Revista de Sistemas de Informação da FSMA
n. 23 (2019) pp. 40-49

SISTEMAS DE INFORMAÇÃO
http://www.fsma.edu.br/si/sistemas.html

TRILHA PRINCIPAL ERI·GO

Autenticação de dispositivos *IoT* baseada nas características do sinal eletromagnético em redes sem fio heterogêneas

Victor Hugo L. Lopes [1] [2], Marcos F. B de Abreu [2], Pabllo F. A. Sousa [2],
Antônio C. de O. Júnior [2] [3], Vinicius da C. M. Borges [2], Kleber V. Cardoso [2]
[1] *NETI - NumbERS (IFG) - Inhumas - Goiás*
[2] *Instituto de Informática (UFG) - Goiânia - Goiás*
[3] *Fraunhofer Portugal AICOS - Porto - Portugal*

***Resumo*—A Internet das Coisas se baseará na coexistência de dispositivos e sensores que implementam diversas tecnologias *wireless*, com diferentes capacidades computacionais, acarretando em novos desafios de segurança. Baseado neste desafio, o presente trabalho apresenta o ADISEH, um método para autenticação de dispositivos *IoT* através das características do sinal eletromagnético em redes sem fio heterogêneas, contribuindo para a segurança nestes novos ambientes. Simulações foram realizadas para a avaliação do seu uso com dispositivos de *ZigBee*, *Bluetooth* e *WiFi*, considerando diversas situações de ruído. Os resultados permitem comprovar a sua viabilidade, tanto para a situação em que um dispositivo autorizado tenta obter acesso à rede, quanto para aquela em que um dispositivo malicioso tenta obter acesso clonando o sinal do dispositivo autorizado.**

***Palavras-chave*—*Gateway IoT*, ADISEH, Autenticação, Máquina de Vetores de Suporte, Rádio definido por Software.**

IoT devices authentication based on the electromagnetic signal characteristics in heterogeneous networks

***Abstract*—IoT will be based on the coexistence of devices and sensors that implements diverse wireless technologies, with different computational capacities, new security challenges arise. Hence, this work presents the ADISEH, a method for the authentication of IoT devices wich uses the characteristics of the electromagnetic signal in heterogeneous wireless networks, rendering these new environments safer. Simulations were performed to evaluate their use with ZigBee, Bluetooth and WiFi devices, considering different noise situations. The results provide evidence of its viability, both when an authorized device attempts to gain access to the network and when a malicious device tries to gain access by cloning the signal from the authorized device.**

***Index Terms*—Gateway IoT, Authentication, ADISEH, Support Vector Machine, Software Defined Radio.**

I. Introdução

INTERNET das Coisas, do inglês *Internet of Things (IoT)*, é um paradigma descrito como um conceito computacional que aponta para ambientes em que os dispositivos físicos terão conectividade com a Internet, de forma autônoma e segura [1]. A *IoT* está ganhando espaço na área das comunicações sem fio modernas devido ao seu grande impacto na vida da população, com sua aplicação para usos industriais e comerciais, automação residencial, monitoramento ambiental, segurança pública e trânsito, por exemplo.

Estimativas apontam que, em 2013, 9 bilhões de dispositivos (coisas) já estavam conectados à Internet, com previsão de se tornar uma tecnologia massiva e pervasiva, com expectativas de se chegar a mais de 20 bilhões de dispositivos já em 2020 [2]. Tais dispositivos devem ultrapassar a quantidade de computadores pessoais e telefones móveis em grande ordem de magnitude [3], de forma a se vislumbrar um complexo ecossistema destes objetos implementando diversas tecnologias de comunicação sem fio, gerando novos desafios.

Dentre estes, observam-se as questões envolvendo a segurança dos dispositivos e da própria rede à qual estejam conectados, de forma que padrões e especificações de segurança se fazem importantes [4]. Em especial, conforme foco deste trabalho, a segurança envolvendo a autenticação e acesso ao meio, em que qual tais "coisas" estão vulneráveis precisa ser devidamente considerada, levando-se em consideração a complexidade nestes novos ambientes, as restrições de recursos computacionais de diversos destes dispositivos, a escalabilidade, mobilidade e heterogeneidade [5].

Dadas tais características, algumas soluções de segurança para autenticação e acesso ao meio podem se mostrar inviáveis, o que deve ampliar a brechas de segurança da rede, ou impedir o uso de uma maior diversidade de dispositivos.

Um ataque bastante conhecido em redes de computado-

res, que se repete em *IoT* é o roubo de identidade [6]. Muitos destes dispositivos limitados não possuem mecanismos de segurança mais sofisticados, que ultrapassem a camada de enlace de dados, em que uma solução de autenticação destes dispositivos sem fio é o uso de identificadores únicos, como endereço MAC (*Media Access Control*). Isto posto, sabido que esses identificadores podem ser facilmente forjados por um invasor, abrindo possibilidades para execução de ataques.

Neste contexto, há a necessidade de se prover mecanismos que permitam que tais dispositivos possam ter acesso ao meio, isto é, à rede a qual pertencem, realizando suas comunicações, considerando o que foi exposto, permitindo que a heterogeneidade e interoperabilidade sejam habilitadas, sem comprometer a segurança da rede.

Uma solução plausível é tratar a segurança em nível de *hardware*, como proposto em [3], mas isso geraria uma dependência de construção de novos dispositivos, além da necessidade da adoção de novos padrões pela indústria, eventualmente elevando os preços de dispositivos que precisam ter baixo custo.

Uma outra alternativa, portanto, é levar a responsabilidade pela autenticação destes dispositivos para um elemento da rede com mais poder computacional, como um *Gateway*, por exemplo. Tal abordagem se torna viável após a proposição da arquitetura dos rádios definidos por software (SDR, do inglês *Software Defined Radio*) [7].

A ideia de um *Gateway* que conecte vários dispositivos de *IoT* e ofereça, através de uma solução centralizada, segurança e o processamento necessário para isso vem evoluindo [8]–[11]. Considerando-se este *Gateway* baseado em SDR, com a habilidade de se ajustar para operar com diferentes frequências e modulações, a resolução dos requisitos de segurança aqui considerados mostra-se promissora.

Dispositivos sem fio podem ser diferenciados ou agrupados com base em sua assinatura eletromagnética [12]–[14], na camada física. Tal diferenciação se dá por diversos fatores que geram imperfeições não propositais, como os que ocorrem nos processos de fabricação dos módulos e circuitos eletrônicos envolvidos na comunicação sem fio, influenciando a geração e/ou a transmissão do sinal. Estas imperfeições geram uma assinatura inequívoca, nos moldes de uma impressão digital.

Assim, este trabalho propõe a identificação de dispositivos de *IoT* com base nas impressões digitais observáveis no sinal eletromagnético emitido por eles, de forma a habilitar o uso da técnica para a construção de um processo de autenticação destes dispositivos, mesmo que estes implementem diferentes tecnologias de comunicação sem fio, conforme requerido para um *Gateway* para comunicação *IoT*.

Embora a proposta explore as características intrínsecas dos dispositivos e mereça um conjunto de testes físicos, como trabalho preliminar, procuramos apresentar um estudo teórico com algumas tecnologias para a avaliação do método proposto, analisando o impacto da interferência na correta identificação destas tecnologias, e obter evidências sobre os ajustes necessários na quantidade de amostras de predição.

Para além da introdução, este trabalho está organizado da seguinte maneira: a seção II apresenta os trabalhos relacionados. A seção III descreve os objetivos gerais e específicos deste trabalho. A seção IV descreve o método empregado e a proposta para a autenticação de dispositivos IoT. Os resultados são apresentados e discutidos na seção V enquanto que as conclusões se encontram na seção VI.

II. Trabalhos Relacionados

REALIZAR a identificação de dispositivos com base nas características eletromagnéticas do sinal não é um tema novo. Na segunda guerra mundial controladores de voo analisavam visualmente ondas dos sinais emitidos por radares para identifiar possíveis transmissões que não vinham de seus dispositivos, baseando-se nas características visualmente observáveis das transmissões.

Os primeiros trabalhos computacionais na área surgem em 1995, como apresentado por Choe et al. [15], que aplicou o processamento *offline* dos dados do sinal transiente de 3 dispositivos do tipo *walkie talkie* em modulação FM com uso de redes neurais artificiais, com apresentação de bons resultados, mas restringindo-se à identificação apenas dos dispositivos de um mesmo fabricante. De forma semelhante, Toonstra e Kinsner [16] empregaram método equivalente, mas com análise computacional do sinal transiente em frequência de 147,84 MHz, resultando em um modelo muito ineficiente em baixa SNR.

Diversos autores já apresentaram propostas para a identificação de dispositivos sem fio através da observação da chamada "impressão digital" contida no sinal eletromagnético. Entretanto, como observado por Zhao e Ge [17] e Mahmoud et al. [18], tais proposições precisam ser validadas em diferentes tecnologias de comunicação sem fio.

A identificação de *NICs* (*Network Interface Card*) em *frames* IEEE 802.11 é proposta por Brik e Banerjee [19]. Dispositivos de redes de sensores sem fio foram identificados no trabalho de Rasmussen e Capkun [20]. Método para a identificação de dispositivos *Bluethooth* foi proposto por Hall, Barbeau e Kranakis [21], [22]. Apesar de apresentarem resultados válidos em determinados cenários, nenhum destes métodos considera múltiplas tecnologias de comunicação sem fio.

Já Bihl, Bauer e Temple [23] propõem um algoritmo de identificação de dispositivos *ZigBee* usando-se as características como Amplitude, Frequência e Fase, e outro trabalho dos mesmos autores [24], que apresentam um esquema colaborativo com vários sensores, e inclui um centro de fusão para tratar das características extraídas, criando um *ranking*, novamente aplicando somente a *ZigBee*.

Abordagens propostas baseadas na inclusão de recursos específicos no transmissor, ou que geram necessidade de *hardware* especializados, com maior capacidade computacional, também não contribuem para o propósito da conexão de dispositivos heterogêneos em um único *Gateway*, dado que impedem a sua adoção pelos dispositivos de

menor capacidade computacional. A inclusão de uma *tag* em forma de um sinal de baixa potência a ser somado com o sinal a ser transmitido, gerando uma assinatura única, facilitando sua identificação e extração é proposta por Verma, Yu e Sadler [14]. Uma prova de conceito foi feita usando um par de SDRs, em que a vialibidade foi comprovada. Entretanto, tal abordagem exigirá que tanto o transmissor quanto o receptor ofereçam suporte, o que excluirá uma grande quantidade de dispositivos *IoT*.

Diferentemente dos trabalhos descritos, que empregam inteligência artificial (incluindo aprendizado de máquina) em dados já conhecidos (conhecimento *a priori*), e assumindo um número conhecido de dispositivos, um limitante para situações reais, Niguyen et al. [25] propõem o uso de algoritmo Bayesiano não parametrizado para a criar grupos de usuários, permitindo descobrir se algum ataque está ocorrendo (clone de *MAC*, por exemplo). Tal técnica também foi empregada em *ZigBee*, e apesar de ter apresentado bons resultados, limita-se a uma pequena quantidade de usuários, sendo que os dispositivos precisam ter as mesmas configurações (mesmo canal, largura de banda, potência de transmissão e frequência da portadora, e mesmo tamanho de *payload* do pacote), e se faz necessário uso de informações do administrador da rede (quais endereços *MAC* estão em uso, por exemplo).

Método que coleta 1.024 características dos sinais de dispositivos com uso de técnicas matemáticas (FFT - *Fast Fourier Transform*, EVM - *Error Vector Magnitude* e PCA - *Principal Component Analysis*), é proposto por Abreu et al. [26], gerando bons resultados no processo de identificação de dispositivos, com taxas de descoberta acima de 90%. Embora o trabalho considere a sua aplicação em um *Gateway* para *IoT*, sendo este desenvolvido a partir de um SDR, levou-se em consideração somente 1 tecnologia em sua validação, módulos *WiFi* nrf24L01 em 2,4 GHz. Desta forma, o presente trabalho visa estender o referido método, validando-o em outras tecnologias de comunicação sem fio.

III. Objetivos

TENDO em vista as potencialidades do uso da identificação das imperfeições nos sinais eletromagnéticos como mecanismo de autenticação dos dispositivos de comunicação *wireless*, o presente trabalho objetivou aplicar técnicas para extração de tais imperfeições, seguido do armazenamento destas em um banco de dados, de forma a permitir a comparação futura, visando checar se o dispositivo tem direito de comunicação com o *gateway* para *IoT*. Para tal, é empregado método para extração das características, incluindo as observadas em [26], mas não se restringindo a elas.

A. Objetivo Geral

O objetivo geral do trabalho é permitir a identificação de dispositivos únicos, ou grupos de dispositivos de uma mesma classe (mesmo fabricante, por exemplo), em ambientes em que múltiplos dispositivos *wireless* de diferentes tecnologias se comunicam com um mesmo *gateway*, permitindo a criação de técnicas para que tais dispositivos possam ser devidamente autenticados nesta rede.

B. Objetivos Específicos

Com o foco em observar o comportamento do método aqui empregado em situações com múltiplas tecnologias de comunicação, como requerido pelo *gateway IoT*, são levadas em consideração as tecnologias de comunicação sem fio *ZigBee* (IEEE 802.15.4), *Bluetooth* (IEEE 802.15) e *WiFi* (IEEE 802.11), sempre considerando tais comunicações em ambiente com ruído AWGN (do inglês *Additive White Gaussian Noise*), como nos casos de ambientes com outros dispositivos comunicando, gerando impactos na comunicação.

Por fim, o trabalho avalia o desempenho do método para autenticação considerando múltiplos dispositivos comunicando ao mesmo tempo, em que os sinais misturados são capturados pelo *gateway IoT*, dificultando o processo de identificação das características de interesse. Para tal, o método dos Mínimos Quadrados (LS, do inglês *Least Squares*) é adotado como técnica de separação do sinal.

Também faz parte dos objetivos do trabalho a utilização de um método de geração de características implementado para simulações em ambiente computacional (Matlab ®), permitindo a simulação de dispositivos com as imperfeições de interesse.

Os objetivos do trabalho podem ser sumarizados como:

1) Aplicar método para identificação de dispositivos *wireless* para autenticação em *gateway IoT* (Objetivo Principal);
2) Realizar a identificação para autenticação de dispositivos *ZigBee*;
3) Realizar a identificação para autenticação de dispositivos *Bluetooth*;
4) Realizar a identificação para autenticação de dispositivos *WiFi*;
5) Aplicar método de geração de características em ambientes de simulação computacional.

IV. ADISEH - Autenticação de dispositivos *IoT* com base nas informações contidas no sinal eletromagnético

DADA a importância da extração das imperfeições dos sinais eletromagnéticos em forma de características, a fornecer uma impressão digital fidedigna do dispositivo, é importante descrever alguns elementos essenciais para trabalhos neste propósito, seguindo para a definição do modelo empregado neste trabalho.

A. O Background

Muitos dos métodos descritos na literatura são baseados, na execução de 4 atividades genéricas: i) Coleta do sinal de interesse (utilizando filtros para selecionar o canal); ii) Extração da(s) característica(s) de interesse e a montagem da "impressão digital" do dispositivo (*Radiometric Signature* [19] ou *RF Fingerprinting* [13]); iii)

Checagem das características extraídas (*Match*); e iv) Aprendizado/retroalimentação. Observa-se, portanto, que a principal diferença entre os métodos propostos encontra-se nas atividades ii e iii.

Uma distinção importante na técnica empregada para a extração de característica(s) é definir qual parte de um sinal eletromagnético vai ser utilizado: sinais transientes ou sinais dos dados. Os chamados sinais transientes (*turn-on transient signals*), sendo o sinal emitido no início de uma comunicação, em que o rádio transmissor estava desligado e começa a transmitir, aumentando gradativamente sua potência de transmissão, ocorrem rapidamente e são difíceis de identificação e coleta, mas podem carregar muitas características únicas do rádio transmissor [15].

Por sua vez, a identificação a partir dos sinais dos dados considera toda e qualquer parte restante de um sinal eletromagnético recebido pelo rádio receptor, podendo ser uma extensa amostra do sinal, ou o preâmbulo, por exemplo.

Observa-se, portanto, que há de se decidir por coletar características mais fidedignas em porções do sinal eletromagnético com menor número de amostras (baixa resolução), como é o caso nos sinais transientes, ou apostar numa maior quantidade de amostras do sinal, para a segunda opção. A presente proposta busca utilizar o menor número possível de amostras para a extração das características, permitindo o ajuste conforme a complexidade do cenário (ruído, por exemplo).

Como a criação do *RF Fingerprinting*, também conhecido como *Physical-layer Identification* [13], baseia-se na coleta das imperfeições dos emissores *wireless* observáveis nas amostras do sinal eletromagnético, faz-se importante definir quais são estas imperfeições observáveis.

Conforme descrito em [12], [13] e [26], as imperfeições no sinal eletromagnético que geram a "impressão digital" do dispositivo são observadas em uma ou mais características: i) Alterações de compensação da frequência da portadora; ii) Correlações nos quadros *SYNC*; iii) distúrbios na compensação IQ; iv) Erro de magnitude dos símbolos; e v) Alterações na compensação de fase na portadora. Tais imperfeições são geradas de forma não intencional por diversos motivos, incluindo micro-imperfeições no circuito do rádio emissor, na antena do dispositivo, entre outros.

Diferentes técnicas empregam diferentes formas para extração das características, incluindo uso de filtros especiais implementados em Rádios Cognitivos (como em [26]), ou com uso de módulos implementados em Matlab ® [27] e [14].

Extraídas as características, elas são armazenadas em um banco de dados para que possa ser utilizada para treinamento do algoritmo de aprendizado de máquina.

Assim que um dispositivo tenta se autenticar na rede, o *gateway* coleta amostras do sinal de interesse, extrai as características de interesse, e compara com o banco de dados. Ocorrendo a identificação (*Match*), o dispositivo estará autorizado.

B. O método proposto

O presente trabalho se apoia na arquitetura de *gateway* proposta em [26], de forma a estender o método a outras tecnologias de comunicação sem fio. O método ADISEH, apresentado na Figura 1, recebe o sinal de interesse, isto é, o sinal do dispositivo que almeja acessar a rede, extrai as características de identificação, que é utilizado para a identificação (*match*) com o uso de um modelo de identificação criado por um algoritmo SVM (do inglês *Suport Vector Machine*). Este modelo é criado com base em um banco de dados de características de dispositivos já conhecidos.

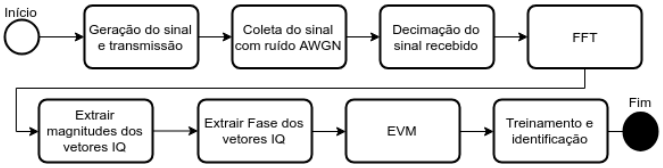

Fig. 1. Fluxo de processos desde a emissão do sinal pelo dispositivo transmissor, sua recepção e autenticação pelo método ADISEH.

Diferentemente do trabalho de [26], que usou a fase de Transmissão (Figura 1) em dispositivos reais, em uma única tecnologia de comunicação (2 módulos *Wifi* nrf24L01 em 2,4 GHz), devido a inviabilidade de implementação em diversos outros tipos de dispositivos reais de outras tecnologias, este trabalho leva em consideração a fase de Transmissão, Coleta e Processamento do sinal em ambiente de simulação computacional (Matlab ®). Tal formulação permite escalar a avaliação do ADISEH em termos de tipos de tecnologias, bem como em número de dispositivos, além de permitir se observar mais facilmente o comportamento do algoritmo frente às variações de SNR.

O conjunto de operações empregado neste trabalho é apresentado na Figura 2, e leva em consideração três entidades: o dispositivo *IoT* que quer obter acesso ao meio, o sistema de identificação de dispositivos, implementado no *gateway*, e a aplicação que requisita a identificação do dispositivo sob análise, que pode ser um protocolo de segurança a ser implementado na rede.

O sinal recebido passa pelo processo de decimação, que reduz a taxa de amostragem por um fator inteiro (*down-sampling*), sendo que o fator de decimação empregado depende do tamanho da amostra utilizada. O sinal decimado é então processado pelo bloco de transformada de Fourier (*FFT*, do inglês *fast fourier transform*) no domínio da frequência, retornando um vetor em que cada amostra é a transformada de Fourier do vetor IQ (do inglês *In-phase and Quadracture*) do sinal sob análise, isto é, considerando os componentes de Fase e Quadratura. As características são, então, extraídas através das magnitudes e fases, para cada vetor IQ, em que a magnitude é o valor absoluto do vetor complexo, conforme:

$$|a + bi| = \sqrt{a^2 + b^2}, \tag{1}$$

que representa o comprimento do vetor desde a origem até a posição do símbolo, e a fase (*Phase Angle*) se dá pelo

cálculo do ângulo do vetor IQ, que é observado na parte imaginária do vetor, conforme:

$$log(z) = log(r) + i\theta, \tag{2}$$

de forma que, ao fim do processo, obtêm-se um vetor de amostras das características do sinal, formando uma espécie de impressão digital do dispositivo.

Como o trabalho também leva em consideração diversos dispositivos comunicando em um mesmo ambiente num mesmo instante T, a fase de coleta do sinal irá receber uma mistura de todos os sinais, juntamente com o ruído, sendo que neste caso, também será aplicado algoritmo de separação dos sinais, antes que estes sejam tratados pelas demais fases do ADISEH, permitindo a extração de características adequadamente.

Para tal separação, adotou-se a técnica dos Mínimos Quadrados, conforme aplicada por Lopes [28], segundo a equação para o cálculo do canal H:

$$H = (X_{pil}S_{pil}^{H})(S_{tx}S_{pil}^{H})^{-1}, \tag{3}$$

sendo X_{pil} a matriz dos sinais recebidos, que carregam as informações referentes aos n_p pilotos, S_{pil} a matriz dos sinais pilotos, $(.)^{-1}$ a operação de matriz inversa e $(.)^{H}$ o operador Hermitiano. De posse deste canal calculado, estima-se o sinal transmitido por cada fonte com o produto:

$$S = (H)^{-1}Y, \tag{4}$$

sendo Y a matriz de sinais misturados recebidos pelo *Gateway*. Então, o sinal S de interesse é comparado com uso do modelo treinado.

Ao fim do processo, em se tratando de um dispositivo autorizado, o *gateway* deve permitir seu acesso à rede, e novos dados das características observáveis podem ser salvas no banco de dados, retroalimentando o sistema, permitindo-o ir aprendendo conforme o uso da rede, e impedir o acesso dos dispositivos não reconhecidos.

Desta forma, o *gateway* é avaliado em comunicações com ruído AWGN (Ruído Gaussiano Branco Aditivo), simulando comunicações em ambientes reais considerando as modulações OQPSK (do inglês *Offset Quadrature Phase Shift Keying*) em 2.450 e 780 MHz, BPSK (do inglês *Binary Phase Shift Keying*) em 868 e 950 MHz, OQPSK e ASK (do inglês *Amplitude Shift Keying*) em 868 e 950 MHz, e GFSK (do inglês *Gaussian Frequency Shift Keying*) e BPSK em 950 MHz, sendo todos implementados pela família do padrão IEEE 802.15.4 (a, b, c e d), cobrindo toda a variedade do *ZigBee*, além de considerar *Bluetooth* e *WiFi*, ambos em 2.450 MHz.

Para a validação, simulações foram realizadas em ambiente computacional (Matlab ®), em que os sinais das tecnologias de interesse são facilmente gerados, de forma fidedigna. Neste tipo de simulação, pode-se adotar o uso de blocos nativos de geração de sinais das tecnologias de interesse, ou gerar os sinais através de aplicação de fórmulas matemáticas.

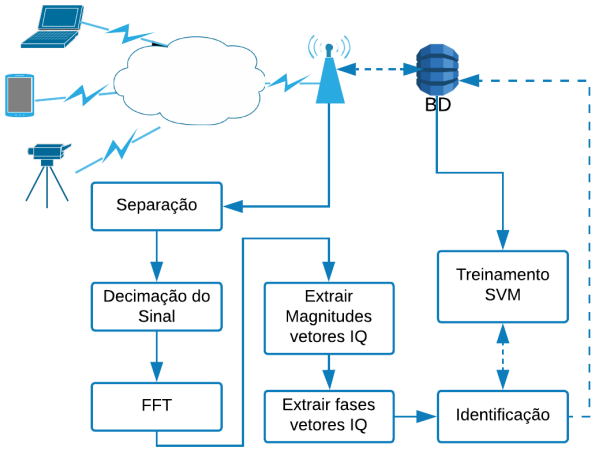

Fig. 2. Arquitetura do *Gateway* implementando o ADISEH em redes sem fio heterogêneas.

Este trabalho levou em consideração a validação utilizando os blocos nativos (*Communications Toolbox Library*), e apresenta um processo para geração de características, no qual uma chave é gerada para cada dispositivo, em que esta chave é usada para gerar uma seleção aleatória de alguns símbolos IQ, que serão modificados para simular as imperfeições. Tal processo criado age sobre o sinal a ser transmitido, permitindo a simulação das características de interesse, deslocando os símbolos IQ ou alterando a fase do sinal em um percentual estabelecido (entre 0,0002% a 0,046%), de forma a permitir explorar sinais com características mais ou menos fortes (com maior ou menor diferenciação), mas sem impactar a qualidade do sinal.

As simulações levam em consideração três cenários: i) um único dispositivo comunicando-se com o *gateway*, sendo que este dispositivo envia mensagens aleatórias de 120 *bytes* em cada comunicação, com foco em avaliar o processo de autenticação em diversos cenários; ii) um único dispositivo invasor, tentando se passar por um dispositivo com direito de acesso, replicando o mesmo sinal do dispositivo autorizado, em diversos cenários; e iii) cenário em que diversos dispositivos estão se comunicando em um mesmo canal, em que seus sinais chegam misturados ao *gateway*, dificultando o processo de autenticação.

A comunicação, simulando a recepção do sinal pelo *gateway* é realizada com a adição de ruído AWGN. A coleta do sinal, a decimação do sinal recebido (reduzir a quantidade de amostras e amplitude) e extração das características também é feita no Matlab, seguido de seu armazenamento localmente para futuro uso, simulando o banco de dados requerido para o *gateway*. Para os casos em que o foco é observar o comportamento do ADISEH com múltiplos sinais, o processo é o mesmo, somente sendo incluídas as técnicas de separação dos sinais de interesse, antes da extração das características e armazenamento (Figura 2).

Logo que o sistema entra em operação, o algoritmo de SVM é executado com dados das características para aprendizado e ajustes (*fit*), gerando um modelo a ser utilizado na identificação dos dispositivos. Assim, o sinal coletado do dispositivo de interesse é avaliado com uso deste modelo, que pode identificar se o sinal sob análise se enquadra no modelo obtido, isto é, possui características de algum dispositivo conhecido, tendo, assim, direito de acesso à rede. Importante ressaltar que o modelo somente é treinado uma única vez, devendo ser utilizado sempre que for necessário checar se um dispositivo é autorizado.

O algoritmo de SVM adotado é o oferecido pelo Matlab (SVM *default*), que recebe como parâmetros apenas a matriz X contendo amostras de treinamento (matriz de predição), e um vetor de etiquetas de identificação y dos dispositivos, de forma que cada linha de X contenha um conjunto de 1.153 características de um único dispositivo conhecido, sendo este valor arbitrário.

Este trabalho emprega método de Monte Carlo para avaliar a Probabilidade de Detecção (P_d), sendo uma métrica de qualidade da presente proposta, em que P_d indica a probabilidade do ADISEH identificar corretamente o dispositivo. Desta forma, se a P_d indicar 90% de descoberta, significa que o *gateway* tem 90% de chances de bloquear o dispositivo não autorizado, e que em 10% dos casos o boqueio não ocorrerá. As simulações empregam 50.000 experimentos, de forma a observar uma tendência de resultados mais realista [29].

V. Resultados e Discussão

Importante ressaltar que as simulações realizadas consideram cenários em que o dispositivo que busca a autenticação é um dispositivo autorizado, que já foi anteriormente identificado pelo *gateway* e possui amostras de suas características já cadastradas no banco de dados, ou é um dispositivo malicioso (invasor), que está clonando o sinal de um dispositivo autorizado, tentando se passar por ele. Tal abordagem se justifica por ser exatamente este o cenário de interesse deste trabalho. São considerados dispositivos de *ZigBee*, *Bluetooth* e *WiFi*.

Inicialmente, busca-se atestar a viabilidade da proposta, em diferentes cenários de ruídos, com diferentes quantidades de amostras utilizadas no treinamento. A Figura 3 apresenta os resultados de um dispositivo invasor clonando o sinal de um dispositivo autorizado, o que dificulta o processo de identificação. Observa-se que a proposta mostra-se viável, permitindo uma identificação da tentativa de invasão próxima a 80% em cenários de baixo ruído.

Adicionalmente, a Figura 3 permite observar que um incremento significativo de amostras no treinamento do modelo de predição (SVM) não gera melhoras proporcionalmente significativas na capacidade de descoberta, ao passo que a melhora do sinal (maior SNR) é definitiva para o sucesso do ADISEH, como era esperado.

Tal comportamento pode ser explicado pela própria natureza do algoritmo de SVM empregado nas simulações, que busca um hiperplano ideal para a classificação (separação das classes), sendo que o seu desempenho depende da capacidade de separação das classes. Para classes perfeitamente separáveis, o hiperplano maximiza a margem de separação, criando limites para as classes. Entretanto, para classes inseparáveis, o algoritmo impõe uma penalidade no comprimento da margem de separação, ampliando o erro. Observa-se, portanto, que o vetor de características empregado no treinamento comporta-se como o segundo caso, sendo que o seu incremento na predição amplia o erro da classificação.

Ressalta-se que o custo computacional para o treinamento do modelo é diretamente proporcional ao tamanho da matriz de preditores, como observado nas simulações realizadas neste trabalho.

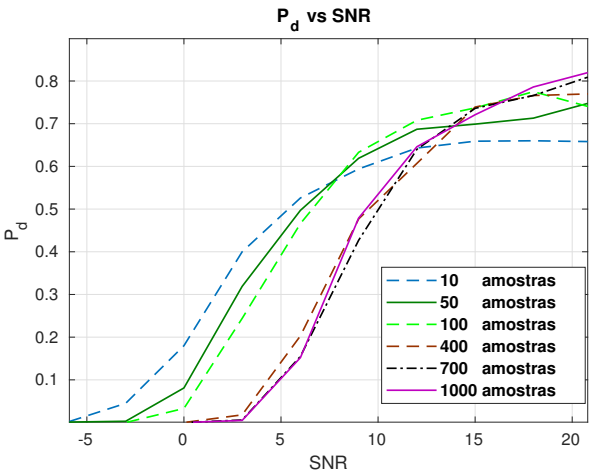

Fig. 3. Avaliação da capacidade de descoberta de um dispositivo invasor clonando o sinal de um dispositivo autorizado frente ao número de amostras utilizadas no treinamento do modelo, para cada SNR.

Diante disso, as demais simulações passam a considerar conjuntos de preditores menores, entre 5 a 50 amostras, incrementando-as em 5, para cada cenário de ruído (SNR), com foco em observar a capacidade de descoberta de dispositivos em diferentes cenários, com diferentes influências do ruído, inicialmente considerando situações em que um dispositivo já autorizado e conhecido pelo *gateway* tenta realizar a autenticação na rede. Também, como pode observado na Figura 3, os resultados não são satisfatórios para SNR muito baixa, não sendo, então, considerado por este trabalho situações abaixo de $-6dB$.

A. Avaliando a identificação de dispositivos autorizados

A Figura 4 apresenta os resultados simulados para *ZigBee*, em que observa-se que o incremento de amostras no treinamento do modelo melhora a P_d em diversos dos cenários de SNR, atingindo valores muito próximos a 100% de descoberta correta para SNR acima de $15dB$. Nestes cenários, observa-se que um pequeno incremento de amostras no treinamento pode elevar a capacidade de descoberta, como pode ser notado no cenário de $18dB$, que possui P_d abaixo de 90% com 5 amostras, mas que atinge 100% quando passa a contar com 25 amostras.

Estando o dispositivo em um cenário de SNR entre $6dB$ e $12dB$, um incremento mais significativo de amostras no treinamento deve ser considerado, sendo que com baixo número de amostras faz com que o modelo seja muito impreciso. Já para os cenários de forte ruído, abaixo de $6dB$, os resultados apontam para uma instabilidade, visto que o ruído degrada significativamente as características observáveis do dispositivo.

Ao se considerar os resultados observados para *Bluetooth* (Figura 5), em que a descoberta se aproxima ou supera os 80% para 50 amostras em até $6dB$, espera-se que o desempenho do *gateway* em ambientes reais seja satisfatório, dada a baixa distância máxima de operação destes dispositivos. Novamente nota-se que os resultados são instáveis para SNR mais baixa.

O comportamento para SNR acima de $9dB$ em *WiFi*, na Figura 6, também apresenta comportamento semelhante, tendo o incremento de amostras de treinamento para $12dB$ garantido uma descoberta próxima de 90%.

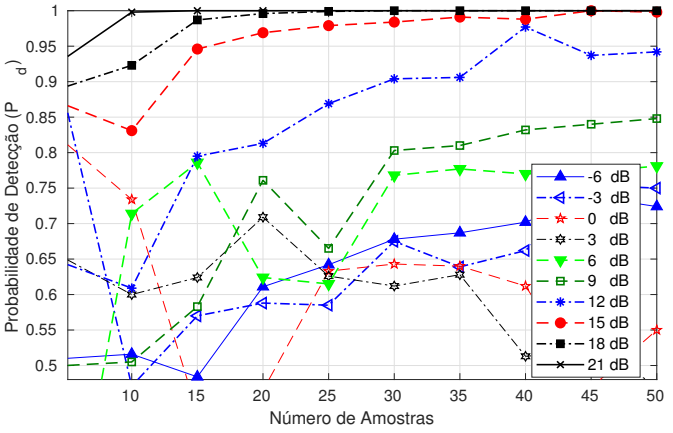

Fig. 4. ZigBee - Impacto do número de amostras no treinamento do SVM na identificação correta de um dispositivo autorizado.

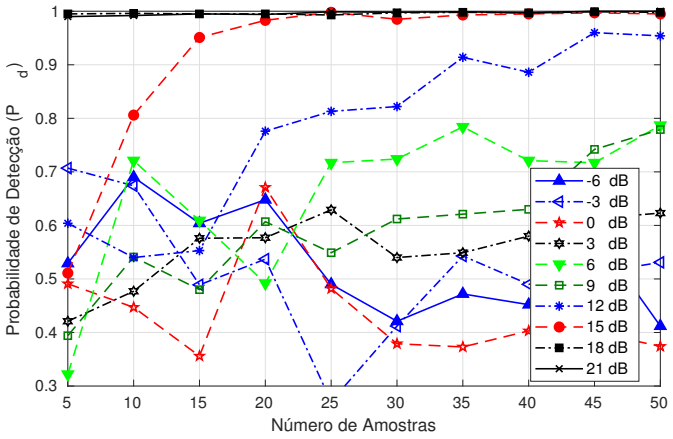

Fig. 5. Bluetooth - Impacto do número de amostras no treinamento do SVM na identificação correta de um dispositivo autorizado.

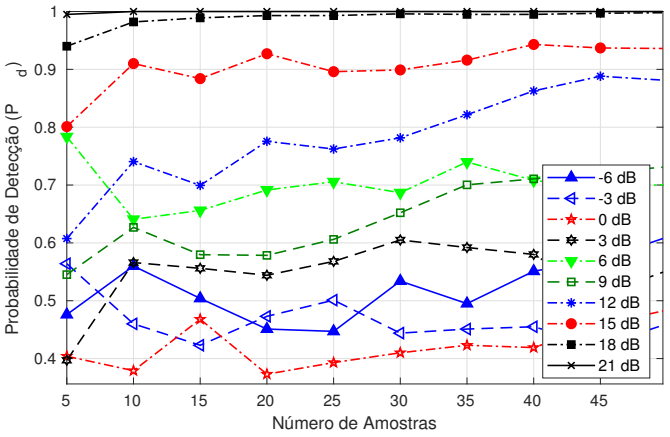

Fig. 6. WiFI - Impacto do número de amostras no treinamento do SVM na identificação correta de um dispositivo autorizado.

B. Avaliando a identificação de dispositivos maliciosos

Abordando um segundo cenário, em que um dispositivo malicioso tenta invadir a rede clonando o sinal de um dispositivo autorizado, como em [25], espera-se que o processo de identificação seja comprometido, visto que o sinal será muito semelhante para os dois dispositivos. Entretanto, a considerar que a "impressão digital" do dispositivo estará presente, diferenciando-o, espera-se que o *gateway* consiga compreender a diferenciação entre os dispositivos.

A Figura 7 apresenta este cenário com uso de dispositivos de *ZigBee*. Nos cenários com SNR abaixo de $9dB$, o incremento de amostras não gera melhorias satisfatórias na descoberta. Quando a SNR está em torno de $9dB$, comprova-se que o incremento de amostras no treinamento do modelo pode melhorar significativamente a capacidade de descoberta. Desta forma, o algoritmo de identificação a ser implementado deve ser parametrizado frente à qualidade do sinal recebido naquele instante, tentando sempre considerar um baixo número de amostras, quando possível. Para os cenários com SNR acima de $9dB$, qualquer quantidade acima de 15 amostras gera resultados acima de 95% de P_d.

Com cenário idêntico, mas considerando agora dispositivos de *Bluetooth* e *WiFi*, a Figura 8 e a Figura 9, respectivamente, apresentam os resultados. Para a identificação de dispositivos de *Bluetooth*, observa-se que o modelo deve ser treinado com um número de amostras mais alto para ambientes com a SNR mais alta, sendo que se a SNR for mais baixa, o incremento de amostras não surte efeito. Já para a identificação de dispositivos de *WiFi*, apesar de ter-se a mesma situação para SNR alta, observou-se que para SNR entre 9 e 12 dB, o uso entre 15 a 30 amostras pode favorecer a descoberta.

Considerando os cenários aqui empregados, comprova-se que o método proposto oferece um meio eficiente para a identificação de dispositivos, tanto para os dispositivos já conhecidos e autorizados, quanto para os cenários em que a rede está sujeita ao uso indevido por dispositivo malicioso. O sucesso da descoberta nos cenários em que o

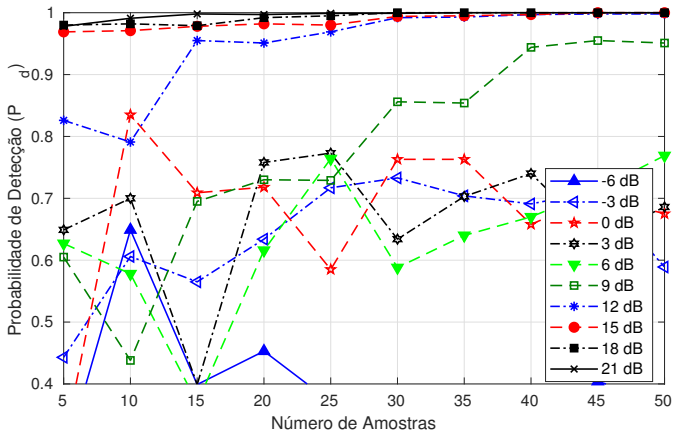

Fig. 7. ZigBee - Impacto do número de amostras no treinamento do SVM com dispositivo invasor clonando o sinal de um dispositivo autorizado.

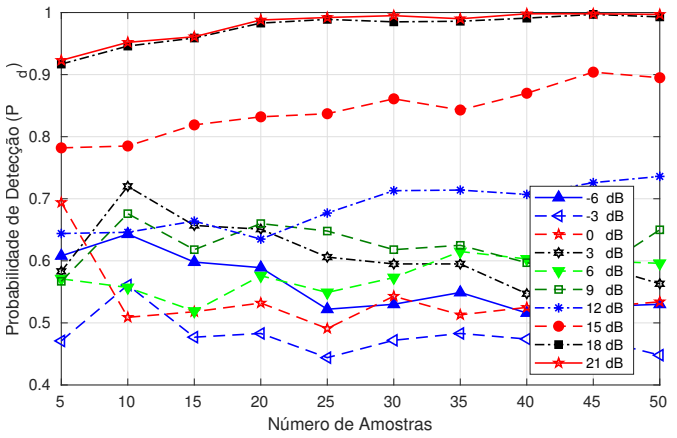

Fig. 8. Bluetooth - Impacto do número de amostras no treinamento do SVM com dispositivo invasor clonando o sinal de um dispositivo autorizado.

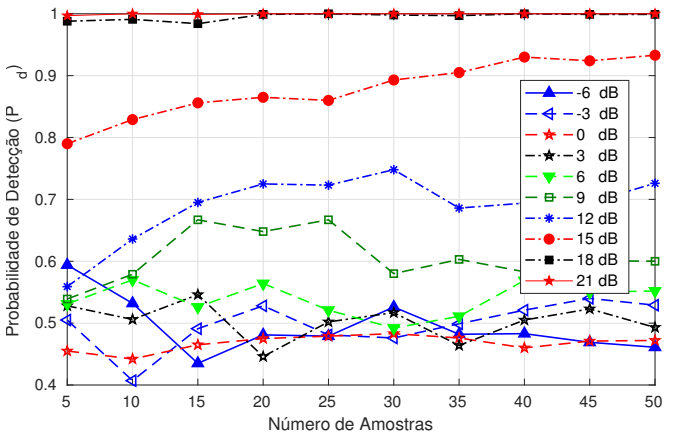

Fig. 9. WiFi - Impacto do número de amostras no treinamento do SVM com dispositivo invasor clonando o sinal de um dispositivo autorizado.

ruído está presente, mas não é muito forte, perante o ajuste no número de amostras no treinamento do modelo permite observar que o algoritmo que implementar o ADISEH deve levar em consideração a SNR, de forma constante.

VI. Conclusão

RESTANDO claro a necessidade de construção de métodos que permitam a coexistência entre dispositivos de tecnologias distintas, em um ambiente de *IoT*, levando-se em consideração que diversos destes dispositivos podem não ter a capacidade computacional necessária para a implementação de métodos de autenticação mais robustos, o presente trabalho apresenta o método ADISEH, que permite a realização de autenticação destes dispositivos com base nas características do sinal eletromagnético contidas em suas comunicações em redes sem fio heterogêneas, de forma a habilitar um protocolo mais robusto, a ser implementado em um *gateway* para *IoT*.

O método foi implementado em ambiente de simulação, considerando múltiplas tecnologias (*ZigBee*, *Bluetooth* e *WiFi*), em diversos cenários, como requerido para o referido *gateway*. Simulações realizadas comprovaram a sua eficácia, principalmente quando a SNR for alta, em termos da probabilidade de descoberta P_d correta dos dispositivos.

Comprovou-se que em ambientes em que o ruído for mais forte, um incremento no número de amostras no treinamento do modelo de predição (SVM) pode compensar a confusão gerada no sinal, garantindo um nível de acertividade mínimo, mas que tal número não precisa ser demasiadamente grande. Também mostrou-se que em um ambiente com a SNR muito baixa, o método não apresenta resultados satisfatórios, apontando para a necessidade de outras abordagens, como o ajuste na potência dos transmissores, por exemplo, caso possível seja.

Como trabalhos futuros, aponta-se a necessidade de implementação do modelo considerando outras tecnologias *wireless*, incluindo outras abordagens de extração das características, com foco em melhorar a capacidade de descoberta nos cenários de ruído alto. Também deve ser considerada a implementação do modelo em múltiplas tecnologias em um SDR físico com múltiplos dispositivos, observando-se o seu comportamento em diversos cenários de ruído real.

Conforme observou-se nos experimentos uma correlação entre o número de amostras utilizadas na predição, a tecnologia e o ruído, os trabalhos futuros também devem incluir tais considerações na proposição de um intervalo para configuração do *gateway*, em um modelo adaptativo, baseando-se nestes parâmetros como métricas para a sua operação.

Referências

[1] S. Singh e N. Singh, “Internet of Things (IoT): Security challenges, business opportunities & reference architecture for E-commerce”, em *Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on*, IEEE, 2015, pp. 1577–1581.

[2] J. Gubbi, R. Buyya, S. Marusic e M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions", *Future generation computer systems*, vol. 29, n.º 7, pp. 1645–1660, 2013.

[3] T. Xu, J. B. Wendt e M. Potkonjak, "Security of IoT systems: Design challenges and opportunities", em *Proceedings of the 2014 IEEE/ACM International Conference on Computer-Aided Design*, IEEE Press, 2014, pp. 417–423.

[4] E. Bertino, K.-K. R. Choo, D. Georgakopolous e S. Nepal, "Internet of Things (IoT): Smart and secure service delivery", *ACM Transactions on Internet Technology (TOIT)*, vol. 16, n.º 4, p. 22, 2016.

[5] D. E. Kouicem, A. Bouabdallah e H. Lakhlef, "Internet of things security: A top-down survey", *Computer Networks*, vol. 141, pp. 199–221, 2018.

[6] M. Nawir, A. Amir, N. Yaakob e O. B. Lynn, "Internet of Things (IoT): Taxonomy of security attacks", em *Electronic Design (ICED), 2016 3rd International Conference on*, IEEE, 2016, pp. 321–326.

[7] J. Mitola, "SDR architecture refinement for JTRS", em *MILCOM 2000. 21st Century Military Communications Conference Proceedings*, IEEE, vol. 1, 2000, pp. 214–218.

[8] Q. Zhu, R. Wang, Q. Chen, Y. Liu e W. Qin, "Iot gateway: Bridgingwireless sensor networks into internet of things", em *Embedded and Ubiquitous Computing (EUC), 2010 IEEE/IFIP 8th International Conference on*, Ieee, 2010, pp. 347–352.

[9] S. Guoqiang, C. Yanming, Z. Chao e Z. Yanxu, "Design and implementation of a smart IoT gateway", em *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, IEEE, 2013, pp. 720–723.

[10] S. K. Datta, C. Bonnet e N. Nikaein, "An IoT gateway centric architecture to provide novel M2M services", em *Internet of Things (WF-IoT), 2014 IEEE World Forum on*, IEEE, 2014, pp. 514–519.

[11] LABORA, *LABORA - SOFTware-defined gateWAY and fog computing for IoT*, 2018. URL: https://software4iot.labora.inf.ufg.br/.

[12] B. Danev, H. Luecken, S. Capkun e K. El Defrawy, "Attacks on physical-layer identification", em *Proceedings of the third ACM conference on Wireless network security*, ACM, 2010, pp. 89–98.

[13] B. Danev, D. Zanetti e S. Capkun, "On physical-layer identification of wireless devices", *ACM Computing Surveys (CSUR)*, vol. 45, n.º 1, p. 6, 2012.

[14] G. Verma, P. Yu e B. M. Sadler, "Physical layer authentication via fingerprint embedding using software-defined radios", *IEEE Access*, vol. 3, pp. 81–88, 2015.

[15] H. C. Choe, C. E. Poole, M. Y. Andrea e H. H. Szu, "Novel identification of intercepted signals from unknown radio transmitters", em *Wavelet Applications II*, International Society for Optics e Photonics, vol. 2491, 1995, pp. 504–518.

[16] J. Toonstra e W. Kinsner, "Transient analysis and genetic algorithms for classification", em *WESCANEX 95. Communications, Power, and Computing. Conference Proceedings., IEEE*, IEEE, vol. 2, 1995, pp. 432–437.

[17] K. Zhao e L. Ge, "A survey on the internet of things security", em *Computational Intelligence and Security (CIS), 2013 9th International Conference on*, IEEE, 2013, pp. 663–667.

[18] R. Mahmoud, T. Yousuf, F. Aloul e I. Zualkernan, "Internet of things (IoT) security: Current status, challenges and prospective measures", em *Internet Technology and Secured Transactions (ICITST), 2015 10th International Conference for*, IEEE, 2015, pp. 336–341.

[19] V. Brik, S. Banerjee, M. Gruteser e S. Oh, "Wireless device identification with radiometric signatures", em *Proceedings of the 14th ACM international conference on Mobile computing and networking*, ACM, 2008, pp. 116–127.

[20] K. B. Rasmussen e S. Capkun, "Implications of radio fingerprinting on the security of sensor networks", em *Security and Privacy in Communications Networks and the Workshops, 2007. SecureComm 2007. Third International Conference on*, IEEE, 2007, pp. 331–340.

[21] J. Hall, M. Barbeau e E. Kranakis, "Radio frequency fingerprinting for intrusion detection in wireless networks", *IEEE Transactions on Defendable and Secure Computing*, vol. 12, pp. 1–35, 2005.

[22] M. Barbeau, J. Hall e E. Kranakis, "Detection of rogue devices in bluetooth networks using radio frequency fingerprinting", em *proceedings of the 3rd IASTED International Conference on Communications and Computer Networks, CCN*, Citeseer, 2006, pp. 4–6.

[23] T. J. Bihl, K. W. Bauer e M. A. Temple, "Feature selection for RF fingerprinting with multiple discriminant analysis and using ZigBee device emissions", *IEEE Transactions on Information Forensics and Security*, vol. 11, n.º 8, pp. 1862–1874, 2016.

[24] T. J. Bihl, M. A. Temple e K. W. Bauer, "Feature selection fusion (FSF) for aggregating relevance ranking information with application to ZigBee radio frequency device identification", em *Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS), 2016 IEEE National*, IEEE, 2016, pp. 80–87.

[25] N. T. Nguyen, G. Zheng, Z. Han e R. Zheng, "Device fingerprinting to enhance wireless security using nonparametric Bayesian method", em *INFOCOM, 2011 Proceedings IEEE*, IEEE, 2011, pp. 1404–1412.

[26] M. F. d. Abreu, P. F. Sousa, V. d. C. Borges, A. C. d. O. Júnior e K. V. Cardoso, "Autenticação de dispositivos de Internet das Coisas baseada

nas características do sinal eletromagnético", em *VI ERIGO, 2018 Anais*, SBC, 2018, pp. 65–75.

[27] M. Pospisil, R. Marsalek e J. Pomenkova, "Wireless device authentication through transmitter imperfections—measurement and classification", em *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, IEEE, 2013, pp. 497–501.

[28] V. H. L. Lopes, "Sensoriamento de espectro contınuo baseado em cancelamento de fontes", 2015.

[29] L. Chwif e A. C. Medina, *Modelagem e simulação de eventos discretos*. Afonso C. Medina, 2006.