless communications. *IEEE Communications Surveys & Tutorials*, IEEE, v. 23, n. 1, p. 282–310, 2020. Citado 2 vezes nas páginas 16 e 34.

XU, Jiuqiang; LIU, Wei; LANG, Fenggao; ZHANG, Yuanyuan; WANG, Chenglong. Distance measurement model based on rssi in wsn. *Wireless Sensor Network*, Scientific Research Publishing, v. 2, n. 08, p. 606, 2010. Citado na página 25.

YANG, Jie; CHEN, Yingying; TRAPPE, Wade; CHENG, Jerry. Detection and localization of multiple spoofing attackers in wireless networks. *IEEE Transactions on Parallel and Distributed systems*, Ieee, v. 24, n. 1, p. 44–58, 2012. Citado 3 vezes nas páginas 16, 31 e 36.

YAQIONG, LV; LEI, TU; LEE, Carman KM; XIN, TANG. Iot based omni-channel logistics service in industry 4.0. In: IEEE. *2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*. [S.l.], 2018. p. 240–243. Citado na página 17.

YU, Haifeng; KAMINSKY, Michael; GIBBONS, Phillip B; FLAXMAN, Abraham. Sybilguard: defending against sybil attacks via social networks. *ACM SIGCOMM Computer Communication Review*, ACM, v. 36, n. 4, p. 267–278, 2006. Citado na página 27.

ZANELLA, Andrea; BUI, Nicola; CASTELLANI, Angelo; VANGELISTA, Lorenzo; ZORZI, Michele. Internet of things for smart cities. *IEEE Internet of Things journal*, Ieee, v. 1, n. 1, p. 22–32, 2014. Citado na página 20.

ZHANG, Junqing; LI, Guyue; MARSHALL, Alan; HU, Aiqun; HANZO, Lajos. A new frontier for iot security emerging from three decades of key generation relying on wireless channels. *IEEE Access*, IEEE, v. 8, p. 138406–138446, 2020. Citado na página 35.

ZHANG, Zhi-Kai; CHO, Michael Cheng Yi; WANG, Chia-Wei; HSU, Chia-Wei; CHEN, Chong-Kuan; SHIEH, Shiuhpyng. Iot security: ongoing challenges and research opportunities. In: IEEE. *2014 IEEE 7th international conference on service-oriented computing and applications*. [S.l.], 2014. p. 230–234. Citado na página 22.

ZHENG, Bojun; MASUDA, Takefumi; SHIBATA, Tsugumichi. An indoor positioning with a neural network model of tensorflow for machine learning. In: IEEE. *2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*. [S.l.], 2021. p. 1–2. Citado na página 16.

ZOU, Yulong; ZHU, Jia; WANG, Xianbin; HANZO, Lajos. A survey on wireless security: Technical challenges, recent advances, and future trends. *Proceedings of the IEEE*, IEEE, v. 104, n. 9, p. 1727–1765, 2016. Citado na página 20.