



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CIN - CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

ALEXANDRE LUÍS BÔA-VIAGEM DE FRANÇA

**UMA ARQUITETURA MULTINÍVEL PARA FUSÃO DE DADOS DE
MONITORAMENTO DO CONSUMO DE ENERGIA ELÉTRICA A PARTIR DE
REDES DE SENSORES SEM FIO**

Recife
2022

ALEXANDRE LUÍS BÔA-VIAGEM DE FRANÇA

**UMA ARQUITETURA MULTINÍVEL PARA FUSÃO DE DADOS DE
MONITORAMENTO DO CONSUMO DE ENERGIA ELÉTRICA A PARTIR DE
REDES DE SENSORES SEM FIO**

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática (CIn) da Universidade Federal de Pernambuco – UFPE, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Área de Concentração: Inteligência Computacional

Orientadora: Dr^a Flávia de Almeida Barros

Recife

2022

Catálogo na fonte
Bibliotecária: Mônica Uchôa, CRB4-1010

F814a França, Alexandre Luís Bôa-Viagem de.
 Uma arquitetura multinível para fusão de dados de monitoramento
 do consumo de energia elétrica a partir de redes de sensores sem fio /
 Alexandre Luís Bôa-Viagem de França . – 2022.
 102 f.: il., fig.

 Orientadora: Flávia de Almeida Barros.
 Dissertação (Mestrado) – Universidade Federal de Pernambuco.
 Cln. Programa de Pós-graduação em Ciência da Computação. Recife,
 2022.
 Inclui referências.

 1. Fusão de dados. 2. Internet das coisas. 3. Sensores. I. Barros,
 Flávia de Almeida (Orientadora). II. Título.

 681.3 CDD (23. ed.) UFPE- CCEN 2022 - 76

ALEXANDRE LUÍS BÔA-VIAGEM DE FRANÇA

**UMA ARQUITETURA MULTINÍVEL PARA FUSÃO DE DADOS DE
MONITORAMENTO DO CONSUMO DE ENERGIA ELÉTRICA A PARTIR DE
REDES DE SENSORES SEM FIO**

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática (CIn) da Universidade Federal de Pernambuco – UFPE, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Área de Concentração: Inteligência Computacional

Aprovada em: 11/02/2022.

BANCA EXAMINADORA

Prof.^a Dr.^a. Flávia de Almeida Barros (orientadora)

Universidade Federal de Pernambuco

Prof.^a Dr. Carlos Andre Guimarães Ferraz (Examinador Interno)

Universidade Federal de Pernambuco

Prof. Dr^o. André Câmara Alves do Nascimento (Examinador Externo)

Universidade Federal Rural de Pernambuco

AGRADECIMENTOS

Primeiramente, agradeço a Deus por todas as oportunidades que tem me proporcionado, sendo esta mais uma conquista fruto de muita dedicação.

À minha família por toda paciência, dedicação e apoio, principalmente nos momentos mais difíceis durante o período do mestrado.

Ao Centro de Informática da Universidade Federal de Pernambuco, pela oportunidade a mim concedida e por toda estrutura disponível.

À minha orientadora, professora Dr^a Flávia de Almeida Barros, por toda dedicação, atenção e paciência. Seus direcionamentos e orientações foram de grande importância para a conclusão deste trabalho.

Aos professores do Centro de Informática da UFPE, por compartilharem seus conhecimentos com grande dedicação.

A todos que não estão listados aqui, mas que de alguma forma fizeram parte desta incrível jornada, deixo aqui meus sinceros agradecimentos!

RESUMO

Em cenários corporativos, o alto consumo de energia elétrica impacta a lucratividade e a qualidade dos produtos e serviços oferecidos. No setor industrial brasileiro, os gastos com energia elétrica podem chegar a 40% dos custos de produção. A utilização racional de energia, ou eficiência energética, torna-se um forte aliado das empresas, sobretudo no setor industrial. Contudo, a eficiência energética requer altos investimentos em automação industrial, que geralmente é invasiva em sua implantação e manutenção. Claramente, obter maior eficiência energética a custo acessível e com portabilidade e escalabilidade apresenta-se como um desafio. Nesse contexto, este trabalho propõe a utilização de redes de sensores sem fio (RSSF), não invasivas e de baixo custo, para monitoramento de equipamentos eletroeletrônicos residenciais ou industriais utilizando a abordagem de IoT (*Internet of Things*). Os dados coletados são fusionados através de uma arquitetura multinível de fusão de dados projetada neste trabalho, gerando assim informações com maior valor agregado (revelando, por exemplo, as principais grandezas elétricas inerentes à demanda e ao consumo de energia). Com o monitoramento contínuo e em tempo real é possível identificar falhas e tomar ações corretivas com mais rapidez, evitando desperdícios e gerando economia. As RSSFs foram implantadas em dois contextos distintos: um ambiente residencial, e um ambiente industrial. O estudo de caso realizado em ambiente doméstico focou na implementação de uma RSSF atrelada a uma geladeira, com vistas a monitorar o consumo de energia por um determinado período. O segundo estudo de caso foi desenvolvido em um ambiente industrial, com objetivo foi monitorar remotamente o consumo de energia de um equipamento denominado “extrusora”, em uma pequena indústria de plástico da região Metropolitana do Recife. Em ambos os casos, os resultados foram muito positivos. Além do grande valor gerencial dos resultados obtidos a partir da fusão dos dados, conseguimos também confirmar a escalabilidade e a independência de domínio da arquitetura de fusão proposta nesse trabalho, que foi utilizada em ambos os estudos de caso com pequenas alterações no *hardware* (mudança de transformadores de corrente, microcontroladores e o rádio).

Palavras-chave: fusão de dados; internet das coisas; sensores.

ABSTRACT

In corporate scenarios, high electricity consumption impacts profitability and the quality of products and services offered. In the Brazilian industrial sector, electricity costs can reach 40% of production costs. The rational use of energy, or energy efficiency, becomes a strong ally for companies, especially in the industrial sector. However, energy efficiency requires high investments in industrial automation, which is usually invasive in its implementation and maintenance. Clearly, achieving greater energy efficiency at an affordable cost and with portability and scalability presents itself as a challenge. In this context, this work proposes the use of wireless sensor networks (WSN), non-invasive and low cost, for monitoring residential or industrial electronic equipment using the IoT (Internet of Things) approach. The collected data are merged through a multilevel data fusion architecture designed in this work, thus generating information with greater added value (revealing, for example, the main electrical quantities inherent to demand and energy consumption). With continuous and real-time monitoring, it is possible to identify failures and take corrective actions more quickly, avoiding waste and generating savings. WSNs were implemented in two different contexts: a residential environment and an industrial environment. The case study carried out in a domestic environment focused on the implementation of a WSN linked to a refrigerator, in order to monitor energy consumption for a certain period. The second case study was developed in an industrial environment, with the objective of remotely monitoring the energy consumption of an equipment called "extruder", in a small plastic industry in the Metropolitan region of Recife. In both cases, the results were very positive. In addition to the great managerial value of the results obtained from the data fusion, we were also able to confirm the scalability and domain independence of the fusion architecture proposed in this work, which was used in both case studies with minor *hardware* changes (change of current transformers, microcontrollers and the radio).

Keywords: data fusion; internet of things; sensors.

LISTA DE FIGURAS

Figura 1- Relação da IoT com Computação Ubíqua, Móvel e Pervasiva	19
Figura 2- Exemplos de topologia de RSSF	25
Figura 3- Exemplo de arquitetura de RSSF	26
Figura 4- Detalhes de uma RSF enfatizando o nó sensor.....	27
Figura 5- Representação gráfica <i>Box Plot</i> de <i>outliers</i> em conjuntos de dados.	31
Figura 6- Taxonomia de técnicas e métodos de detecção de <i>outliers</i>	36
Figura 7- Relacionamento entre termos ligados a fusão.	40
Figura 8- Fusão baseada na relação entre fontes.....	43
Figura 9- Métodos baseados em entrada e saída	45
Figura 10- Modelo JDL	49
Figura 11- Modelo DFD	50
Figura 12- Modelo OODA.....	51
Figura 13- Arquitetura Multinível de Fusão de Dados	60
Figura 14- Esquema Funcional da Arquitetura Multinível de Fusão de Dados.....	61
Figura 15- AMnFD – Camada de baixo nível	62
Figura 16- AMnFD – Camada de médio nível	64
Figura 17- AMnFD – Camada de alto nível	65
Figura 18- Arquitetura da RSSF para gestão de consumo de energia elétrica	73
Figura 19- Sensores utilizados nos casos de uso	77
Figura 20- Quantidade de leituras por dia.	81
Figura 21- Quantidade de leituras por hora.....	82
Figura 22- Quantidade de leituras por dia da semana.	82
Figura 23- Leituras por classificação.....	83
Figura 24- Gráfico de Leituras por Intervalo de tempo.	84
Figura 25- Visão geral de uma extrusora	86
Figura 26- Quantidade de leituras por dia	87
Figura 27- Gráfico de fator de cargas.....	89
Figura 28- Gráfico do Fator de potência.....	89
Figura 29- Gráfico de Demanda Máxima	90
Figura 30- Gráfico de Consumo em KW no dia 10Dez2020.	91
Figura 31- Gráfico de Consumo total em KW.....	91

SUMÁRIO

1	INTRODUÇÃO	11
1.1	TRABALHO REALIZADO.....	12
1.2	ESTRUTURA DO DOCUMENTO.....	15
2	ÁREAS CORRELATAS - IOT E RSSF	17
2.1	INTERNET DAS COISAS (IoT)	17
2.1.1	Visão Geral	17
2.1.2	Tipos de rede IoT	20
2.1.3	Aplicações da IoT	21
2.2	REDES DE SENSORES SEM FIO – RSSF	22
2.2.1	Visão Geral	22
2.2.2	Arquitetura das RSSF	24
2.2.3	Estrutura do nó	26
2.2.4	Classificação das RSSF	27
2.3	CONSIDERAÇÕES FINAIS	29
3	OUTLIERS E FUSÃO DE DADOS EM RSSF.....	30
3.1	OUTLIERS	30
3.1.1	Tipos de Anomalias / outliers	31
3.1.2	Deteção de outliers.....	33
3.1.3	Abordagens e Técnicas de Deteção de outliers	34
3.2	FUSÃO DE DADOS	38
3.2.1	Visão Geral	38
3.2.2	Fusão de informações, fusão de sensores e fusão de dados	39
3.2.3	Limitações	40
3.2.4	Classificação de fusão de dados / informações	41
3.2.4.1	Classificação baseada na relação entre fontes de dados.....	41
3.2.4.2	Classificação baseada em níveis de abstração de dados	43

3.2.4.3	Classificação baseada nos níveis da entrada e saída de dados	44
3.2.4.4	Classificação baseada na arquitetura (centralizada x descentralizada)	46
3.2.5	Modelos de Fusão	47
3.2.5.1	Modelo JDL.....	48
3.2.5.2	Modelo DFD.....	50
3.2.5.3	Modelo Durrant-Whyte.....	51
3.2.5.4	Modelo OODA (<i>Observe, Orient, Decide e Act</i>).....	51
3.3	TRABALHOS RELACIONADOS	52
3.4	CRITÉRIOS DE AVALIAÇÃO DOS TRABALHOS NA ÁREA.....	55
3.5	CONSIDERAÇÕES FINAIS	57
4	AMNFD: UMA ARQUITETURA MULTINÍVEL DE FUSÃO DE DADOS	58
4.1	REQUISITOS DESEJÁVEIS x QUESTÕES DE PESQUISA	59
4.2	ARQUITETURA MULTINÍVEL PROPOSTA.....	59
4.2.1	Camada de Baixo Nível	62
4.2.2	Camada de Nível Médio.....	64
4.2.3	Camada de Alto Nível	65
4.3	CONSIDERAÇÕES FINAIS	66
5	IMPLEMENTAÇÃO E ESTUDOS DE CASO	66
5.1	CONCEITOS BÁSICOS NA ÁREA DE GERENCIAMENTO DE ENERGIA ELÉTRICA	67
5.1.1	Consumo de Energia Elétrica	67
5.1.2	Demanda de Energia Elétrica	68
5.1.3	Outros Indicadores	69
5.2	UMA RSSF PARA MONITORAMENTO DE ENERGIA ELÉTRICA	72
5.2.1	Infraestrutura de comunicação	74
5.2.1.1	Área de conectividade dos dispositivos	74
5.2.1.2	Área Gestão de Mensagem	74

5.2.1.3	Área Gerenciamento e análise de processamento de dados	76
5.2.1.4	Área Apresentação e conectividade de negócios	76
5.2.2	Infraestrutura de <i>Hardware</i>	77
5.2.3	<i>Software</i> desenvolvido	78
5.3	ESTUDOS DE CASO	80
5.3.1	Contexto Residencial (Geladeira)	80
5.3.1.1	Coleta de dados	80
5.3.1.2	Interpretação dos Dados Obtidos	81
5.3.2	Contexto Industrial (Extrusora)	85
5.3.2.1	Coleta dos dados	86
5.3.2.2	Interpretação dos Dados Obtidos	87
5.3.3	Comparação com os Trabalhos relacionados	92
5.4	CONSIDERAÇÕES FINAIS	94
6	CONCLUSÃO	96
6.1	PRINCIPAIS CONTRIBUIÇÕES	96
6.2	TRABALHOS FUTUROS	97
	REFERÊNCIAS	98

1 INTRODUÇÃO

Segundo a Empresa de Pesquisa Energética¹ (BEN, 2021), o Brasil dispõe de uma matriz elétrica essencialmente de origem renovável, representando 84,8% da oferta interna de eletricidade. Dentre as fontes renováveis, a fonte hídrica representa 62,5%. Esses dados deixam clara a grande dependência de disponibilidade de água na geração de energia elétrica no Brasil.

A princípio, esse fato não parece trazer problemas, uma vez que a água é um recurso natural renovável. No entanto, a distribuição desse recurso não é equilibrada no nosso país, concentrando-se na região Norte (80%), que abriga apenas 5% da população brasileira. Assim, a energia gerada na região norte sem grandes custos a partir da abundância de água tem perdas e alto custo de distribuição para regiões de maior demanda. Por fim, vale ressaltar que, no contexto nacional, ações que tenham como objetivo alcançar a economicidade de consumo de energia favorecem a economia de água e vice-versa.

Em diversos cenários corporativos, constata-se que as empresas enfrentam sérios problemas relacionados ao consumo de energia, o que impacta diretamente a lucratividade das empresas e a qualidade dos produtos e serviços entregues ao mercado. De acordo com a Firjan (Federação das Indústrias do Estado do Rio de Janeiro)², no setor industrial brasileiro, os gastos associados à energia elétrica podem representar cerca de 40% dos seus custos de produção. Evidentemente, sendo um dos principais insumos dessa cadeia é de fundamental importância sua utilização de maneira racional e econômica. Com isso, as indústrias estão cada vez mais atentas ao gerenciamento do consumo de energia.

Existem também outros fatores não menos importantes: o aumento dos preços da energia no Brasil (19,3% acumulado entre janeiro de 2019 até junho de 2021); o uso sustentável de recursos naturais (uma vez que a escassez desses recursos, e consequente impacto na geração de energia elétrica podem

¹ Empresa de Pesquisa Energética (EPE) - <https://www.epe.gov.br/pt>

² Firjan <https://www.firjan.com.br/pagina-inicial.htm>

levar a riscos de suspensão de operação das indústrias); a necessidade de adaptação rápida da produção a determinadas condições operacionais visando maior lucratividade dos negócios, entre outros.

A utilização racional de energia, ou simplesmente eficiência energética, torna-se um forte aliado das empresas, sobretudo no setor industrial. Contudo, a eficiência energética requer altos investimentos em automação industrial, que geralmente é invasiva e necessita de alguma obra civil e elétrica, representando altos investimentos em sua implantação e manutenção.

Claramente, obter maior eficiência energética a custo acessível e com portabilidade e escalabilidade apresenta-se como um desafio. Uma abordagem que tem se mostrado eficaz para tratar essa lacuna é o uso de tecnologias como IoT (*Internet of Things*) (ASHTON, 2009). Com IoT é possível a implantação de uma rede de sensores sem fio (RSSF), de baixo custo e não invasiva, para sensoramento e monitoramento de equipamentos residenciais ou industriais, de forma individualizada e/ou setorizada. Os dados coletados podem ser então fusionados e processados com o auxílio de ferramentas de ETL (*Extract, Transform and Load* ³). Por fim, os resultados podem ser apresentados através de *dashboards* providos por ferramentas de *Business Intelligence* (BI), a fim de prover informações gerenciais úteis ao controle do consumo de energia elétrica.

O processo sugerido acima é o estado da arte na área de monitoramento de equipamentos usando IoT. Contudo, durante a revisão da bibliografia relacionada, não encontramos nenhum trabalho na área de eficiência energética que atenda a todos os requisitos desejáveis para tal atividade. Destacamos aqui: baixo custo de implantação e manutenção das RSSF, independência de domínio da aplicação (reuso do projeto), escalabilidade, fusão dos dados coletados, identificação de *outliers* (que podem revelar problemas nos equipamentos), visualização dos resultados através de relatórios e/ou *dashboards*, e gestão de ações por parte do usuário.

1.1 TRABALHO REALIZADO

Nesse contexto, o objetivo principal deste trabalho foi propor um processo automático de tratamento de dados referentes a grandezas energéticas,

³ ETL https://pt.wikipedia.org/wiki/Extract_transform_load

coletados por sensores acoplados a equipamentos, que seja eficaz na identificação de dados anômalos (*outliers*), a fim de prover informação gerencial para apoiar e agilizar a tomada de decisão na busca de melhorar a utilização racional de energia, favorecendo assim a eficiência energética.

As principais Questões de Pesquisa que este trabalho investigou podem ser descritas como a seguir:

- (1) É possível efetuar medição e monitoramento do consumo de energia de equipamentos eletroeletrônicos utilizando sensores de baixo custo, de modo escalável e independente da aplicação (doméstico ou industrial)?
- (2) É possível prover uma maior acurácia na identificação de *outliers* a partir dos dados coletados através de uma arquitetura multinível de fusão de dados, que seja independente de domínio da aplicação?
- (3) A análise final dos dados fusionados (com foco na identificação de *outliers*) auxilia de forma objetiva na tomada de decisão, a fim de favorecer a eficiência energética dos equipamentos?

Visando adequar as questões de pesquisas citadas, esse trabalho vem propor uma arquitetura multinível de fusão de dados utilizando a abordagem de IoT. Como primeiro passo, este trabalho propõe a utilização de redes de sensores sem fio (RSSF), não invasivas e de baixo custo, para monitoramento de equipamentos eletroeletrônicos residenciais ou industriais utilizando a abordagem de IoT (*Internet of Things*) (**QP1**).

Os dados brutos coletados pelos sensores são fusionados e tratados através de uma arquitetura de fusão de dados projetada neste trabalho, gerando assim informações com maior valor agregado (**QP2**). Esses dados revelam os valores das principais grandezas elétricas inerentes ao consumo de energia, tais como: fator de potência, energia ativa e energia reativa. A identificação de *outliers* nos valores dessas grandezas pode indicar comportamentos anômalos do equipamento. Por exemplo, variações de valores do fator de potência podem indicar defeitos no equipamento, que causam consumo desnecessário de energia.

A arquitetura multinível de fusão de dados proposta neste trabalho é composta por três camadas: camada de baixo nível ou local, camada de nível médio e camada de alto nível ou de gerenciamento. Cada camada trata os dados com um foco específico, buscando garantir maior acurácia dos resultados que

serão apresentados aos usuários finais. Usamos como base os níveis de abstração de dados apresentados por (DASARATHY, 1997).

A camada de baixo nível pode ser bidirecional (monitora e pode agir sobre o ambiente quando o hardware tiver atuadores). Inicialmente, essa camada recebe os dados brutos coletados pelos sensores instalados nos aparelhos, realizando o processo inicial de fusão de dados e identificação inicial de *outliers* (ruídos e erros). Para monitoramento, essa camada pode ser implementada dentro dos nós sensores, ou posteriormente (*off-line*). Caso a camada seja implementada para também agir sobre o ambiente, o processamento deve ser realizado diretamente nos nós sensores.

A camada de nível médio recebe os dados tratados na camada anterior e realiza novas fusões, com o objetivo de derivar informações úteis para a tomada de decisão (i.e., as grandezas elétricas). Os dados são então armazenados em bancos de dados locais ou em nuvem.

Por fim, a camada de alto nível provê a interface com o usuário. Através dessa interface, o usuário pode definir parâmetros de normalidade e regras de funcionamento esperado dos equipamentos, usados nos processos de análise das informações. Essa interface apresenta os resultados gerados pelas análises realizadas e os painéis que apoiam a tomada de decisão. Essas análises sugerem ações ao usuário (por exemplo, realizar manutenção em um equipamento que apresenta variação no fator de potência). Algumas ações também podem ser realizadas autonomamente pela camada de baixo nível, se a implementação da rede prover essa funcionalidade.

Com a identificação de *outliers* a partir do monitoramento contínuo e em tempo real é possível identificar possíveis problemas nos equipamentos e tomar ações corretivas com mais rapidez, evitando desperdícios e gerando economia (**QP3**).

Como prova de conceito, dois estudos de caso foram realizados. Buscando verificar o requisito de independência da solução proposta com relação à aplicação (**QP1** e **QP2**), as RSSFs foram implantadas em dois contextos distintos: um ambiente residencial e um ambiente industrial.

O estudo de caso realizado em ambiente doméstico focou na implementação de uma rede de sensores atrelada a um equipamento eletrodoméstico (geladeira), com vistas a medir o consumo de energia por um determinado período. Utilizando a mesma abordagem, o segundo estudo de

caso foi desenvolvido em um ambiente industrial. O objetivo foi monitorar remotamente o consumo de energia de um equipamento industrial, denominado “extrusora”, em uma pequena indústria de plástico da região Metropolitana do Recife.

Através de pequenas alterações no *hardware* (mudança de transformadores de corrente), foi verificada a independência de domínio da solução proposta (**QP1**). Em ambos os casos, os resultados revelaram ser possível efetuar leituras remotamente com sensores de baixo custo, e enviar esses dados para uma plataforma em nuvem. Através da arquitetura de fusão dados, foi possível obter maior acurácia na análise dos dados a partir da identificação de *outliers* (**QP2**). Com a identificação de outliers a partir do monitoramento contínuo e em tempo real, é possível identificar possíveis problemas nos equipamentos e sugerir ações corretivas com rapidez, evitando desperdícios e gerando economia. (**QP3**).

1.2 ESTRUTURA DO DOCUMENTO

O presente trabalho está dividido em mais 5 capítulos, além da Introdução. Eles estão organizados da seguinte forma:

- No Capítulo 2, são apresentados alguns conceitos que servem como base para o entendimento deste trabalho, com destaque para IoT e RSSF.
- O Capítulo 3 apresenta conceitos sobre *outliers* e Fusão de dados, trazendo ainda uma breve revisão de trabalhos relacionados, que têm como ideia central arquiteturas de fusão de dados implementadas sobre RSSF e fazendo uso de abordagens de IoT.
- O Capítulo 4 descreve em detalhes a arquitetura de fusão de dados multinível proposta neste trabalho, que foi utilizada para construção de duas RSSFs para monitoramento e consumo de energia.
- O Capítulo 5 apresenta dois estudos de caso realizados detalhando a Metodologia e desenho experimental, além da descrição do ambiente, coleta de dados para avaliação, detalhes da análise e validação da arquitetura de fusão de dados proposta.

- No Capítulo 6 são exibidas as considerações finais e indicações de trabalhos futuros.

2 ÁREAS CORRELATAS - IoT e RSSF

Este capítulo apresenta uma visão geral de áreas fortemente relacionadas ao tema da pesquisa desenvolvida no mestrado. A seção 2.1 traz um panorama da IoT, apresentando conceitos básicos, tipos de redes e aplicações. Essa é uma área de interesse, por englobar o tema central desta pesquisa.

Em seguida, a seção 2.2 trata das redes de sensores sem fio, com uma breve explicação sobre os conceitos principais abordados nesta dissertação.

2.1 INTERNET DAS COISAS (IoT)

Veremos a seguir uma descrição geral sobre a área de IoT, incluindo conceitos como computação móvel, computação pervasiva e computação ubíqua. A seguir, temos um detalhamento sobre os tipos de redes existentes. Por fim, veremos aplicações nessa área de interesse.

2.1.1 Visão Geral

IoT é acrônimo do termo inglês “*Internet of Things*” (Internet das Coisas) e foi citado pela primeira vez por Kevin Ashton, pesquisador do *Massachusetts Institute of Technology* (MIT) em uma apresentação da empresa Procter & Gamble no ano de 1999. Na ocasião, Ashton apresentava a sua ideia de etiquetar produtos da empresa com RFID (do inglês, *Radio-Frequency Identification* – identificação por radiofrequência), com o intuito de melhorar seus processos logísticos e de produção. Posteriormente, foram adicionados sensores e atuadores, mesmo com limitações de processamento, energia e memória.

A evolução em áreas como eletrônica, microeletrônica, sistemas embarcados, comunicação e sensoriamento tornaram os custos de *interfaces* de comunicação mais acessíveis, possibilitando a diminuição do tamanho dos componentes eletrônicos. Outro avanço importante foi a introdução de telecomunicações nesses componentes eletrônicos, tornando-os objetos conectados à Internet e inteligentes. Nascia assim o conceito de Internet das Coisas.

Segundo entrevista realizada pela Revista Inovação da Pauta (RANGEL, 2014), Kevin Ashton, ao ser questionado sobre “do que se tratava a IoT”, respondeu que se baseava na ideia de que a Internet e o mundo real das coisas precisavam se comunicar; que o computador não seria mais usado e sim passaria a “se usar”, com o objetivo de tornar as atividades do dia-a-dia mais eficientes. As “coisas”, ou objetos estariam conectados entre si de forma inteligente e passariam a perceber o mundo ao seu redor.

Em outras palavras, IoT trata-se da interconexão de objetos do cotidiano com a Internet, ou ainda uma rede entre objetos capaz de reunir e transmitir dados.

Computação móvel se baseia na portabilidade dos recursos computacionais que, como dito acima, tiveram seu tamanho diminuído, bem como na oferta de acesso permanente à Internet em qualquer lugar e a qualquer hora (ARAUJO, 2003).

Por sua vez, Computação pervasiva oferece recursos computacionais distribuídos no ambiente de forma perceptível e/ou imperceptível ao ser humano (ARAUJO, 2003). Dessa forma, a computação está em todo o ambiente (sensores e computadores) e pode extrair informações detalhadas desse ambiente (conectividade plena). Também é possível controlar, configurar e ajustar aplicações para melhor atender às necessidades de um dispositivo ou usuário.

Por fim, destacamos outro conceito intimamente relacionado às duas áreas citadas acima, a computação ubíqua. Segundo (ARAUJO, 2003), “a computação ubíqua se beneficia da computação móvel e da computação pervasiva”, tendo como objetivo tornar a interação homem-computador invisível. Basicamente, ela busca integrar a informática a ações e comportamentos naturais das pessoas, tornando mais fáceis as tarefas do dia a dia.

Nesse contexto, a Internet das Coisas pode ser considerada como uma rede ubíqua (Figura 1) que viabiliza a integração com o mundo físico, através de processos como coleta, processamento e análise de dados oriundos de seus sensores, que estão presentes em todos os objetos/coisas e se comunicando por meio da Internet.

Figura 1- Relação da IoT com Computação Ubíqua, Móvel e Pervasiva



Fonte: Adaptado de (ARAUJO, 2003)

Em geral não se faz mais distinção entre os conceitos de computação pervasiva e ubíqua, sendo a última mais utilizada.

Segundo a Associação Brasileira das Empresas de *Software* (ABES,2019), Internet das Coisas pode ser definida como “uma agregação de pontos de extremidade ou ‘coisas’, que são identificáveis unicamente e que se comunicam através de uma rede sem interação humana usando alguma forma de conectividade automatizada, local ou globalmente”.

Segundo o Ministério de Ciência, Tecnologia, Inovações e Comunicações, (MCTIC,2016), IoT “é uma rede de todos os objetos que se comunicam e interagem de forma autônoma via Internet, permitindo o monitoramento e gerenciamento desses dispositivos via *software* para aumentar a eficiência de sistemas e processos, habilitar novos serviços e melhorar a qualidade de vida das pessoas”.

Um estudo apresentado pela ABES, feito em parceria com a empresa de consultoria IDC Brasil, apresentado em 2019 com dados em 2018, informou que a tendência de crescimento do ecossistema de IoT no Brasil deverá ser de 20% ao ano até 2022. O estudo apontou ainda que o mercado de IoT global deveria atingir US\$ 745 bilhões em 2019 e passará da marca de US\$ 1 trilhão em 2022 (ABES,2019).

O surgimento da IoT está relacionado diretamente a vários avanços tecnológicos em áreas como *software*, *hardware* e infraestrutura de comunicação, que ofereceram a base para sua viabilização, aproveitamento de seu potencial e popularização.

Em *software*, a evolução aproveitou-se dos avanços na área de *hardware*, que propiciou aumento do poder computacional (armazenamento, processamento e memória). Houve também aumento de funcionalidades, padrões de desenvolvimento, aumento de usabilidade (ou facilidade de uso), e facilidade para a criação de novas soluções embarcadas, o que ajudou a democratizar e popularizar a IoT.

Em relação à infraestrutura, houve evoluções das redes e nas telecomunicações, de forma a viabilizar o aumento expressivo de possibilidades de integração.

A possibilidade de integrar objetos do dia a dia, incluindo inteligência em seus processos, gerou uma corrida contra o tempo para as empresas, na busca de se tornar mais competitivas. Muitos setores aderiram a essa nova tecnologia, sendo necessário gerar tipificações diferentes para cada domínio de aplicação.

2.1.2 Tipos de rede IoT

A IoT teve seu crescimento tendo como base muitas tecnologias de comunicação pré-existentes, em sua maioria, por radiofrequência e sem fio. As tecnologias variam e são aplicadas de acordo com sua área de alcance ou cobertura, que podem variar de centímetros a grandes áreas geográficas.

Na Tabela 1 são apresentadas algumas dessas tecnologias, agrupadas por área de cobertura.

Tabela 1. Exemplo de tecnologias IoT (LEITE et al.,2021)

Tipo da Rede	Tecnologia utilizada
P2P (<i>Peer to Peer</i>)	RFID (<i>Radio Frequency Identification</i>) NFC (<i>Near Field Communication</i>)
WPAN (<i>Wireless Personal Area Network</i>)	Bluetooth, Zigbee
WLAN (<i>Wireless Local Area Network</i>)	Wi-Fi, Adhoc, Wi-GIG
WAN-MAN (<i>Wireless Metropolitan Area Network</i>)	WiMAX 802.16 d,e,s, LTE GSM, NB-IoT, SIGFOX, LoRa, MIMO-MAX
WAN-RAN (<i>Wireless Area Satellite Network</i>)	VSATA

Outros avanços tecnológicos tiveram papel importante, seja para a viabilização quanto para o aproveitamento do potencial da IoT. São eles:

- Computação ubíqua

- Computação em nuvem
- Computação cognitiva
- *Big data*
- *Machine learning e deep learning*
- *Analytics*
- Inteligência preditiva e cognitiva
- *Crowdsourcing*
- *Blockchain*
- *Bring your own device (BYOD)*

2.1.3 Aplicações da IoT

Muitos domínios de aplicação se beneficiaram com o crescimento da utilização da IoT, tais como citado por (LEITE et al.,2017):

- Consumidores e residências (infraestrutura, segurança, conforto e conveniência);
- Edifícios (comercial / institucional);
- Indústria (automação de recursos, fluidez dos processos, distribuição);
- Saúde e Ciências da vida (Cuidados, Pesquisa);
- Energia (demanda de fornecimento petróleo e gás);
- Varejo (Loja, hospitalidade e especialidade);
- Transportes;
- Segurança privada e pública (serviços de emergência, infraestrutura pública, rastreamento, equipamentos de vigilância);
- TI e Redes (Corporativas e Públicas) - servidores, armazenamento, PCs, roteadores, *switches* etc.

É fato que a IoT vem realizando uma transformação digital, conectando dispositivos, integrando pessoas/objetos, alavancando tecnologias, incrementando valores de negócios, redefinindo, transformando organizações e gerando enorme quantidade de oportunidades.

Apesar de todos os aspectos já citados de melhoria e oportunidades, existem fatores referentes ao armazenamento das informações geradas por IoT

que precisam ser analisados sob os pontos de vista social e de segurança. Com relação a esses dois aspectos, a IoT pode ser considerada uma tecnologia invasiva, tendo em vista que coleta e gera informações em residências, nas pessoas (dispositivos vestíveis), em empresas, muitas vezes imperceptíveis.

Esses grandes volumes de dados brutos gerados, de uma forma ou de outra, poderão ser analisados, tratados, e poderão fornecer inteligência para diversos negócios. No entanto, para isso é necessário garantir a integridade desses dados brutos. Segundo (LEITE et al.,2017), imperfeições (erro em sensores, calibragem etc.), inconsistências (devido a *outliers*) ou discrepâncias podem surgir nos dados, seja por defeitos dos sensores, erros em redes ou por falha do meio de comunicação. Sendo assim, os algoritmos e aplicações devem ser preparados para lidar com esses desafios.

Como relatado anteriormente, estamos cada vez mais conectados à IoT, sendo necessário, além de garantir a integridade dos dados, também observar os conceitos de conectividade e privacidade sob a ótica da Lei Geral de Proteção de Dados (LGPD). Nesse contexto, e visando o alinhamento com o tema desta pesquisa, serão apresentados em capítulos posteriores conceitos como RSSF e *outliers*.

2.2 REDES DE SENSORES SEM FIO – RSSF

Esta seção traz uma breve apresentação sobre as Redes de Sensores sem Fio, suas topologias, arquitetura, e estrutura interna dos nós. Por fim, veremos uma classificação proposta para essas redes.

2.2.1 Visão Geral

Os grandes avanços em áreas como comunicação sem fio, microprocessadores e microssistemas eletromecânicos (MEMS, do inglês *Micro Electro-Mechanical Systems*) viabilizaram o surgimento e a evolução de novas tecnologias, como as Redes de Sensores Sem Fio (RSSF). Segundo (YINBIAO et al.,2014), uma RSSF é uma rede formada por nós autônomos que, de forma cooperativa, são capazes de detectar fenômenos físicos do ambiente, tais como temperatura, luz, umidade, pressão, sinais acústicos, sísmicos, infravermelho, dentre outros.

Segundo (BENEFIT, 2013), a primeira rede sem fio que tem alguma semelhança real com as RSSFs atuais é o Sistema de Vigilância Sonora (SOSUS), desenvolvido pelos militares dos Estados Unidos na década de 1950 para detectar e rastrear submarinos soviéticos. As décadas de 1960 e 1970 foram de investimentos na área, e apenas na década de 1980 a Agência de Projetos de Pesquisa Avançada de Defesa dos Estados Unidos (DARPA) iniciou o programa Rede de Sensores Distribuídos (DSN, do inglês *Distributed Sensor Networks*) para explorar formalmente os desafios na implementação de redes de sensores distribuídos/sem fio. Com o surgimento desse programa, as redes DSN logo se tornaram tema de estudos científicos. Governos e Universidades iniciaram a utilização de RSSFs em aplicações como monitoramento da qualidade do ar, detecção de incêndios florestais, prevenção de desastres naturais, monitoramento estrutural e em estações meteorológicas.

Na ocasião, os trabalhos valorizavam a funcionalidade e o desempenho das RSSFs, deixando de lado outros fatores como: custos de *hardware* e implantação, padrões de rede, consumo de energia e escalabilidade. Isso inviabilizou a adoção dessas redes em uma gama mais ampla de aplicações. Entretanto, o potencial das RSSFs gerou interesse da academia e da indústria, levando a um esforço conjunto em prol de resolver os desafios da engenharia. Com o tempo, esforços e avanços levaram ao desenvolvimento de RSSFs com nós sensores cada vez menores e mais baratos, capazes de efetuar comunicação sem fio, sensoriamento e de executar instruções.

Segundo (YICK;MUKHERJEE;GHOSAL,2008), podemos dividir as aplicações de RSSFs em duas categorias: rastreamento e monitoramento. Aplicações de monitoramento podem atuar em ambientes internos ou externos, bem como os dispositivos encontrados nesses ambientes. Já as aplicações de rastreamento incluem rastreamento de animais, pessoas, objetos, veículos, dentre outros.

Com os recursos de observação e controle, as RSSF expandem a sua aplicabilidade em vários domínios do mundo real, tais como: monitoração ambiental, gerenciamento de infraestrutura, monitoração e controle industrial, segurança pública, transporte, medicina, aplicações militares, consumos de água, energia, gás, dentre várias outras.

2.2.2 Arquitetura das RSSF

Como já mencionado, RSSFs são redes formadas por nós autônomos que cooperam entre si a fim de detectar/monitorar/controlar um ambiente, geralmente com o mínimo de intervenção humana. Uma RSSF pode ser composta por dezenas, centenas e até milhares de nós sensores, que são dispostos próximo do local que se deseja monitorar (YINBIAO et al.,2014).

Muitos fatores influenciam um projeto de RSSF, dentre eles: escalabilidade, topologia, custo de produção, transmissão, consumo de energia, tolerância a falhas e o ambiente no qual a rede será implantada (RUIZ et al.,2004)). Todos os fatores citados terão requisitos específicos em sua concepção e implantação no projeto. Além disso, todos os fatores citados acima dependerão da aplicação para a qual o projeto será desenvolvido.

A título de ilustração, tomemos como exemplo dois projetos: o primeiro tem o objetivo de monitorar grandezas físicas em áreas remotas; e o outro tem o objetivo de efetuar leitura de energia de uma determinada máquina industrial e, com isso, controlar o seu consumo.

No primeiro caso, para conseguir maior acurácia, o ideal seria espalhar nós sensores em uma grande área geográfica, enquanto que no segundo caso, apenas um único sensor poderá atender a demanda. Ainda, em relação à capacidade de alimentação dos nós, no primeiro caso é pré-requisito a implantação de baterias de longa duração; já no segundo caso, isso não é necessariamente fator determinante, pois é possível aproveitar a alimentação da própria planta industrial.

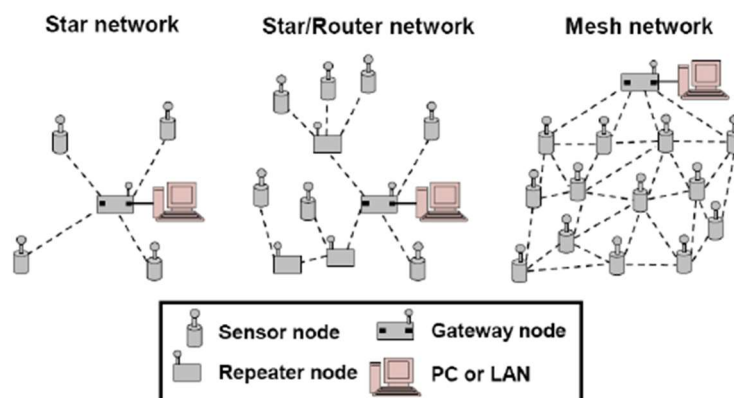
No contexto de RSSFs, um nó é um elemento computacional com capacidade de processamento, memória, *interface* de comunicação sem fio, contendo um ou mais sensores do mesmo tipo ou não (LOUREIRO,2003). Os nós em uma RSSF têm como objetivo analisar o ambiente, de forma a coletar informações de um fenômeno desejado, efetuar algum processamento e compartilhar a informação.

São três as topologias de rede utilizadas em RSSF (Figura 2): estrela (*star network*), árvore (*star/router network*) e malha (*mesh network*). (ANDRE,2017)

Na topologia estrela, os nós transmitem as informações diretamente para um receptor ou *gateway*. Essa topologia tem como limitação a distância de alcance da rede, pois os nós precisam estar no alcance do *gateway*. Na topologia

em árvore ou *cluster* trata a limitação da distância, possibilitando estender a rede, de forma que os nós se comuniquem com outros nós roteadores e estes se comuniquem com o *gateway*. Já a topologia em malha possibilita que cada nó se comunique com outros nós, permitindo redundância de rotas e maior fluidez dos dados.

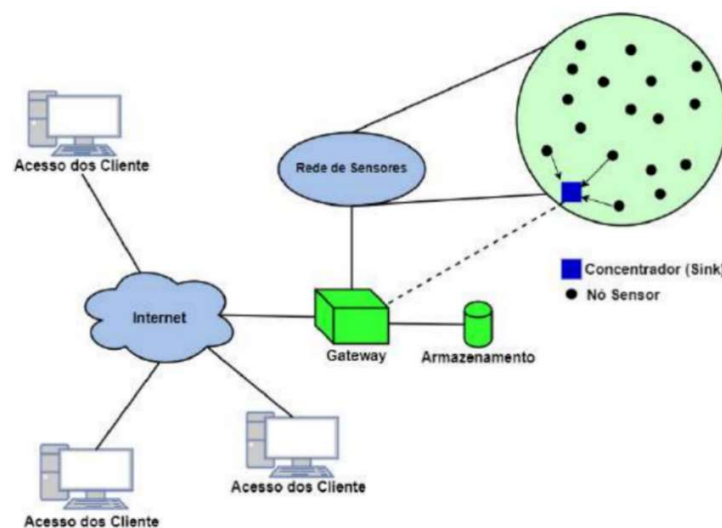
Figura 2- Exemplos de topologia de RSSF



Fonte: (ANDRE,2017)

A Figura 3 ilustra um exemplo de uma arquitetura de RSSF, contendo vários nós sensores, um nó *sink* (também denominado de roteador ou concentrador), um *gateway* e uma unidade de armazenamento. Os nós sensores também se comunicam com o concentrador, que envia as mensagens para o *gateway*, cujo objetivo é integrar a RSSF com outras redes -no exemplo acima, com a Internet. O *gateway* tem papel fundamental no fluxo de informações de entrada e saída, por isso deve ter *interfaces* de comunicação distintas para a RSSF e a rede externa.

Figura 3- Exemplo de arquitetura de RSSF



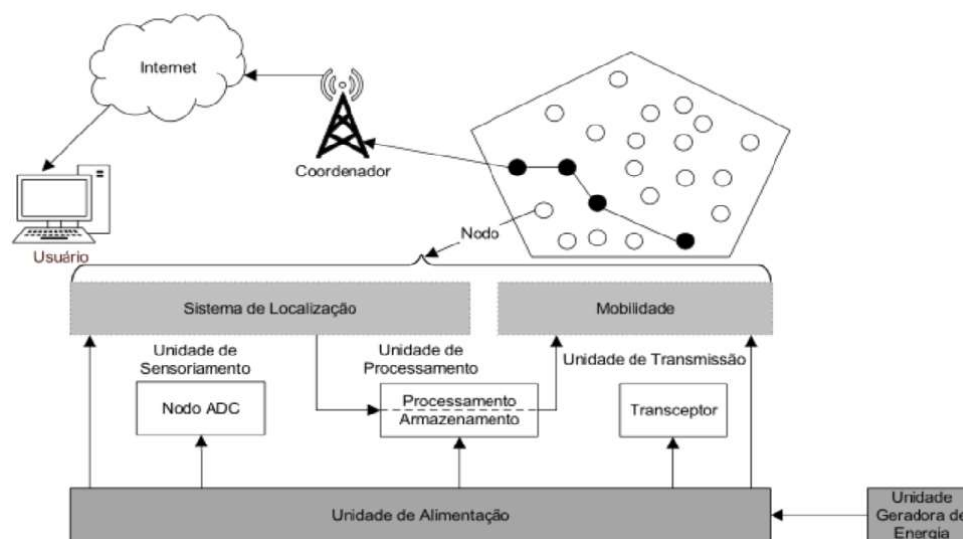
Fonte: (SILVA,2011)

2.2.3 Estrutura do nó

Segundo (AKYILDIZ *et al.*,2002) e (RASSAM *et al.*, 2013), os nós que compõem uma RSSF são dispositivos autônomos equipados com quatro unidades básicas a saber:

- **unidade de sensoriamento** - é aqui que ocorre a leitura de eventos físicos do ambiente monitorado, tais como temperatura, umidade, pressão e luminosidade, transformando essas grandezas físicas em sinais digitais (através de um conversor analógico digital ADC) que serão processados pela unidade de processamento.
- **unidade de processamento** (processador e memória) - responsável por processar as grandezas capturadas dos sensores.
- **unidade de comunicação/transmissão** (emissor e receptor) - responsável por transmitir as informações para o nó *sink* ou concentrador, outros nós ou ainda outras redes através do *gateway*.
- **unidade de alimentação/energia** - responsável por fornecer autonomia energética ao nó.

Figura 4- Detalhes de uma RSF enfatizando o nó sensor



Fonte: (RASSAM *et al.*,2013)

Outras unidades podem ser acopladas ao nó, como por exemplo GPS (do inglês, *Global Positioning System*) para geolocalização, assim como diversos tipos de sensores de temperatura, pressão, umidade e qualidade do ar, sensores de leitura de energia, água, gás, presença, detector de fumaça, etc. Além disso, os nós podem ter capacidade de processamento e se portarem como atuadores, executando tarefas como por exemplo: liga/desliga ou aumenta/diminui a temperatura em um dispositivo de ar-condicionado.

2.2.4 Classificação das RSSF

Segundo (LOUREIRO,2003), as características de uma RSSF devem ser definidas em conformidade com sua aplicação. Assim, os parâmetros de configuração, operação e manutenção variam com o tipo de aplicação em foco.

De acordo com (RUIZ, 2004), uma RSSF pode ser classificada segundo sua configuração, o sensoriamento, o tipo de comunicação e o tipo de transmissão que ela executa.

Para (RUIZ, 2004), na definição da configuração é importante determinar características quanto à composição da rede: se ela será formada por nós com as mesmas características de *hardware*, denominada homogênea, ou se será heterogênea. Em relação à mobilidade, a rede pode ser fixa, quando não existe alteração na posição dos nós ou a rede ser móvel, quando os nós podem se

mover para locais diferentes de onde foram colocados inicialmente. As redes ainda podem ser planas, quando não estão organizadas em grupos ou *clusters*; já as redes hierárquicas encontram-se organizadas em *clusters* e um “líder” para cada *cluster*. Em relação à densidade, as redes podem ser densas, quando concentram muitos nós numa determinada área, esparsas, com pouca concentração de nós, ou ainda balanceadas, onde respeitando o objetivo da rede, há uma distribuição adequada dos nós.

A forma de sensoriamento é baseada na forma de coleta das informações que segundo (RUIZ, 2004), pode ser periódica quando os nós sensores coletam dados em períodos regulares, contínua quando os nós sensores coletam dados continuamente ou reativa, ocorre a coleta dos dados quando há um evento.

Segundo a comunicação, a RSSF pode ter característica de disseminação, quando os nós enviam as informações de forma programada (em intervalos regulares), contínua (enviam dados continuamente), sob demanda (quando é solicitado, seja por um atuador, por ocorrência de eventos, pelo observador etc.). Quanto ao tipo de conexão, a rede pode ser simétrica (nós com mesmo alcance excetuando o nó *sink*), ou assimétrica (nós com alcances diferentes). O tipo de transmissão pode ser *simplex* (os nós sensores são preparados apenas para enviar informações), *half-duplex* (os nós sensores podem enviar e receber informações em um determinado instante) ou *full-duplex* (quando os nós sensores podem enviar e receber informações ao mesmo tempo).

Outras características que são percebidas nas RSSF são:

- **Escalabilidade** - diz respeito ao número de nós da rede e sua capacidade de suportar o crescimento desse número de nós.
- **Auto-organização** - diz respeito à inclusão ou recuperação de nós nas RSSF de forma automática. A posição dos nós sensores não precisam ser projetada ou predeterminada, permitindo com isso a implantação dos nós em áreas remotas e de difícil acesso. Isso obriga as RSSF a possuírem recursos de auto-organização (AKYILDIZ *et al.*, 2002).
- **Conectividade entre os nós** - todos os nós da rede devem ser capazes de transmitir informações.
- **Endereçamento dos sensores ou nós** - dependendo da aplicação, cada sensor tem um identificador único.
- **Agregação dos dados** – é uma técnica de fusão de dados que visa diminuir a quantidade de dados trafegados.

- **Tarefas colaborativas** - um nó sozinho tem uma capacidade limitada de recursos, mas em conjunto ele pode executar tarefas maiores. Possuindo um processador, os nós são capazes de efetuar cálculos simples localmente, antes de enviar os dados aos nós responsáveis pela fusão de dados, transmitindo assim apenas os dados necessários e parcialmente processados, economizando recurso computacional.
- **Limitação da energia disponível** - dependendo da aplicação da RSSF é possível que os sensores sejam alocados em regiões remotas de difícil acesso e, conseqüentemente, sem acesso a “recarga” de baterias ou algo semelhante. Sendo assim, protocolos e algoritmos escolhidos devem considerar minimizar ao máximo o consumo de energia e o uso de recurso computacional.

Como já citado, os nós são distribuídos pelas áreas a serem monitoradas e os seus dados são coletados e encaminhados a uma estação base. Em virtude disso, grande volume de dados brutos precisa ser processado. Isso pode gerar um grande tráfego na rede. Junte-se a isso o fato de os nós terem recursos computacionais limitados.

2.3 CONSIDERAÇÕES FINAIS

Este capítulo foi dedicado a apresentar duas áreas relacionadas ao trabalho de mestrado relatado neste documento (IoT e RSSF). O objetivo foi trazer conceitos básicos utilizados no desenvolvimento deste trabalho.

Foi apresentado um breve histórico de IoT, bem como sua tipologia, tipos de redes e tecnologias, e ainda algumas áreas de aplicação da IoT. Também foi apresentada uma visão geral de RSSF, arquitetura, estrutura de nós, características quanto a configuração, sensoriamento e comunicação.

A seguir, o Capítulo 3 apresenta dois temas fortemente relacionados à pesquisa desenvolvida durante o curso de mestrado: *outliers* e fusão de dados.

3 OUTLIERS e FUSÃO DE DADOS EM RSSF

Como mencionado acima, este capítulo se dedica a apresentar dois temas intimamente relacionados ao trabalho descrito neste documento: anomalias (*outliers*) e fusão de dados. Inicialmente, a seção 3.1 traz uma visão geral sobre *outliers*, as principais causas de anomalias, sua classificação, e abordagens e técnicas para sua detecção. Por fim, a seção 3.2 apresenta o problema da fusão de dados, muito relevante para o trabalho de mestrado desenvolvido.

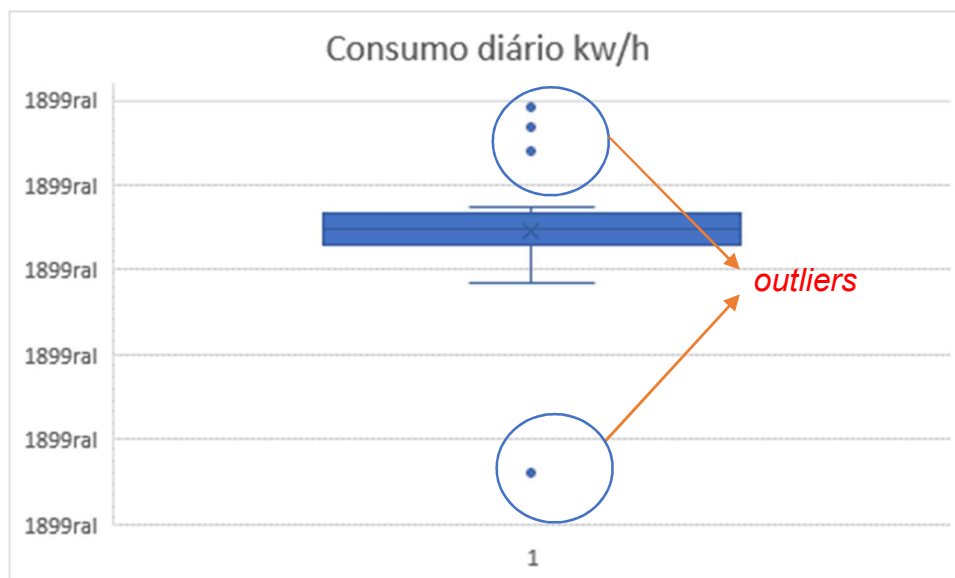
3.1 OUTLIERS

As RSSFs vêm sendo usadas em muitas áreas, como industrial, militar, de saúde, finanças, entre outras. Muitas dessas aplicações são críticas, necessitando de confiabilidade dos dados coletados.

Outliers (dados anômalos, discrepantes ou atípicos), como as próprias denominações expressam, são valores que diferem muito dos demais valores em estudo (BHOJANNAWAR,2013). Já (SICSU,2017) definiu *outliers* como sendo uma observação que se desvia tanto das outras observações que levanta suspeitas sobre se foi gerada por um mecanismo diferente. Segundo (BHOJANNAWAR,2013), *outliers* podem ocorrer devido a erros de *hardware*, codificação, problemas de amostragem ou simplesmente podem ser um resultado atípico justificável.

No contexto de RSSF, *outlier* pode ser considerado uma leitura que parece uma exceção ou uma inconsistência no conjunto de dados lidos por sensores. A figura 5 traz uma representação simples de *outliers* num conjunto de dados.

Figura 5- Representação gráfica *Box Plot* de *outliers* em conjuntos de dados.



Fonte: Autor

Dependendo do contexto, um *outlier* não necessariamente representa um problema ou um erro para análise de dados, mas apenas algo que se deve analisar. Um exemplo disso seria um senso que apresentasse a renda *per capita* de uma determinada região onde predominam pessoas de classe média. Agora, digamos que nessa região morasse Bill Gates ou Warren Buffet. Com certeza o resultado seria alterado e, nesse caso, eles seriam os *outliers* dessa série de dados.

Pela natureza das RSSF, elas estão sujeitas a *outlier*, seja pelos seus nós estarem expostos em áreas hostis ou por terem recursos computacionais limitados (por exemplo, fonte de energia ou bateria). Além disso, RSSFs são normalmente densas, ou seja, são formadas por diversos nós sensores que juntos podem gerar um grande tráfego na rede. Se essa configuração não for bem gerenciada, poderá criar um cenário propenso à ocorrência de *outliers* ou anomalias. Sendo assim, visando minimizar os dados “sujeitos”, evitar contaminação dos dados e garantir que esses dados venham agregar valor são importantes a identificação e o tratamento dos *outliers* o quanto antes.

3.1.1 Tipos de Anomalias / *outliers*

De acordo com (ZHANG,2010), são três os tipos de fontes que causam

outliers:

- **Ruídos e erros** - São normalmente causados por nós sensores com defeito. Eles são aleatórios e devem ser identificados e tratados (corrigidos ou eliminados), pois comprometem diretamente a análise de dados. Normalmente os ruídos são removidos para melhorar a análise de dados.
- **Eventos** - são causados por provável mudança no ambiente no qual está sendo monitorada a grandeza física, normalmente afetando vários nós da rede. Esses *outliers* têm importância para a análise de dados, e devem ser investigados. Um exemplo de evento é um incêndio em uma área monitorada, onde a informação obtida pelos sensores, apesar de atípica é válida e será detectada por vários sensores. Nesse caso, o incêndio é o evento que causou a anomalia de leitura dos sensores, seja de detecção de fumaça, temperatura ou qualidade do ar.
- **Ataques maliciosos** - são *outliers* causados por intrusão na rede com objetivo de contaminar os dados da rede. São importantes para seu domínio de aplicação.

As anomalias podem ocorrer na rede nos nós sensores, nos canais de transmissão e nos dados dos aplicativos. Podem ser causados por erros sistemáticos, erros aleatórios ou ataques maliciosos.

As anomalias de nós ocorrem por falha em um nó, e normalmente têm como origem falhas nos componentes de *hardware* (bateria, processador, memória, sensor etc) ou no *software*. As anomalias de rede ocorrem quando a comunicação entre os nós é interrompida.

A detecção de anomalias de dados depende de irregularidades estatísticas nos dados. Pode se apresentar em três categorias: temporal, espacial e espaço-temporal. Anomalia temporal ocorre variação nos dados de um mesmo nó ao longo do tempo, podendo indicar um evento físico ou um erro no sensor. Anomalia espacial ocorre quando os valores obtidos de um nó sofrem alterações em comparação aos seus nós vizinhos. Anomalia espaço-temporal combinam alterações em relação aos seus nós vizinhos e ao longo do tempo.

3.1.2 Detecção de *outliers*

Como já visto, *outlier* é uma observação que se desvia de um padrão do conjunto de dados ao qual pertence (NAKAMURA,2007). Eles ocorrem em pequenas proporções em relação ao seu conjunto de dados. Uma estratégia de detecção de *outliers* seria definir uma região padrão, onde todo elemento fora dessa região seria denominado de *outlier*.

Segundo (NAKAMURA,2007), a detecção de *outlier* deve considerar alguns fatores, tais como:

- **Natureza dos Dados** - A natureza dos dados está relacionada ao conjunto de dados de entrada, e é um dos principais fatores na seleção de técnicas de detecção de *outliers*. Esses dados de entrada podem ser compreendidos como o conjunto de atributos ou características. Os atributos podem ser tipificados (binários, contínuos, categóricos). As instâncias desses dados podem ser univariável ou multivariável. Quando o conjunto de valores dos dados é de um mesmo tipo, temos a instância univariável (exemplo de leituras apenas da temperatura). Já os dados multivariável ocorrem quando o conjunto de valores tem vários tipos de dados (exemplo, um mesmo sensor ou vários sensores gerando informações sobre pressão e umidade, além da temperatura).
- **Tipos de *Outliers*** - Os *outliers* podem ser considerados globais ou locais em relação ao conjunto de dados. *Outliers* globais são aqueles que representam anomalias em relação a todos os dados disponíveis. Já os *outliers* locais são dados que consistem em anomalias em relação aos seus vizinhos locais mais próximos. Outros autores (BELL,2005) consideram ainda os tipos de *outlier* contextual e coletivo. *Outliers* são ditos contextuais quando a instância de dados é anômala em apenas um determinado contexto. Por exemplo, um dado de temperatura de 25 graus celsius, por si só, não determina ser um *outlier*. Porém, se sabemos que essa temperatura foi medida no inverno de Toronto no Canadá, então existe algo estranho, e possivelmente ele dado será um *outlier*. Esses tipos de *outliers* também podem ser considerados condicionais, pois estão condicionados em um contexto específico. Já os *outliers* são ditos coletivos quando um subconjunto dos dados se desvia como um todo do conjunto de dados. Um bom exemplo disso é uma fábrica que recebe o

pedido de produção, executa e despacha para entrega. Se ocorrer um atraso em um dos pedidos, isso não pode ser considerado um *outlier* coletivo. Contudo, se no mesmo período de tempo existirem 100 pedidos atrasados, isso pode ser considerado *outlier* coletivo.

- **Disponibilidade de Rótulos** - Para o contexto de detecção de *outliers*, a disponibilidade de rótulos está vinculada a um conjunto de dados ter instâncias classificadas como normais ou anômalas. Dependendo do grau em que os rótulos estão disponíveis, a detecção de *outliers* pode operar de três maneiras: (1) O modo supervisionado assume que o conjunto de treinamento contempla rótulos normais e anômalos; (2) O modo semisupervisionado compreende que o conjunto de dados contém somente dados com rótulos normais; (3) Já o modo não supervisionado não exige dados de treinamento (com rótulos), mas considera que os dados normais são mais frequentes (BELL,2005).

Segundo (WANG *et al.*,2019), em RSSFs, os modos de operação para as técnicas de detecção de *outlier* são *on-line* e *off-line*. O modo está relacionado ao local onde se dará o processamento da técnica de identificação de *outliers*.

- *On-line*: nesse modo, a detecção de *outliers* é feita após a leitura no nó sensor. Esse modo exige um poder de processamento ainda na RSSF, havendo com isso restrições de técnicas com base na disponibilidade de recurso computacional (*hardware* e *software*).
- *Off-line*: Nesse modo, a detecção é feita após um período de tempo, quando os dados forem enviados para uma base de processamento centralizada. Como há um atraso na análise, essa abordagem não é aconselhada para aplicações críticas, como detecção de incêndio, monitoramento de pacientes, dentre outras.

3.1.3 Abordagens e Técnicas de Detecção de *outliers*

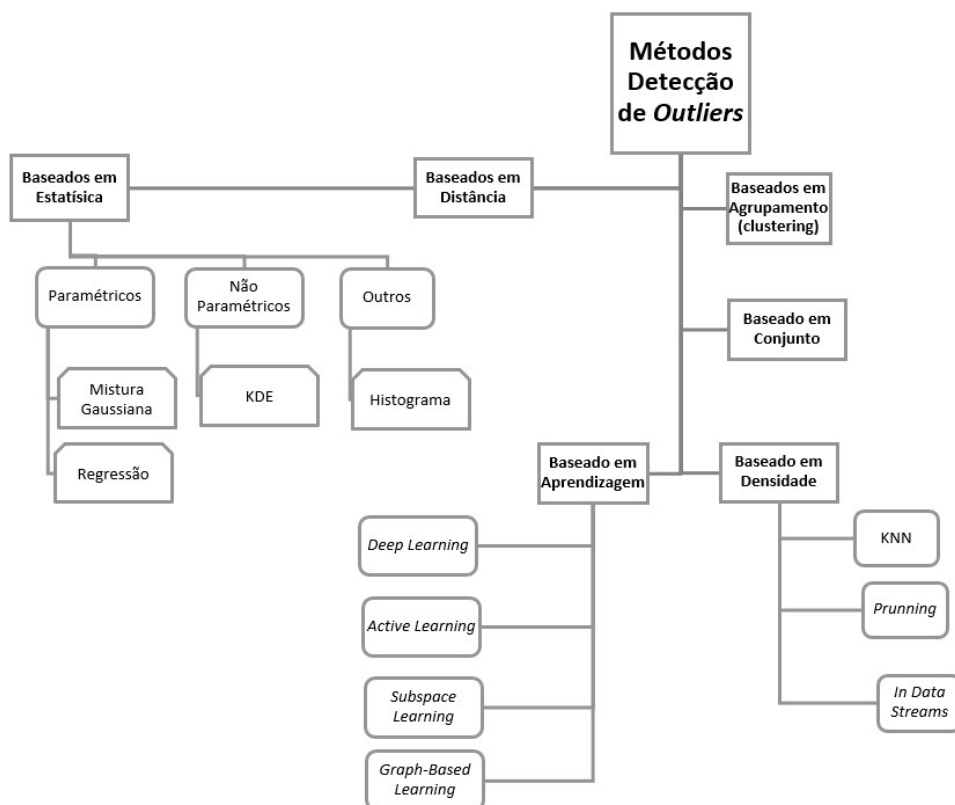
Na literatura relacionada, são relatadas diversas técnicas de detecção de *outliers* ou anomalias. Para qualquer uma das técnicas é relevante considerar as propriedades de redução dos dados trafegados, modo de detecção e a arquitetura, como segue:

- **Redução dos dados que trafegam na rede** - Em virtude das limitações de recursos computacionais, as técnicas devem considerar redução dos dados que são transmitidos pela RSSF. Os dados que trafegam nas RSSF variam de acordo com sua dimensão e podem ser univariados e multivariados.
- **Modo de detecção** – sempre que possível, as técnicas de detecção de *outliers* devem considerar o tempo de verificação como fator primordial. Sendo assim, a detecção preferencialmente deverá ser feita praticamente em tempo real. Para isso, as soluções devem utilizar o modo de detecção *on-line*.
- **Arquitetura** - As técnicas atuais fazem uso de modelos estruturais de três tipos: centralizado, distribuído e local (WANG *et al.*,2019).
 - A arquitetura local ocorre quando os próprios nós, individualmente e dentro de suas limitações, executam a detecção de *outliers*. Nessa arquitetura, os nós devem utilizar seus históricos de leituras e podem também fazer uso leituras de seus vizinhos, efetuando assim uma correlação espaço-temporal.
 - Na abordagem centralizada, a detecção ocorre após os nós efetuarem suas leituras e enviarem a uma estação base, normalmente com maior capacidade computacional, que efetuará a detecção e o tratamento dos *outliers*. Em virtude de tratar uma grande massa de dados, essa abordagem pode usar técnicas que normalmente requerem maior poder de processamento e histórico de outras leituras. A desvantagem é fazer uso maior da infraestrutura da rede, podendo causar um maior tráfego na rede.
 - Já na abordagem distribuída, cada nó tem um agente de detecção que monitora o comportamento de seus nós vizinhos, e dentro de uma faixa de transmissão conseguem detectar qualquer comportamento fora do padrão.

Em uma pesquisa recente, os autores de (WANG *et al.*,2019) apresentaram uma revisão abrangente e organizada do progresso dos métodos de detecção de *outliers* de 2000 a 2019. As técnicas de detecção de *outliers* visam “limpar” os dados coletados pelos nós sensores de forma a melhorar as informações que auxiliarão em decisões junto aos usuários. Ao longo do tempo, muitas técnicas

foram projetadas para identificar *outliers*. Em (WANG *et al.*,2019), os autores dividiram as técnicas de detecção de *outliers* em seis grupos principais, baseados em Aprendizagem, Conjunto, Agrupamento, Estatística, Densidade e Distância (ver Figura 6).

Figura 6- Taxonomia de técnicas e métodos de detecção de *outliers*.



Fonte: Adaptado (WANG *et al.*,2019)

Em (WANG *et al.*,2019), os métodos são descritos como sendo estatísticos, baseados em densidade, baseados em distância, agrupamento, em conjuntos e em aprendizagem como se segue:

- **Métodos estatísticos** - Na abordagem estatística, são criados modelos de dados de referência. O valor de cada leitura é comparado a esse modelo e, havendo desvio, a leitura será classificada como anômala. Os autores ainda classificaram as pesquisas que usaram abordagem estatística em três categorias – métodos paramétricos, não paramétricos e outros tipos. Essa classificação, descrita abaixo, está baseada na disponibilidade de haver conhecimento prévio do modelo de referência (CHANDOLA,2009).
 - Os métodos paramétricos assumem o conhecimento prévio do

modelo de dados. Esses métodos dividem-se ainda em modelo de mistura gaussiana, que assume uma distribuição normal dos dados, e o modelo de regressão, que utiliza outras funções de densidade.

- Os métodos não paramétricos desconhecem, a princípio, o modelo de referência, e utilizam a técnica de estimador de kernel (KDE).
 - Existem outros métodos, sendo os histogramas um dos mais conhecidos, pela simplicidade em detectar *outliers*. Segundo (CHANDOLA,2009), histogramas fazem a contagem da frequência de ocorrência dos valores lidos pelos sensores, estimando a probabilidade de ocorrência dessa leitura, comparando-a a cada categoria no histograma e avaliando se pertence a alguma delas.
- **Métodos baseados em densidade** - O princípio básico desses métodos é que um *outlier* pode ser encontrado numa região de baixa densidade, enquanto *inliers* estão em uma vizinhança densa.
 - **Métodos baseados em distância** - Os métodos baseados em distância detectam *outliers* através das distâncias entre os pontos. Um ponto é considerado *outlier* porque está longe de seus vizinhos mais próximos.

Esse é o princípio básico dos algoritmos de detecção que se concentram no cálculo da distância entre as observações. As técnicas mais utilizadas nessa abordagem são *K-Nearest Neighbor (KNN)*, *In Data Streams e Pruning*.

- **Métodos baseados em Agrupamento (*clustering*)** - Nesse tipo de abordagem, os dados são agrupados em *clusters* com base em suas distâncias de similaridade e comportamentos similares. Um ponto é considerado *outlier* se não pertencer a nenhum *cluster* ou se um *cluster* for significativamente menor que os outros (SHAHID; NAQVI; QAISAR, 2012)(SHAHID,2015).
- **Métodos baseados em conjunto** - Visando detectar *outliers* de forma mais eficiente, os métodos de conjunto (*ensemble*) focam na ideia de combinar os resultados de modelos diferentes para produzir modelos

mais robustos. Eles ajudam a responder se um *outlier* deve ser baseado em modelo linear, baseado em distância ou em outro tipo de modelo.

3.2 FUSÃO DE DADOS

Esta seção tem como foco descrever de Fusão de dados, que é um dos tópicos mais relevantes para esse trabalho de mestrado.

3.2.1 Visão Geral

O termo fusão de dados surgiu na literatura pela primeira vez em 1960, no contexto de modelos matemáticos para manipulação de dados. Posteriormente, em 1986, o Departamento de Defesa dos Estados Unidos criou o grupo de trabalho *Joint Directors of Laboratories* (JDL) com a missão de tratar as principais questões inerentes à fusão de dados, dentre elas a unificação de terminologias e procedimentos. Em 1991, o JDL definiu um léxico, com vistas a facilitar a troca de informações e a cooperação dentro da comunidade de fusão de dados por meio da comunicação e da compreensão (*U.S Departamento Defense* 1991) (WHITE,1991).

Nesse sentido e conforme citado em (BOSTRÖM et al., 2007), o JDL define a fusão de dados como sendo:

“um processo que lida com a associação, correlação e combinação de dados e informações de fontes únicas e múltiplas para alcançar posições refinadas e estimativas de identidade, e avaliações completas e oportunas de situações e ameaças, e seu significado. O processo é caracterizado por refinamentos contínuos de suas estimativas e avaliações, e pela avaliação da necessidade de fontes adicionais, ou modificação do próprio processo, para alcançar melhores resultados”.
(BOSTRÖM et al., 2007)

Em (DSTO,1994), os autores definem a fusão de dados como sendo “um processo multinível e multifacetado que lida com a detecção automática, associação, correlação, estimativa e combinação de dados e informações de fontes únicas e múltiplas”.

Em (BOSTRÖM et al., 2007), os autores estudaram várias definições

existentes para a fusão de dados, fusão de sensores e fusão de informações, buscando analisar se elas estabeleciam o objetivo de forma clara para o campo de pesquisa e fundamentos científicos. Constataram que várias definições continham expressões que as deixavam ambíguas e não claras. Dessa forma, os autores apresentaram uma nova definição, que será considerada nesse trabalho (BOSTRÖM et al., 2007):

“A fusão de informações é o estudo de métodos eficientes para transformação automática ou semiautomática de informações de diferentes fontes e em diferentes momentos em uma representação que fornece suporte eficaz para tomada de decisão humana ou automatizada.”

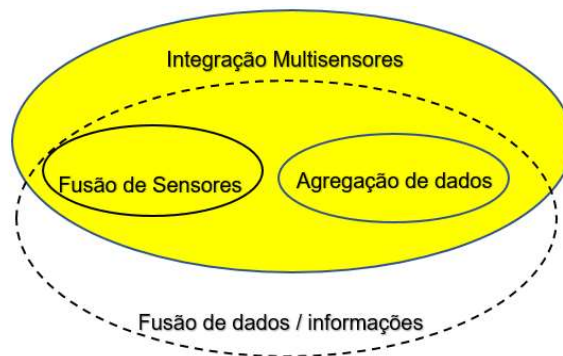
De forma geral, no contexto deste trabalho, que tem como foco a fusão de dados em redes de sensores sem fio, fusão é um termo que define o uso de técnicas que combinam dados brutos de múltiplas fontes, reunindo as informações para colaborar com a tomada de decisão. Dessa forma é esperado que as informações resultantes da fusão sejam melhores, completas e com maior acurácia do que as originais. A melhoria na qualidade das informações é sem dúvida um dos benefícios esperados pelas técnicas de fusão de dados.

A fusão pode ocorrer em sistemas distribuídos de forma *on-line*, ou seja, nos próprios nós sensores, usando os seus dados e de seus vizinhos mais próximos. Nessa abordagem, vale ressaltar a limitação computacional dos nós sensores. Já na abordagem centralizada, todos os dados são enviados para um nó responsável pelo processamento, ou ainda para uma área de processamento central (exemplo servidor em nuvem), onde ocorrerá o processamento e a fusão.

3.2.2 Fusão de informações, fusão de sensores e fusão de dados

Como relatado anteriormente, a fusão de dados é bastante estudada em áreas distintas e contextos diferentes. Isso vem causando certa confusão com relação à terminologia utilizada, segundo (NAKAMURA, 2007). O autor também fez uma comparação entre os termos sinônimos de fusão de informações e suas relações (ver Figura 7).

Figura 7- Relacionamento entre termos ligados a fusão.



Fonte: Adaptado (NAKAMURA,2007)

No contexto de RSSF, os termos fusão de dados e fusão de informações são sinônimos, e dizem respeito a adquirir dados através de múltiplas fontes, inclusive sensores. Segundo (NAKAMURA,2007), a fusão de informação é um termo que engloba fusão de sensores e agregação de dados, tratando da fusão de dados brutos oriundos de múltiplas fontes de dados, inclusive de sensores. Fusão de sensor é apresentado como um subgrupo de fusão de dados que lida com a união de fontes sensoriais. Já a agregação de dados, outro subconjunto de fusão de dados ou de informações é um termo que trata da capacidade de resumir e, conseqüentemente, reduzir esses dados, que são obtidos de múltiplas fontes de dados.

A integração multissensor é um termo que se refere à aplicação da fusão de informação usando sensores e informações associadas para interagir com o ambiente.

3.2.3 Limitações

Em RSSF, as falhas decorrentes de problemas de conectividade, problemas em sensores e eventos isolados, são comuns. Além disso, existem limitações inerentes à cobertura espacial e temporal. A limitação espacial diz respeito à capacidade limitada de detecção de um nó sensor em uma dada região. Por exemplo, um sensor que mede a temperatura em uma sala está restrito ao seu alcance, ou seja, a área que está próxima a ele, e isso pode não representar a temperatura geral dentro da sala.

A cobertura temporal, segundo (NAKAMURA,2007), diz respeito à capacidade do nó de cumprir o propósito da rede enquanto ela existir. Ou seja,

se uma rede visa detectar eventos, deve-se garantir que nenhum evento relevante seja perdido.

Nesse contexto, a fusão de informações pode ser utilizada como ferramenta para suprir algumas limitações apresentadas em RSSF, tais como falhas de sensores, limitações de tecnologia, ou cobertura espacial e temporal.

Segundo (DURRANT-WHYTE,1990), para ajudar na solução das limitações citadas acima, as RSSF devem garantir três propriedades: cooperação, redundância e complementaridade. A cooperação é indicada quando se deseja cobrir uma determinada área a ser monitorada, na qual cada sensor é responsável por cobrir uma pequena parte dessa área. Dessa forma, a fusão pode ser utilizada para se obter uma visão mais abrangente, composta por todas as leituras de todos os sensores que cooperaram para essa visão completa. A redundância torna as informações menos vulneráveis. Com isso, a premissa é ter leituras de nós sensores de uma mesma área e grandeza de forma que, combinando essas leituras, obtenha-se uma maior acurácia, quando comparada com a leitura de um único sensor. Já a complementariedade utiliza-se de vários sensores, efetuando leituras de diferentes grandezas com vistas a efetuar inferências de novas características, o que seria impossível a partir de um único nó.

Em virtude das propriedades de colaboração e redundância, as RSSF são compostas por um grande número de nós, que normalmente ocasionam um tráfego intenso com colisões e transmissões redundantes de dados. Deste modo, a fusão traz benefícios ao reduzir a carga causada pela transmissão excessiva na rede, podendo colaborar inclusive com a economicidade de energia e a vida útil da rede.

3.2.4 Classificação de fusão de dados / informações

A classificação de fusão de informações pode ser baseada em diversos aspectos, tais como relacionamento entre fontes de dados, níveis de abstração de dados e com base nos níveis de abstração de entrada e saída de dados.

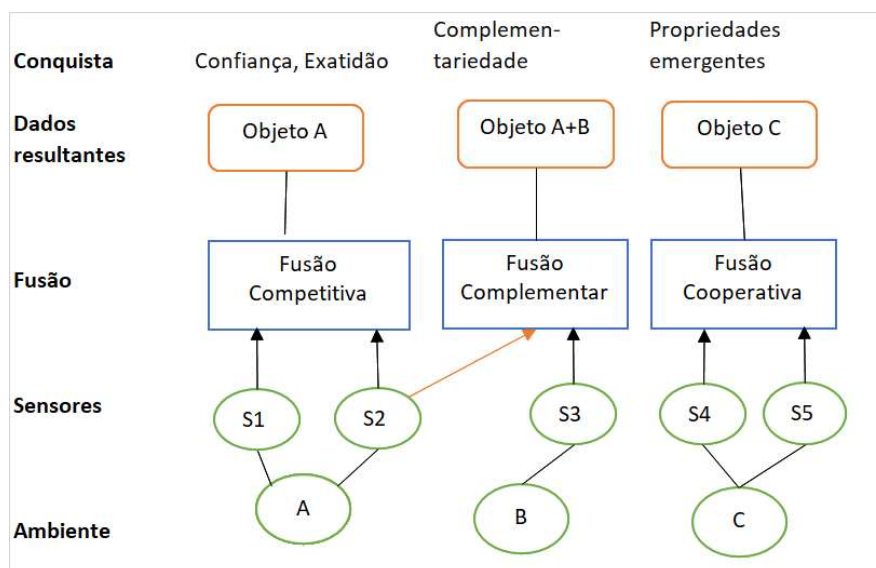
3.2.4.1 Classificação baseada na relação entre fontes de dados

Segundo (DURRANT-WHYTE,1990), de acordo com a relação entre as

fontes, não existe dependência direta entre os sensores e a fusão de dados pode ser classificada como:

- **Dados cooperativos** - quando informações de sensores diferentes são fundidas, gerando uma nova informação com maior relevância de acordo com a aplicação. Na Figura 8, os sensores S4 e S5 representam uma configuração cooperativa, pois apesar de observarem o mesmo ambiente C, as suas medições são utilizadas para formar uma observação sobre o ambiente C que não poderia ter sido obtida a partir das medições individuais de S4 ou S5. Um exemplo desse tipo de fusão são os GPS, que combinam a triangulação de informações de latitude, longitude dos satélites.
- **Dados competitivos ou redundantes** - quando os sensores apresentam o mesmo tipo de informação do mesmo cenário. A Figura 8 mostra a observação do ambiente A pelos sensores S1 e S2. Essas partes são fundidas para aumentar a acurácia. Um exemplo disso é a implementação de mais de um sensor de temperatura num mesmo local, para garantir uma melhor acurácia da informação.
- **Dados complementares** - sensores fornecem trechos diferentes de um mesmo cenário. A fusão, quando aplicada, obtém uma informação mais completa. Na Figura 8, verifica-se que S2 e S3 fornecem informações dos ambientes A e B. O Resultado da fusão é a observação de A + B de forma complementar.

Figura 8- Fusão baseada na relação entre fontes



Fonte: (NAKAMURA,2007)

3.2.4.2 Classificação baseada em níveis de abstração de dados

Em (LUO,2002), os autores apresentam quatro níveis de abstração para classificar a fusão de informações: sinal, pixel, características e símbolo. A fusão no nível de sinal trata de sinais sensoriais de apenas um sensor ou de sinais multissensoriais. Podem ser usados em sistemas de tempo real e tratam basicamente dos dados brutos, tais como: temperatura, pressão, dentre outros.

Já a fusão no nível pixel opera em imagens, sendo usadas para aprimorar a qualidade das imagens obtidas de câmeras e sensores infravermelhos. A fusão no nível de características lida com atributos (feições, traços, padrões, aspectos) extraídos a partir dos dados/imagens de sensores. Por fim, a fusão de símbolo ou decisão permite que decisões sejam tomadas a partir de informações conjuntas oriundas de sensores, conhecimentos a priori e de outros níveis de fusão.

Segundo (DASARATHY,1997), a classificação, quando feita pelo nível de abstração, pode ocorrer em três níveis: baixo nível ou sinal, nível médio ou característica e alto nível ou decisão. Em (NAKAMURA,2007), os autores complementam com mais uma camada, denominada multinível.

- Fusão de baixo nível ou de sinal - As entradas são oriundas de dados brutos oriundos diretamente dos sensores. Esses dados podem ser

combinados em novos dados melhores que as entradas individuais. O dado gerado é da mesma natureza que os dados originais. Alguns métodos que podem ser utilizados para este tipo de fusão de baixo nível são: média aritmética, média ponderada, filtragem Kalman, sensores em Consenso.

- Fusão de nível médio ou de características – os dados de entrada são a saída da fusão do nível de sinal, onde serão abstraídos a fim de representarem o objeto de forma mais precisa. Métodos que podem ser utilizados para este tipo de fusão são: filtros lógicos, morfologia matemática, álgebra de imagens.
- Fusão de alto nível ou de decisão - nível onde ocorre a fusão de decisões para gerar uma nova decisão ou para aumentar a acurácia de uma decisão. Alguns métodos que podem ser utilizados para este tipo de fusão são: estimativa Bayesiana, teoria dos conjuntos, lógica booleana, lógica Fuzzy.
- Fusão Multinível - Nesse nível, ocorre a fusão com dados de diferentes níveis de abstração. Ou seja, tanto a entrada quanto a saída da fusão podem ser de níveis diferentes (NAKAMURA,2007).

3.2.4.3 Classificação baseada nos níveis da entrada e saída de dados

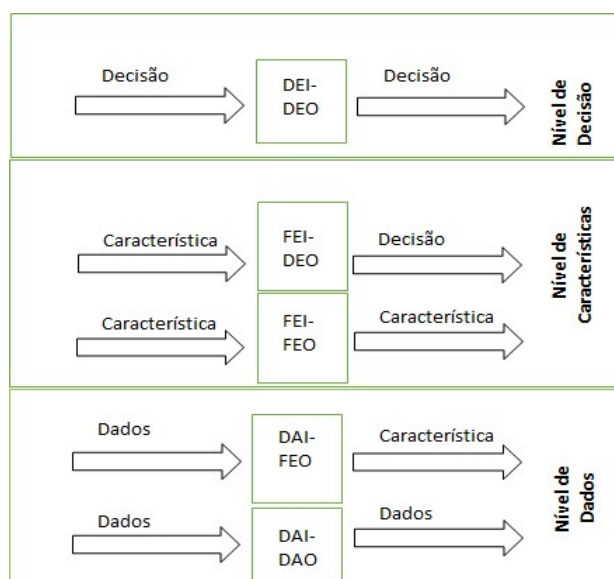
Em (NAKAMURA,2007), os autores classificaram os métodos de fusão em cinco categorias, de acordo com as entradas e saídas, como se segue:

- Dados entram – Dados Saem (*Data In – Data Out* ; DAI-DAO): a fusão de informações lida apenas com dados brutos na entrada e na saída. No entanto, os resultados são possivelmente mais precisos ou confiáveis.
- Dados entram – Características Saem (*Data In – Feature Out* ; DAI-FEO): usa dados brutos de fontes para extrair características ou atributos que descrevem uma entidade. Neste caso, o termo “entidade” significa qualquer objeto, situação ou abstração do mundo.

- Características entram – Características Saem (*Feature In – Feature Out*; FEI-FEO): é aplicado em um conjunto de recursos para melhorar/refinar uma característica ou novas.
- Características Entram – Decisões Saem (*Feature In – Decision Out*; FEI-DEO): nesta classe, a fusão de informação leva a um conjunto de características de uma entidade a gerar uma representação simbólica ou uma decisão.
- Decisões entram – Decisões Saem (*Decision In - Decision Out*; DEI-DEO): decisões podem ser fundidas, a fim de obter novas decisões ou dar ênfase as anteriores.

A Figura 9 apresenta de forma comparativa os métodos baseados em entrada e saída, bem como o baseado nos três níveis de abstração.

Figura 9- Métodos baseados em entrada e saída



Fonte: Adaptado (NAKAMURA,2007)

É importante observar que no método baseado em entrada e saída descrito acima, não é possível utilizar um dado bruto para tomada de decisão. Para haver essa possibilidade, deve-se entrar com características e sair decisões.

De acordo com (NAKAMURA,2007), a principal contribuição da classificação de Dasarathy é “que ela especifica o nível de abstração de entrada e saída de um processo de fusão, evitando possíveis ambiguidades”.

3.2.4.4 Classificação baseada na arquitetura (centralizada x descentralizada)

Essa classificação está diretamente relacionada ao local onde pode ocorrer a fusão. A forma descentralizada ocorre no próprio sensor ou em algum nó concentrador, no entanto, ainda ocorre na rede. Esse tipo de fusão requer um gasto computacional adicional no sensor, além de uma maior complexidade no seu desenvolvimento. A forma de fusão centralizada é feita na Central do Sistema, que executa um processamento mais preciso, alcançando um nível maior de abstração de informações. Isso ocorre pelo fato de ela fazer uso das leituras de todos os sensores da rede, obtendo uma visão mais global. Além disso, esse tipo de fusão utiliza recursos computacionais maiores, o que garante um poder de processamento maior, podendo assim fazer uso de técnicas de fusão mais avançadas.

Na abordagem centralizada, existe uma maior confiança no processamento das informações, uma vez que o sensor pode estar sujeito às condições físicas e mecânicas que podem deteriorar o resultado.

Vale ressaltar que existem sistemas que apresentam tanto fusão centralizada como descentralizada, de forma híbrida, utilizando assim as vantagens de ambas as implementações.

Na literatura, em [HIMEUR,2020] o autor apresenta quatro grupos de arquitetura para sistemas baseados em fusão de dados:

- Arquitetura de computação de borda (*edge computing*): nesta plataforma de arquitetura o processamento e a fusão de informações são conduzidos no nível de borda e executado na camada física, onde vários tipos de informações são realmente coletados. Esta arquitetura faz uso de diferentes plataformas, incluindo microcontroladores, plataformas de computação (Arduino, Raspberry Pi) e plataformas embarcadas de múltiplos núcleos.
- Arquitetura de computação de névoa (*fog computing*): nesta arquitetura, o processamento e a fusão da informação são conduzidos na camada intermediária situada entre a borda e a nuvem.
- Arquitetura de computação em nuvem (*cloud computing*): neste caso, a fusão e o processamento da informação são gerenciados no nível da nuvem. A computação em nuvem apresenta um conjunto de benefícios, como o acesso imediato a fontes de informação *on-line* e *off-line* para

conduzir processamento e fusão adicionais. As principais limitações da computação em nuvem estão relacionadas ao aumento de custos e sobrecargas de comunicação.

- **Arquitetura de computação híbrida:** trata-se do caso em que o processamento de dados é dividido em diferentes camadas, como borda, névoa e nuvem. A depender dos dispositivos de computação disponíveis e os objetivos da aplicação, a fusão e / ou processamento de informações de baixo nível pode ser realizada na borda e / ou névoa, enquanto o gerenciamento de dados de alto nível pode ser realizado na nuvem

3.2.5 Modelos de Fusão

A importância dos modelos para fusão de dados vem do fato deles padronizarem conceitos e os fluxos dos processos necessários para a fusão. Além disso, eles são genéricos, não dependendo de um domínio de aplicação específico (CALLEGARO , 2014).

Muitos modelos foram propostos para auxiliar nos projetos de sistemas de fusão de informações. Segundo (NAKAMURA,2007), esses modelos, inicialmente baseados em informações, evoluíram cronologicamente para modelos baseado em papéis. Segundo os autores, os modelos são agrupados como a seguir:

- **Modelos baseados em informação** - representam os primeiros modelos para fusão de informações, e se concentram no nível de abstração dos dados gerados na fusão
 - Modelo JDL
 - *Dasarathy Model* ou DFD
- **Modelos baseados em atividades** - esses modelos são baseados nas atividades que devem ser executadas pelo sistema de fusão de informações
 - *Loop* de controle de Boyd
 - Ciclo de inteligência
 - Modelo Omnibus

- **Modelos baseados em papéis ou funções** - são modelos que especificam os atores e seus papéis na tarefa de fusão
 - Modelo Orientado a Objetos
 - Arquitetura Frankel-Bedworth
 -

3.2.5.1 Modelo JDL

Este modelo surgiu na área militar, na década de 1980, pelo JDL e Departamento de Defesa Americano, apresentou léxico de terminologias padrões e procedimentos. É composto por cinco níveis de processamento, sistema de banco de dados e um barramento de informações conectando todos os componentes. Sua estrutura está ilustrada na Figura 10 e seus componentes são descritos a seguir.

Descrição dos componentes do modelo, segundo (CASTANEDO, 2013):

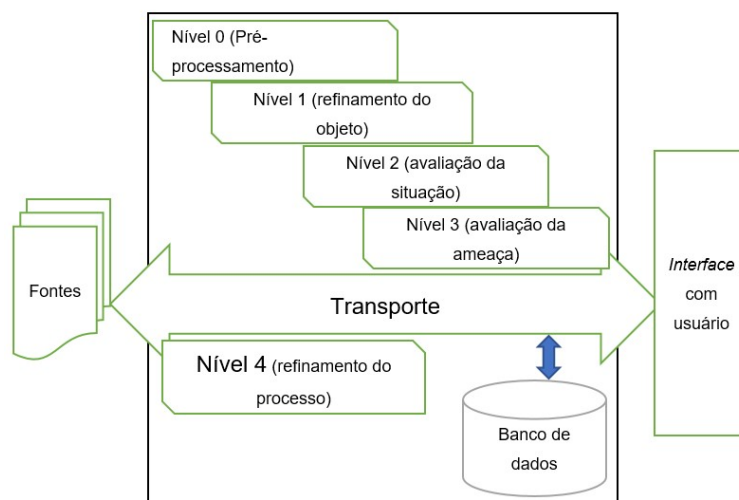
- **Fontes de informação.** O lado esquerdo da figura 10 indica que uma série de fontes de informação podem estar disponíveis como entrada, podendo ser sensores, outros dados (como informações de referência), informações geográficas, banco de dados ou entradas humanas.
- **Interação Humano-Computador (HCI).** O lado direito da figura 10 mostra a função de interação com usuário, ou interação humano-computador (HCI) para sistemas de fusão. É um mecanismo através do qual o sistema de fusão comunica os resultados por meio de alertas, telas de monitoria, *dashboards* e sons.
- **Nível 0** - pré-processamento da fonte (Atribuição de processo). Esta etapa visa observar o objeto (sensor, por exemplo) e a função de pré-processamento dos seus dados, de forma a normalizar, formatar, ordenar e comprimir dados para outros níveis de fusão subsequentes.
- **Nível 1** (Refinamento do objeto). Transforma os dados em uma estrutura de dados consistente ou efetua um refinamento nos dados. Um exemplo é combinar dados de múltiplos sensores com vistas a obter estimativas confiáveis de localização e características. As saídas desse nível são a

discriminação (classificação e identificação) e o rastreamento de objetos (estado do objeto e orientação).

- **Nível 2** (Refinamento da Situação). A partir dos resultados do nível 1, o objetivo nessa etapa é desenvolver uma interpretação contextual dos dados, buscando entender como as entidades se relacionam com o ambiente e com os demais ativos.
- **Nível 3** (Refinamento de ameaças). Preocupa-se em projetar a situação atual no futuro, para identificar potenciais ameaças, vulnerabilidades e oportunidades associadas ao contexto. Técnicas desse nível envolvem simulações, previsões e modelagem.
- **Nível 4** (Refinamento do processo). É responsável por gerenciar e endereçar processos, analisando o resultado de outros níveis, a fim de melhorá-lo em algum aspecto (como precisão, tempo) através de redirecionamento de sensores, fontes de informação e mudança de parâmetros ou tipos de dados. Envolve funções como modelagem de sensores, modelagem de redes de comunicação, medidas de performance e otimização de recursos.

O modelo JDL foi o primeiro a buscar uma terminologia comum e processos para o domínio de fusão. No entanto, ele nasceu especificamente para o contexto militar, e sua terminologia e tarefas de fusão do modelo refletem a especificidade do domínio militar.

Figura 10- Modelo JDL



Fonte: Adaptado (CASTANEDO,2013)

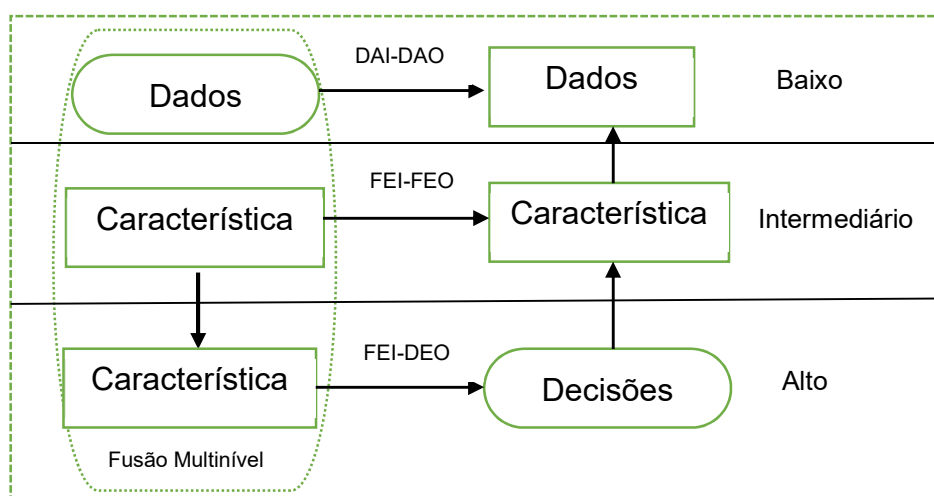
3.2.5.2 Modelo DFD

O modelo DFD (*Data-Feature-Decision*) definido por (DASARATHY,1997) é um modelo centrado em refinamento de informações, no qual os elementos de fusão são focados nas entradas e saídas dos dados. Basicamente, a entrada principal é um dado bruto, que sofre um processamento, gerando uma decisão como saída principal. Os componentes desse modelo responsáveis pelas fusões são os elementos DAI-DAO, DAI-FEO, FEI-FEO, FEI-DEO e DEI-DEO, e já foram devidamente descritos na seção 3.2.4.3.

A vantagem desse modelo se dá porque ele é totalmente constituído por blocos de execução que podem ser interligados ou não, e quanto mais interligados, mais complexa é a fusão dos dados executada. Também ele independe do domínio da fusão, ao contrário do JDL, que é orientado para aplicações militares.

A principal desvantagem é que para gerar blocos de fusão mais complexos, possivelmente os blocos funcionais precisam ser combinados. Um exemplo disso é que não se pode gerar diretamente uma decisão DEO a partir de um dado bruto de entrada DAI. Para isso, devem-se gerar duas operações DAI-FEO e posteriormente o resultado FEI-DEO. A figura 11, apresenta o esquema do modelo.

Figura 11- Modelo DFD



Fonte: Adaptado (DASARATHY.,1997)

3.2.5.3 Modelo Durrant-Whyte

Esse modelo categoriza a fusão de sensores considerando o tipo de configuração dos sensores e Durrant-Whyte (DURRANT-WHYTE,1990) classifica os três tipos de configurações em complementar, redundante e cooperativa. Os tipos de configurações desse modelo de fusão foram detalhados no tópico 3.2.4.1.

3.2.5.4 Modelo OODA (*Observe, Orient, Decide e Act*)

Esse modelo, também conhecido como *Boyd Control Loop* (BOYD,.1987), é baseado em atividades e dividido em quatro fases: Observar (*observe*), Orientar (*Orient*), Decidir (*Decide*) e agir (*Act*), figura 12.

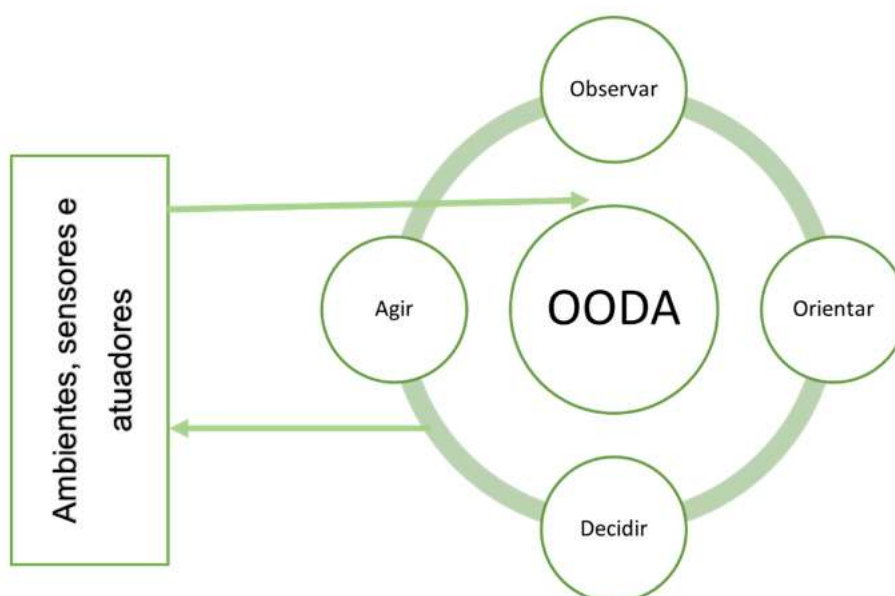
Na fase **observar** o modelo recebe os dados coletados do ambiente. No contexto de RSSF, esses dados brutos são oriundos dos sensores.

Na fase **orientar** ocorre a fusão dos dados para uma avaliação e interpretação da situação atual.

Na fase de **decidir** é especificado um plano de ação em resposta as análises e decisões tomadas em níveis anteriores;

A fase de **agir** é executado o plano definido na fase decidir.

Figura 12- Modelo OODA



Fonte: Adaptado (BOYD.,1997)

Esse modelo tem como vantagens a sua velocidade de execução e seu grau de eficiência para sistemas táticos embarcados. Além disso, poder ser empregado em qualquer domínio de aplicação. No entanto, ele falha ao fornecer uma representação adequada de tarefas específicas de um sistema de fusão de dados (NAKAMURA,2007).

3.3 TRABALHOS RELACIONADOS

Foi realizada uma revisão bibliográfica a fim de identificar os trabalhos mais relevantes e alinhados ao tema desta pesquisa, na área de fusão de dados e identificação de *outliers* onde tivessem como proposta uma arquitetura multinível.

Muitos trabalhos foram encontrados inerentes a construção de tecnologias que tinham como objetivo a eficiência energética a partir de leituras de sensores objetivando o consumo de energia seja em residências, edifícios e ambientes comerciais diverso. Em alguns havia alinhamento com o tema de fusão de dados e identificação de *outliers*, dos quais citamos (CALLEGARO et al.,2014; TORRES,2017; MOHOMED et al.,2008; NESA,2017). Já outros trabalhos, apesar de tratarem de eficiência energética a partir de sensoriamento, pouco tinham aderência com o tema de fusão de dados. Eles focavam apenas na construção de suas arquiteturas e funcionalidades e não implementavam nenhum modelo de fusão de dados e nem tratavam *outliers*. Um bom exemplo disso são os trabalhos de (YOKOGAWA,2020) e (SEVERINO et al. ,2019).

Em (DA SILVA; YOKOGAWA ,2020), o objetivo geral foi o desenvolvimento de um dispositivo capaz de monitorar o fornecimento de energia elétrica, de forma remota e em consequência acompanhar seu consumo. Além disso, aborda a comparação entre as tensões mínimas e máximas obtidas com as estabelecidas pela concessionária de energia. Com isso foi possível efetuar o monitoramento do consumo por 21 dias, demonstrando as oscilações da tensão. Os objetivos foram atendidos, porém como citado, não houve alinhamento com o tema de fusão de dados.

Já em (SEVERINO et al. ,2019), foi apresentado um MVP (do inglês *Minimum Viable Product*) de monitoramento inteligente de consumo de energia de forma não invasiva para uso residencial. A estratégia era analisar as

componentes harmônicas na rede elétrica residencial (tensão eficaz, corrente eficaz, potência real, instantânea, reativa e ativa) de forma permitir a identificação dos aparelhos e se os mesmos estão ligados ou não, além de fornecer informações sobre o consumo de energia da residência. A abordagem se faz interessante, pois, implementa apenas um único ponto de coleta, não necessitando de ter um nó sensor para cada equipamento da casa, no entanto, não foi testado um único ponto de coleta centralizado no cenário real, onde numa residência, normalmente tem vários pares de eletrodomésticos iguais, exemplo, 2 TVs, 2 ventiladores e ligados simultaneamente, o que pode dificultar a identificação unívoca do equipamento. Além disso e apesar do contexto ser similar ao proposto nessa dissertação, consumo de energia elétrica, o trabalho não abordou o tema da fusão de dados e identificação de *outliers*.

Em (CALLEGARO et al.,2014), os autores tiveram como principal motivação saber se um conjunto de sensores de baixo custo conseguiria substituir os tradicionais sensores centralizados e de alto custo, mantendo ou até superando a confiabilidade obtida nos dados monitorados. Os autores ressaltam que com a grande adoção das RSSF, problemas como a grande quantidade de dados e o tratamento de erros (*outliers*) são necessários precisam ser solucionados de forma sistemática. Com isso, propôs uma arquitetura multinível para fusão de informações em 3 camadas: camada de fusão local, camada de fusão de baixo nível e camada de gerenciamento e interface com usuário.

Para testar a arquitetura proposta foi feito caso de uso, onde foi criada uma aplicação de monitoramento ambiental, usando pressão atmosférica (PA) como grandeza principal. Comumente é utilizado para monitoramento da PA, sensores de alto custo.

Basicamente foram implementados sensores de baixo custo e suas leituras, foram comparadas a mais dois sensores de alta precisão e alto custo. Foi implementado tratamento de *outliers*, sendo descartados antes de serem submetidos aos métodos de fusão. Os resultados tiveram desempenho superior quando comparado a um único sensor de alto custo.

Apesar de sua relevância esse trabalho não abordou técnicas de fusão em camadas de níveis superiores, concentrando-se em especificar o funcionamento da camada de baixo nível.

Em (NESA,2017), é realizado um monitoramento remoto do consumo de energia em um edifício através do controle de aquecimento, ventilação e ar

condicionado, com base na detecção de ocupantes. Para detectar a ocupação é realizada a fusão a partir de multisensores das condições ambientais, como temperatura, umidade, luz e CO₂. Para a implementação da solução foi utilizado modelos de classificação tendo como propósito a detecção de ocupação. Os dados brutos de cada sensor são enviados para uma central de fusão de dados, onde é adotada a Teoria da Evidência *Dempster-Shafer* (DSET) (DEMPSTER, 1968) como modelo matemático para fusionar os dados dos sensores com vistas a obter uma conclusão sobre o estado de ocupação. Após obtida a conclusão a mesma é enviada através da Internet para o proprietário do edifício que poderá regular o controle de climatização. Como resultado foi criado um modelo que quando comparado com três outros (RF, GMB, CART) (JAMES et al., 2013) obteve resultado superior. Em consequência o modelo gerado pode ser implementado em qualquer ambiente com vistas a identificar a ocupação, sendo que para isso se faz necessário efetuar um treinamento do modelo.

Em (TORRES, 2017) é apresentado uma arquitetura de fusão de dados em três camadas denominada Hydra, com aplicação no domínio da Agricultura Inteligente. Foram efetuados experimentos para o monitoramento de dois cultivos: caju anão-precoce e coqueiros.

Foram apresentados três experimentos, onde no primeiro foi verificado uma melhora na qualidade dos dados ao efetuar a fusão de sensores de umidade do solo. No segundo experimento foi possível verificar as decisões obtidas baseadas nos eventos detectados. No terceiro experimento foram utilizados métodos de aprendizado de máquina com vistas a criar um novo modelo de evapotranspiração utilizando como base o método *Hargreaves-Samani*.

Em (MOHOMED et al., 2008) é proposta uma arquitetura híbrida para a fusão de dados em três camadas, com vistas a melhoria da eficiência operacional no domínio de monitoramento remoto de paciente. Na primeira camada encontram-se os sensores biomédicos, na segunda é composta por um dispositivo móvel e a terceira camada foi implementada num ambiente remoto onde encontram-se os servidores de aplicações.

Com vistas a diminuir consideravelmente o volume de dados trafegados na rede, é efetuada uma filtragem dos dados. Para a implementação dessa técnica de filtragem foi criado um protótipo denominado HARMONI (*Healthcare oriented Adaptive Remote Monitoring*) .

O funcionamento do protótipo inicia quando os sensores biomédicos conectados a um paciente, coletam os dados e os enviam para um *gateway* (nesse caso um *tablet* ou *smartphone*). O *gateway*, efetua uma filtragem de contexto consciente, utilizando como entrada os dados dos sensores (como exemplo, oximetria e frequência cardíaca) e aplicando regras que são pré-definidas pelo servidor de aplicação. Como resultado da filtragem de contexto, é possível saber quais dados do conjunto de dados, fogem aos padrões de normalidade para aquele paciente. Os dados resultantes da filtragem são encaminhados para os servidores de aplicações.

Esses servidores efetuam o armazenamento e também o processamento desses dados, que consiste na nova análise das regras atuais considerando os novos dados dos sensores e o contexto atual (condição do paciente).

Havendo necessidade, as regras são atualizadas e encaminhadas novamente para o dispositivo móvel que na próxima filtragem irá considerar essas novas regras.

Para avaliar a proposta foi efetuado um estudo onde foi monitorado a frequência cardíaca de usuários em contextos diferentes. O resultado é que a filtragem sensível ao contexto reduziu a quantidade de dados transmitidos pelo dispositivo móvel em até 72%.

3.4 CRITÉRIOS DE AVALIAÇÃO DOS TRABALHOS NA ÁREA

Relembramos aqui as principais Questões de Pesquisa que este trabalho investigou:

- (1) É possível efetuar medição e monitoramento do consumo de energia de equipamentos eletroeletrônicos utilizando redes de sensores de baixo custo, de modo escalável e independente da aplicação (doméstico ou industrial)?
- (2) É possível prover uma maior acurácia na identificação de *outliers* a partir dos dados coletados através de uma arquitetura multinível de fusão de dados, que seja independente de domínio da aplicação?
- (3) A análise final dos dados fusionados (com foco na identificação de *outliers*) auxilia de forma objetiva na tomada de decisão, a fim de favorecer a eficiência energética dos equipamentos?

Tabela 2. Critérios de avaliação dos trabalhos analisados

TRABALHOS RELACIONADOS					
QP		Nesa (2017)	(Torres, A.B., (2017)	(Mohomed, I.; 2016)	(Callegaro, 2014)
QP1					
	Aplicação	Monitoramento remoto do consumo de energia em um edifício a partir da detecção ocupantes.	Hydra Agricultura inteligente (outdoor)	Harmoni Monitoramento de Saúde (indoor)	Monitoramento de ambiente (Outdoor)
	Objeto do monitoramento	temperatura ,umidade, luz e CO2	Umidade do solo, sensor de chuva e temperatura	Oximetria de Pulso e frequência cardíaca	Pressão Atmosférica
	Tecnologia IoT das redes	ZigBee entre sensores e centro de fusão de dados(DC). Wifi entre o DC e a aplicação em nuvem.	Comunicação via radio wireless dos sensores para o nó concentrador e wifi do concentrador à internet.	Sensores comunicam-se com o PDA via Bluetooth PDA via wifi	Não especificada
	Monitoramento remoto e em tempo real	Remoto em tempo real	Remoto em tempo real	Remoto em tempo real	Remoto em tempo real
	Topologia	estrela	estrela com nó concentrador	estrela	estrela com nó concentrador
	Avaliação qualitativa da rede		++	-	
QP2					
	Arquitetura de Fusão de dados	Não informado	Multinível em 3 camadas	Multinível em 3 camadas	Proposta multinível em 3 camadas, no entanto, implementada solução na camada de baixo nível.
	Identificação e descarte de outliers	Não informado sobre tratamento de outliers .	Descarte de ruídos e erros online no sensor; identificação de eventos off-line	Descarte online no PDA e identificação de eventos off-line	Descarte de ruídos no sensor. Não tratou identificação de eventos.
	Melhora da acurácia dos resultados		(+) descarte inicial de ruídos e erros;	(+) descarte inicial de ruídos e erros;	(+) descarte inicial de ruídos e erros;
QP3					
	Definição de regras e parâmetros relevantes para o processo de análise dos dados	Não especificado	regras definidas por parâmetros	O usuário pode definir essas regras a qualquer tempo via interface. As regras geram novos parâmetros	Não implementado
	Visualização dos resultados das análises	Não especificado	Gráficos		
	O sistema realiza ações diretamente no ambiente	Não	Controla o tempo de irrigação do solo online a partir dos dados sensoriais	Não	Não

A tabela 6, pagina 95 apresenta uma breve comparação desses trabalhos (descritos na seção 3.3) com a solução proposta nesse documento.

3.5 CONSIDERAÇÕES FINAIS

Este capítulo foi dedicado ao referencial teórico, abordando o conteúdo necessário para embasar a proposta e o desenvolvimento do trabalho de mestrado. Foi apresentado um breve histórico de IoT, bem como sua tipologia, tipos de redes e tecnologias e ainda algumas áreas de aplicação da IoT. Também foi apresentada uma visão geral de RSSF, arquitetura, estrutura de nodo, características quanto a configuração, sensoramento e comunicação. Em seguida, foi apresentada uma visão de *outliers*, bem como causas, classificação, quanto a detecção de *outliers*, foi visto principais características, abordagens e técnicas.

Foram revisados termos inerentes à fusão de dados, bem como suas limitações, classificação e tipos de modelos. Foram citadas algumas vantagens da utilização de fusão de dados, como a cobertura espacial e temporal, maior disponibilidade proporcionando maior acurácia das informações obtidas.

Por fim, o capítulo apresenta alguns trabalhos relacionados, que foram analisados observando os critérios destacados na tabela 2.

4 AMnFD: UMA ARQUITETURA MULTINÍVEL DE FUSÃO DE DADOS

De forma geral, os processos de gerenciamento de energia nas empresas têm como objetivo otimizar as operações, a fim de reduzir o consumo de energia, os desperdícios e o impacto ambiental. Contudo, como já foi discutido no capítulo de Introdução, a eficiência energética no setor industrial requer investimentos em automação industrial, que geralmente é invasiva e tem alto custo de implantação e manutenção.

As tecnologias relacionadas a IoT têm-se mostrado uma boa alternativa para tratar esse problema, através do uso de redes de sensores sem fio para monitoramento de equipamentos eletroeletrônicos. Os dados coletados, quando fusionados e processados, oferecem informações gerenciais úteis ao controle do consumo de energia elétrica.

Nesse contexto, este trabalho buscou identificar alternativas para alcançar eficiência energética através de soluções de baixo custo, independente da aplicação, seja residencial ou industrial. Como resultado, este trabalho propõe um processo automático de tratamento de dados referentes a grandezas energéticas, coletados por sensores acoplados a equipamentos, com o objetivo principal de identificar dados anômalos (*outliers*), a fim de prover informação gerencial para apoiar e agilizar a tomada de decisão na busca de melhorar a utilização racional de energia. Assim, busca-se obter maior eficiência energética a custo acessível e com portabilidade e escalabilidade.

A seção 4.1 busca relacionar as questões de pesquisa vistas no capítulo de Introdução com os critérios desejados para uma solução que auxilie na obtenção de eficiência energética. A seção 4.2 apresenta a arquitetura proposta, detalhando cada uma das suas camadas. A seção 4.3 apresenta alguns conceitos básicos relacionados à área de gerenciamento de energia elétrica, com o objetivo de facilitar a compreensão do trabalho realizado. Por fim, a seção 4.4 conclui o capítulo com algumas considerações finais.

O capítulo 5 traz uma instanciação da AMnFD para gerenciamento de Energia Elétrica, com detalhes de implementação da RSSF de baixo custo baseada nessa arquitetura. Essa rede foi usada em dois estudos de caso descritos no Capítulo 5.

4.1 REQUISITOS DESEJÁVEIS x QUESTÕES DE PESQUISA

A Tabela 2 (seção 3.4) apresentou os principais critérios/características desejáveis para soluções de apoio à obtenção de eficiência energética: baixo custo de construção e manutenção, independência de domínio, portabilidade, monitoramento remoto e em tempo real, análise de dados fusionados através de *dashboards*, identificação e descarte de *outliers*.

A seguir vamos relacionar esses critérios com as Questões de Pesquisa centrais deste trabalho.

QP1: É possível efetuar medição e monitoramento do consumo de energia de equipamentos eletroeletrônicos utilizando sensores de baixo custo, de modo escalável e independente da aplicação?

- Aqui estamos tratando dos critérios de escalabilidade e independência de domínio com baixo custo.

QP2: É possível prover uma maior acurácia na identificação de *outliers* a partir dos dados coletados através de uma arquitetura multinível de fusão de dados, que seja independente de domínio da aplicação?

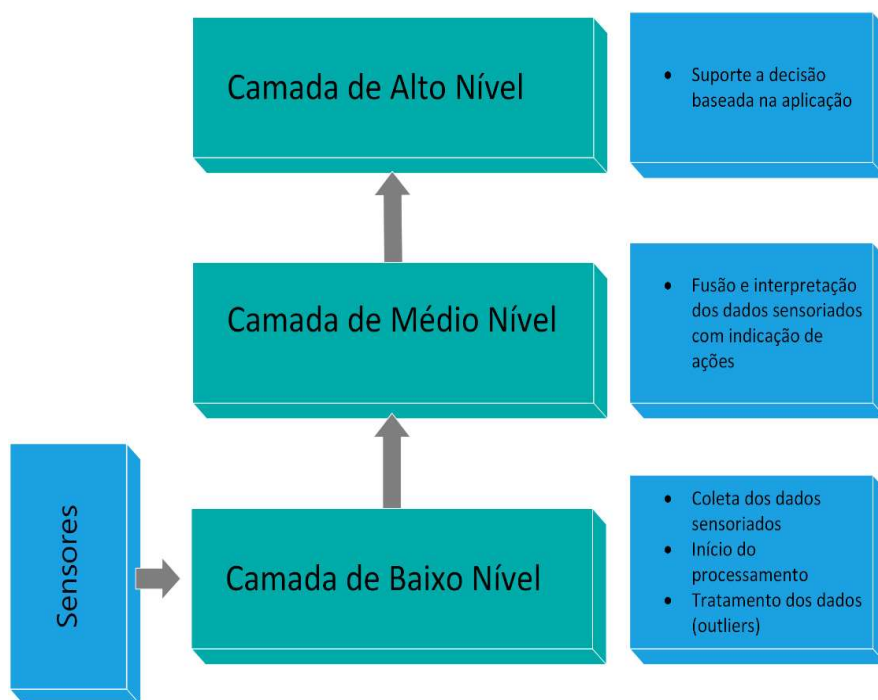
QP3: A análise final dos dados fusionados (com foco na identificação de *outliers*) auxilia de forma objetiva na tomada de decisão, a fim de favorecer a eficiência energética dos equipamentos?

4.2 ARQUITETURA MULTINÍVEL PROPOSTA

Esta seção se dedica à apresentação detalhada da AMnFD, uma arquitetura multinível para fusão de dados proposta neste trabalho (Figura 13). Essa arquitetura é composta por três camadas: camada de baixo nível ou local, camada de médio nível e camada de alto nível ou gerenciamento. Usamos como base os níveis de abstração de dados apresentados por (DASARATHY,1997).

Figura 13- Arquitetura Multinível de Fusão de Dados

Arquitetura Multinível de Fusão de Dados



Fonte: Autor

Como pode ser visto na Figura 13, cada camada trata a informação com foco específico. O processamento dos dados tem início a partir da leitura das grandezas elétricas nos sensores (baixo nível), objetivando o cálculo do consumo de energia, identificação e remoção de *outliers* e, posteriormente, a análise de indicadores para apoio à tomada de decisão (médio e alto níveis).

A AMnFD foi projetada para ser independente do contexto (domínio) da aplicação (por exemplo, monitoramento de consumo de energia, de variações de temperatura, entre outros), e do ambiente (residencial ou industrial). No entanto, alguns fatores devem ser considerados na implementação dessa arquitetura para cada aplicação, mais especificamente, onde será executado o processamento de dados referente a cada camada.

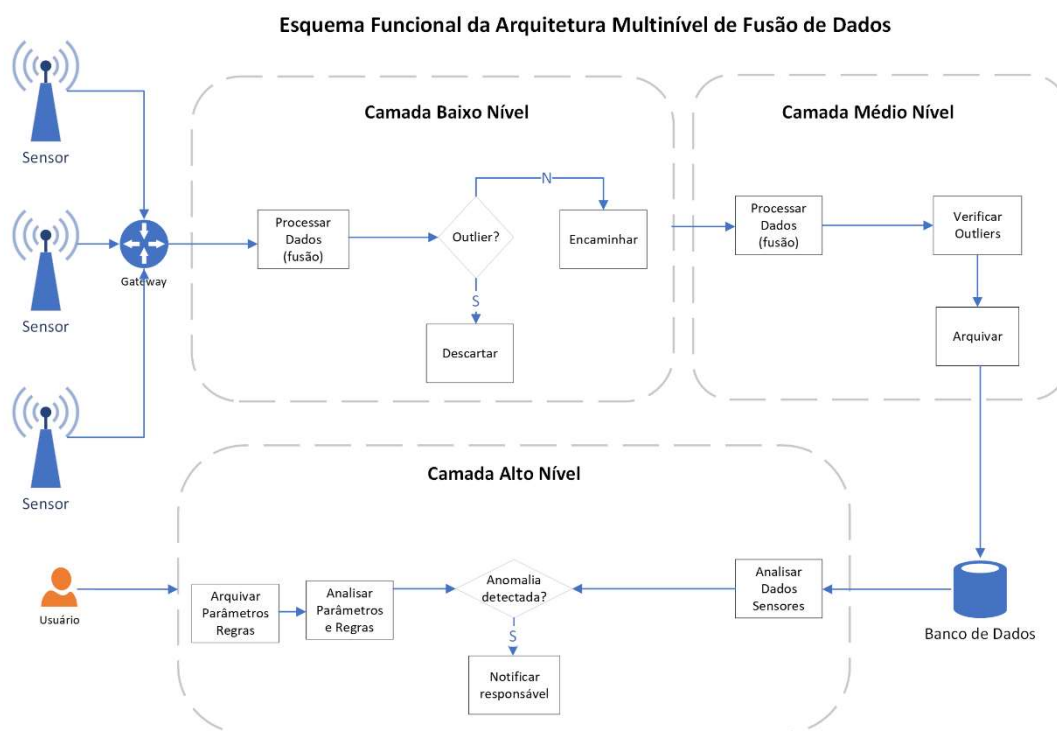
Por exemplo, considere uma rede de sensores implementada para ser usada em locais abertos (*outdoor*), com limitações de conectividade e energia. Essas condições tornam os nós sensores mais dependentes de baterias internas. Contudo, apesar do maior consumo de baterias internas, o processamento dos dados brutos deve ser realizado dentro dos nós sensores ou

nós concentradores (que são mais robustos), devido às dificuldades com conectividade. Assim, é mais indicado implantar as camadas de baixo e médio nível na RSSF. Já o processamento mais pesado, da camada de alto nível, será realizado em um ambiente externo à RSSF.

Já em redes *indoor*, onde presumivelmente não há limitação de energia e conectividade, é possível implementar as camadas de forma distribuída, ou seja, a camada de baixo nível pode ser implementada nos nós sensores e as demais camadas são implementadas fora da RSSF.

A figura 14 apresenta uma visão geral da AMnFD, detalhando os processos executados em cada uma de suas camadas.

Figura 14- Esquema Funcional da Arquitetura Multinível de Fusão de Dados



Fonte: Autor

4.2.1 Camada de Baixo Nível

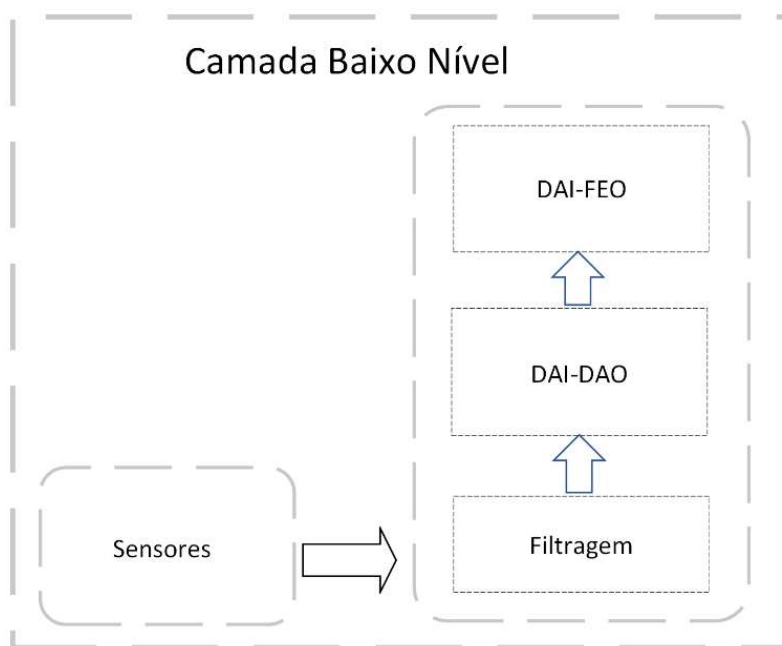
A camada de baixo nível, ou de sinal, recebe os dados dos sensores em seu estado mais bruto. É nesta camada que é feita uma filtragem preliminar, de acordo com a natureza da leitura de cada sensor (por exemplo, grandezas elétricas, temperatura, etc).

A filtragem nessa camada tem o objetivo de verificar se nas leituras há valores considerados díspares (*outliers*) em relação ao escopo de cada sensor. Várias técnicas simples podem ser utilizadas para obter dados mais acurados, como por exemplo: média, mediana, média móvel, dentre outras.

Também é possível realizar fusão dos dados brutos para obter valores mais precisos, e ainda obter outras características (por exemplo, transformar o valor numérico da leitura da temperatura em graus celsius) – ver Figura 15.

Os dados filtrados e processados seguem para a próxima camada, ou camada de médio nível.

Figura 15- AMnFD – Camada de baixo nível



Fonte: Autor

Abaixo temos alguns exemplos que ilustram a fusão de dados na camada de baixo nível para o domínio de monitoramento de energia elétrica:

- **Fusão DAI-DAO:** Dados brutos são gerados a partir de outros dados brutos oriundos dos sensores. Por exemplo, é possível aplicar filtros aos

dados sensoriais referentes a grandezas elétricas, a fim de obter dados brutos mais acurados.

- **Fusão DAI-FEO:** A partir dos dados brutos dos sensores são obtidos dados interpretados. Os sensores podem ser iguais (fusão redundante) ou diferentes (fusão cooperativa) - ver seção 3.2.4. Em nosso contexto de energia elétrica, temos um exemplo de fusão cooperativa: a leitura da corrente em ampere (A) e a tensão em volts (V) resultam na potência em watts (W) ou quilowatts (kW).

Filtragem dos dados – identificação de *outliers*

Muitas vezes, os sensores podem sofrer alterações na leitura dos dados brutos, por motivos tais como ruídos, erros ou eventos válidos (conforme descrito no capítulo 3).

Algumas vezes, os dados brutos obtidos pelos sensores podem apresentar alterações em relação aos valores esperados, tais como ruídos, erros (falha no sensor) ou eventos válidos (conforme descrito no Capítulo 3). Também é importante lembrar que trafegar dados exige um grande recurso computacional, sendo necessário garantir que esses dados estejam dentro dos parâmetros normais do contexto da aplicação.

Medidas simples, como a delimitação de valores válidos de leitura, podem ajudar a identificar e eliminar *outliers* precocemente. Um exemplo em nosso contexto é que as tensões elétricas de corrente alternada utilizadas comumente são de 110 ou 220 volts, podendo sofrer pequenas variações para mais ou para menos. Dessa forma, valores muito diferentes desses podem ser considerados *outliers*. Como a tensão é uma variável utilizada para o cálculo do consumo, uma diferença nessa grandeza pode causar um erro de interpretação no contexto de consumo de energia.

Uma vez identificado o *outlier*, deve-se observar se esse dado deve ser desconsiderado, ou se deve ser considerado para uma análise mais profunda, como por exemplo verificar se trata de um *outlier* causado por evento.

Como mencionado no Capítulo 3, a identificação de *outliers* pode ser *on-line*, ou seja, nos nós sensores ou concentradores, ou de forma *off-line*, onde a verificação é feita posteriormente, nas camadas de médio e/ou alto nível. Os dados que saem da filtragem podem ser considerados dados mais acurados.

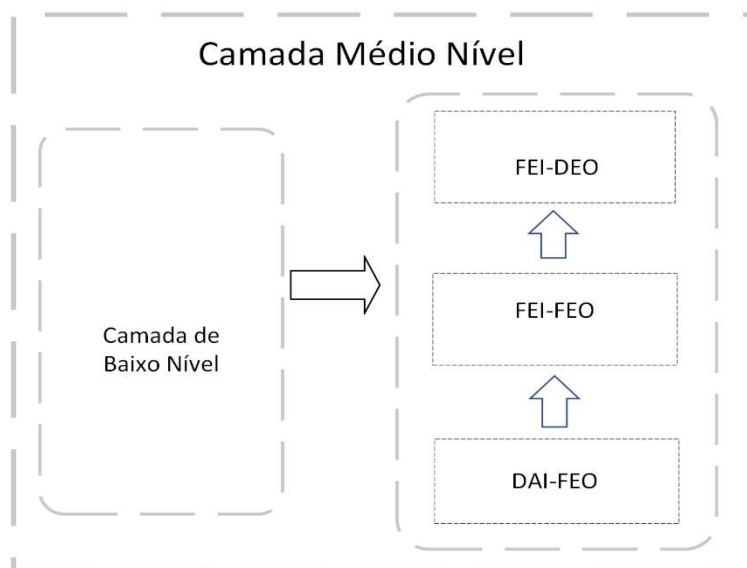
4.2.2 Camada de Nível Médio

Conforme ilustrado na Figura 16, os dados oriundos da camada de baixo nível servirão de entrada para a camada de médio nível, que tem a função de realizar fusão desses dados, de modo a sugerir ações para apoio à tomada de decisões.

Estando de posse das informações dos sensores já devidamente filtradas, é possível efetuar fusões do tipo FEI-FEO e DEI-DEO (camada de alto nível). Assim, características como o valor do consumo de energia podem ser calculadas, possibilitando assim a obtenção de novas características, como por exemplo o fator de potência, o fator de carga, dentre outras. Essas grandezas são detalhadamente definidas na seção 4.3.

A interpretação dos valores dessas características pode ajudar o sistema a realizar ações preventivas, como enviar mensagens ao responsável, informando a variação do fator de potência, e que isso pode indicar uma possível falha no equipamento. As ações referentes à tomada de decisão são realizadas na camada de alto nível, descrita a seguir.

Figura 16- AMnFD – Camada de médio nível



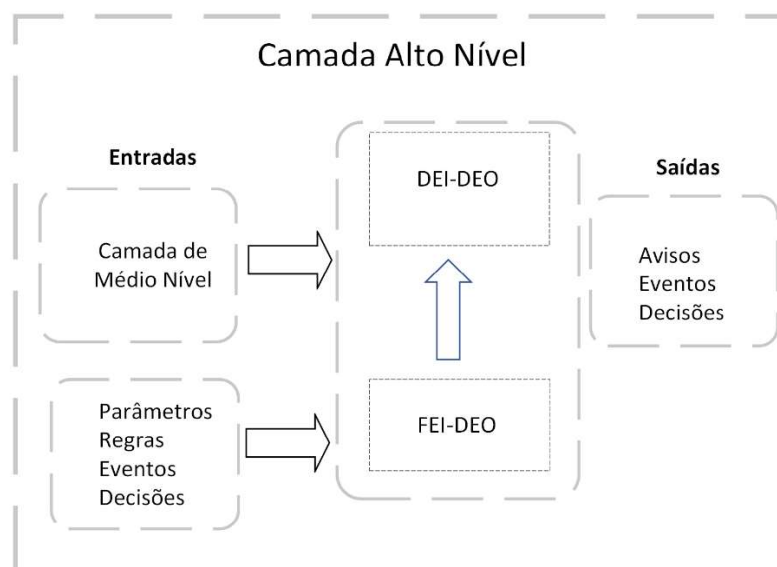
Fonte: Autor

- **Fusão DAI-FEO:** A partir de dados brutos, novos dados interpretados são obtidos. No contexto de energia, grandezas elétricas como fator de potência, energia ativa e energia reativa podem ser obtidas a partir de dados brutos interpretados.
- **Fusão FEI-FEO:** Os dados interpretados são gerados a partir de outros dados interpretados. Por exemplo, a potência, originária do produto da corrente pela tensão, agora pode ser usada para obtenção do consumo, bastando para isso efetuar o produto dessa grandeza por um intervalo de tempo (minuto, hora etc).

4.2.3 Camada de Alto Nível

- A Figura 17 apresenta o esquema da camada de alto nível, responsável pela interface com usuário. Através dessa interface, o usuário pode definir parâmetros que delimitam valores aceitáveis pelos sensores (por exemplo, tensão de 110 ou 220 volts), bem como regras que servirão de base para decisões nas camadas inferiores (se a temperatura exceder 70 graus celsius, enviar mensagem ao responsável pelo equipamento).
- Também é nessa camada que os usuários acompanham os dados gerados e painéis que apoiam em decisões (ver Seção 5.3). As decisões são executadas por meio de eventos configurados pelo usuário.

Figura 17- AMnFD – Camada de alto nível



Fonte: Autor

4.3 CONSIDERAÇÕES FINAIS

Este capítulo apresentou a AMnFD proposta, descrevendo os processos executados em cada uma das suas camadas. Essa arquitetura foi proposta para ser independente de domínio e de contexto (residencial ou industrial). Contudo, como mencionado, nossos estudos de caso foram realizados no domínio de controle de consumo de energia elétrica. Assim sendo, o detalhamento das camadas foi influenciado por essa aplicação.

O capítulo 5, a seguir, apresentará em detalhes a implementação dessa arquitetura com foco no monitoramento do consumo de energia, bem como sua utilização em dois estudos de caso em contextos distintos: residencial e industrial.

Por fim, como veremos a seguir, vale salientar que a AMnFD também foi usada no contexto do monitoramento de ambientes refrigerados (para controle de temperatura). Contudo, os resultados dos testes ainda são preliminares, e não serão apresentados neste documento.

5 IMPLEMENTAÇÃO E ESTUDOS DE CASO

Com base na AMnFD, a arquitetura multinível para fusão de dados proposta neste trabalho, foi desenvolvida uma aplicação com o objetivo de monitorar o consumo energético em tempo real. A partir dos dados de monitoramento, é possível calcular grandezas elétricas ou indicadores úteis, principalmente, para clientes não residenciais. O acompanhamento desses indicadores possibilita disciplinar o consumo racional de energia, auxiliando na identificação de falhas em equipamentos instalados e na tomada de ações preventivas com mais rapidez, evitando desperdícios e mitigando riscos operacionais.

Como resultado direto da implantação da solução proposta, espera-se obter economia em consumo de energia elétrica e diminuição de custos relacionados aos processos rotineiros de manutenção reativa, além de evitar acidentes e incidência de multas nas contas de energia.

A seguir, a seção 5.1 apresenta alguns conceitos inerentes à energia elétrica relevantes para este trabalho. Na seção 5.2 apresenta detalhes de

implementação da aplicação proposta, uma RSSF baseada na AMnFD para monitoramento de consumo de energia elétrica. A seção 5.3 apresenta os dois estudos de caso realizados com a RSSF criada, sendo um estudo realizado em ambiente doméstico (monitoramento de uma geladeira), e o segundo estudo sendo realizado em uma indústria de pequeno porte. O objetivo principal desses estudos de caso foi verificar se essa implementação da AMnFD é capaz de oferecer uma maior acurácia dos dados monitorados, permitindo a identificação de *outliers* considerados como espúrios ou como eventos, auxiliando assim nos processos de tomada de ações corretivas e/ou preventivas a partir do monitoramento dos equipamentos.

5.1 CONCEITOS BÁSICOS NA ÁREA DE GERENCIAMENTO DE ENERGIA ELÉTRICA

A AMnFD proposta acima foi instanciada para o contexto de monitoramento de energia elétrica, servindo de base para uma implementação detalhada nas seções posteriores. Essa implementação objetiva medir o consumo de energia de equipamentos, sejam residenciais ou industriais, efetuando leituras das grandezas elétricas básicas, como tensão e corrente. A partir dos valores dessas grandezas, é possível derivar medidas que servem como indicadores e auxiliam a melhor gestão energética.

Essa implementação será apresentada em detalhes nas próximas seções, juntamente com resultados de dois estudos de caso realizados (um residencial e outro em contexto industrial). No entanto, faz-se necessário o entendimento prévio de alguns conceitos inerentes ao tema. Assim, esta seção apresenta alguns conceitos básicos relacionados à área de gerenciamento de energia elétrica, com o objetivo de facilitar a compreensão do trabalho realizado neste curso de mestrado.

5.1.1 Consumo de Energia Elétrica

Inicialmente, é importante relembrar o conceito de consumo energético, que está presente em nosso dia-a-dia. Quando um equipamento elétrico fica em funcionamento por um determinado intervalo de tempo, é possível calcular a energia elétrica consumida por ele (geralmente, medido em KWh). Para isso

basta multiplicar a potência do equipamento (P, medida em watts –W, ou KW) pelo tempo de funcionamento (geralmente, medido em horas).

$$\text{Consumo Energia Eletrica} = \text{Potência}(W) * \Delta t$$

A potência elétrica indica a quantidade de energia que é transformada em outro tipo de energia em uma unidade de tempo (ALEXANDER,2013). No nosso caso, por exemplo, estamos falando da transformação de energia elétrica em energia luminosa (no caso das lâmpadas), em energia mecânica pelo acionamento de equipamentos elétricos, ou até em energia térmica (ferro de passar roupa).

Para calcular a potência de um equipamento, basta multiplicar a tensão submetida ao equipamento (em volts - V) pela intensidade de corrente que percorre o equipamento (em A). Logo,

$$\text{Potência}(W) = \text{Corrente}(A) * \text{Tensão}(V)$$

A soma da potência de todos os equipamentos de um ambiente sendo monitorado é denominada potência instalada (Pinst), ou carga instalada.

Os sensores utilizados na solução implementada efetuam a leitura da tensão e da corrente do equipamento sendo monitorado, e a partir dessas grandezas, calculam o consumo de energia elétrica. Dessa forma, executam uma fusão DAE-FEO, onde dados brutos oriundos dos sensores são transformados em características de fácil interpretação (seção 4.2.1).

5.1.2 Demanda de Energia Elétrica

Outra grandeza relacionada à conta de energia é a demanda de potência, e tem como unidades de medida mais comuns o KW (quilowatt) e o MW (megawatt).

Podemos definir variações dessa grandeza.

- Demanda máxima (Dmáx): A soma das potências dos equipamentos em uso normal, quando todos são ligados simultaneamente;

- Demanda média: A média do consumo de energia durante um dado período de tempo (potência utilizada). Note que o consumo é acumulado com o passar do tempo, diferentemente da demanda.

A demanda média pode ser usada na predição do consumo esperado de cada unidade (residência, indústria, ponto comercial), auxiliando o fornecedor de energia a calcular a demanda de potência global da região.

Em residências e pequenas empresas, caracterizadas como consumidores de baixa tensão (grupo B), apenas o consumo é tarifado. Já para os consumidores de média e alta tensão (denominado grupo A), além do consumo, uma parcela da demanda contratada também é tarifada.

A demanda contratada corresponde ao valor estimado de demanda de potência, informado pelos consumidores do grupo A à distribuidora. Essa informação é muito importante, pois serve como um contrato para que o sistema elétrico esteja preparado para atender a toda carga nacional. Sendo assim, as empresas desse grupo devem informar a carga total (de seus equipamentos) à distribuidora, para que ela possa garantir o fornecimento, sem com isso sobrecarregar o sistema. Caso a empresa ultrapasse o consumo acordado, deverá pagar multas sobre o excedente, como num plano pré-pago. Vale ressaltar que esse valor é bem mais caro do que a tarifa contratada inicialmente.

Nossa implementação calcula a demanda média a partir do monitoramento dos valores do consumo. Contudo, o valor da demanda máxima só pode ser obtido a partir de um parâmetro de entrada informado pelo usuário.

5.1.3 Outros Indicadores

A partir do consumo e da demanda de energia, é possível obter outros indicadores que são importantes para a gestão energética em indústrias e empresas. Destacamos aqui o **fator de potência**, **fator de carga** e **fator de demanda**, descritos a seguir.

Fator de Potência

Esse indicador é muito importante para consumidores não residenciais, sobretudo os industriais. Para esses consumidores, valores de fator de potência

abaixo de 0,92 indicam consumo de potência reativa (definida abaixo) além do permitido, sendo passível de multa pela concessionária na fatura de energia.

Para o entendimento do fator de potência, é importante definirmos antes as grandezas de potência aparente, potência ativa e potência reativa.

- Potência aparente pode ser considerada como toda a carga entregue à unidade consumidora. Essa grandeza é conhecida como potência instantânea, sendo calculada pela multiplicação da tensão pela corrente (medida em kVA - quilo volt-ampere).
- Potência ativa é a energia utilizada na conversão de energia elétrica em térmica, mecânica etc. Ou seja, a energia consumida para realizar movimento, calor, emissão de luz, etc., sendo medida em kW.
- Potência reativa, ao contrário da ativa, não é transformada, e sim apenas consumida pelos equipamentos a fim de formar campos eletromagnéticos necessários para seu funcionamento (por exemplo, serve para induzir as bobinas dos equipamentos elétricos). Sua unidade de medida é o kVAr (quilo volt-ampere reativo).

Por fim, fator de potência pode ser calculado como a razão entre as potências ativa e aparente, e indica quanto da potência elétrica consumida está sendo efetivamente transformada em trabalho. Essa grandeza varia entre 0 e 1, e valores mais próximos de 1 indicam uma maior eficiência do equipamento.

$$\text{Fator de Potência} = \frac{\text{Potência Ativa}[kW]}{\text{Potência Aparente}[kVA]}$$

A implementação realizada neste trabalho é capaz de calcular o fator de potência a partir das leituras dos sensores para os equipamentos monitorados (ver Capítulo 5, seção 5.3).

Fator de Demanda

Para evitar o superdimensionamento da instalação elétrica, é importante considerar que nem todos os equipamentos estarão ligados simultaneamente. Para contornar isso, deve-se considerar no projeto os fatores de demanda e de simultaneidade.

O fator de demanda é calculado a partir de grandezas já definidas: potência instalada (P_{inst}), que é o total de potência de todos os equipamentos do ambiente é denominado; e demanda máxima ($D_{máx}$), que é a soma das potências dos equipamentos em uso normal (quando todos são ligados simultaneamente). O fator de demanda é a razão entre a demanda máxima e a potência instalada.

$$Fator\ de\ Demanda = \frac{D_{máx}[kW]}{P_{inst}[kW]}$$

Assim como o fator de potência, o fator de demanda varia de 0 a 1. Valores próximos de 1 indicam que a demanda máxima está bem próxima da potência instalada, revelando que a instalação elétrica pode ficar sobrecarregada.

Esse indicador é de grande relevância para o dimensionamento dos equipamentos elétricos que irão alimentar uma instalação, considerando a análise de simultaneidade dos equipamentos. Conhecendo a demanda máxima, é possível a contratação mais assertiva da demanda necessária (i.e., a potência demandada) junto à distribuidora, sem correr o risco de superdimensionar ou subestimar essa demanda.

Nossa implementação atual não tem dados suficientes para realizar o cálculo do fator de demanda (ver Capítulo 5). Para tanto, seria necessário que o usuário informe a potência instalada. Assim, esse cálculo, apesar de simples, fica indicado como trabalho futuro.

Fator de Carga

Esse indicador mede a uniformidade com que a energia está sendo consumida. É calculado pela razão entre o consumo total de energia elétrica e a demanda máxima medida dentro de um intervalo de tempo. Esse valor varia entre 0 e 1, e valores mais próximos de 1 são mais desejados, uma vez que indicam um consumo de energia mais uniforme.

$$Fator\ de\ Carga = \frac{Consumo\ Total[kWh]}{Demanda\ Máxima\ Medida[kW] * Tempo [h]}$$

Esse indicador, quando bem gerenciado, leva a um melhor aproveitamento da instalação elétrica, além de viabilizar a redução no preço médio da energia (R\$ / kWh). Podemos concluir que o preço médio da energia é inversamente proporcional ao fator de carga, ou seja, quanto maior o fator de carga, menor o preço pago pelo kWh (ver equação abaixo).

$$\text{Preço Médio [R\$ / kWh]} = \frac{\text{Valor da Fatura [R\$]}}{\text{Demanda Máxima [kW]} * \text{Tempo[h]} * F.Carga}$$

Com a gestão desse indicador, é possível analisar os horários de maior consumo de energia e buscar um uso mais uniforme, garantindo um melhor aproveitamento da instalação elétrica e da demanda contratada.

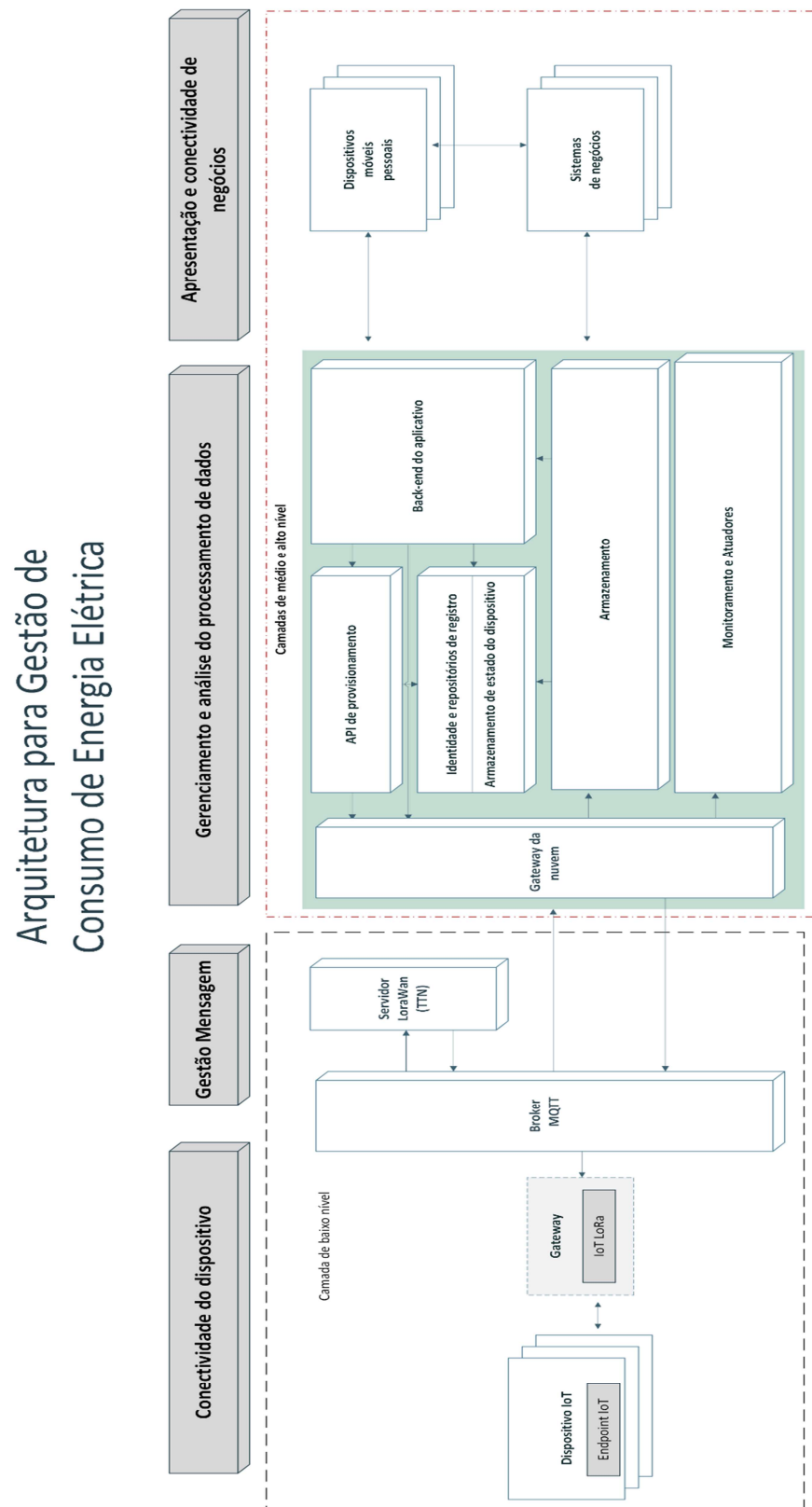
Nossa solução ainda não implementou o cálculo desse indicador, uma vez que não estamos tratando da demanda no momento. Porém, trata-se de um cálculo simples, e de grande relevância para consumidores industriais. Assim, sua inclusão na nossa implementação fica indicada como trabalhos futuros.

5.2 UMA RSSF PARA MONITORAMENTO DE ENERGIA ELÉTRICA

Esta seção apresenta a aplicação desenvolvida para gestão de consumo de energia elétrica, uma RSSF implementada a partir da AMnFD (Capítulo 4). A figura 18 traz uma visão geral da arquitetura da RSSF, com seus componentes e o fluxo de dados. As seções a seguir detalham a infraestrutura de comunicação, o *hardware* (equipamentos utilizados) e *software* desenvolvido.

Como já mencionado, a solução proposta foi testada em dois estudos de caso (seção 5.2), com o objetivo de monitorar o consumo de energia a partir das leituras das grandezas de corrente e tensão. Vale salientar que nos dois estudos de caso foram utilizados a mesma infraestrutura de rede e os mesmos sensores, apenas com pequenas alterações no nó sensor para oferecer leituras tanto em redes monofásicas quanto trifásicas.

Figura 18- Arquitetura da RSSF para gestão de consumo de energia elétrica



Fonte: Autor

5.2.1 Infraestrutura de comunicação

A infraestrutura de comunicação implementada está dividida em quatro áreas distintas (figura 18), com o objetivo de garantir a coleta e transmissão dos dados brutos dos sensores até o seu armazenamento, possibilitando futuras análises e consultas pelos usuários.

5.2.1.1 Área de conectividade dos dispositivos

É nessa área que se encontram os nós sensores, ou *endpoints*, que utilizam a tecnologia de modulação de rádio frequência LoRa (do inglês, *Long Range*)⁴ para se comunicarem. A tecnologia utilizada permite uma comunicação a longas distâncias e baixa potência. Essa tecnologia foi criada pela empresa Semtech, visando uma padronização das LPWANS (do inglês *low-power wide area networks*).

Os nós sensores foram dispostos numa topologia estrela, onde todos se comunicam com o dispositivo *gateway* específico, similar ao que ocorre numa rede *Wi-Fi*. O *gateway*, que deve estar conectado com a internet (seja por *Wi-Fi* ou telefonia móvel), envia as informações dos nós sensores para o servidor de mensageria em nuvem numa frequência de tempo de minutos.

Para a comunicação do *gateway* com o servidor de mensageria *broker*, foi utilizado o protocolo de comunicação LoRaWAN⁵, que tem a capacidade de suportar milhões de mensagens, a depender da quantidade de *gateways* da rede. Um único *gateway* de 8 canais é capaz de suportar cerca de 1,5 milhão de mensagens num período de 24 horas (SEMTECH,2019).

Para fins deste trabalho, foi utilizado o servidor de mensageria denominado *broker TTN* (do inglês *The Things Network*)⁶, que tem o papel de intermediário na comunicação entre o *gateway* e os servidores de aplicação. O *gateway* repassa as mensagens dos sensores para a área de Gestão de mensagem.

5.2.1.2 Área Gestão de Mensagem

⁴ <https://www.semtech.com/products/wireless-rf>

⁵ <https://lora-alliance.org/about-lorawan/>

⁶ <https://www.thethingsnetwork.org/>

É nesta área onde está localizado o servidor de mensageria LoRaWAN e o *broker mqtt*. A comunicação entre o *broker mqtt*, servidor LoRaWAN e os clientes é feita utilizando um protocolo de comunicação denominado MQTT (*Message Queue Telemetry Transport*) (OASIS,2019). Esse protocolo foi inicialmente desenvolvido para ser proprietário da IBM, mas em 2010 foi liberado para uso gratuito e em 2014 tornou-se um padrão OASIS (*Organization for the Advancement of Structured Information Standard*), um consórcio global sem fins lucrativos que conduz o desenvolvimento e adoção de padrões para *e-business* e *webservices* e logo virou padrão das redes IoT. Atualmente, o MQTT encontra-se na versão 5 (OASIS,2019).

Ainda segundo (OASIS,2019),, algumas características do protocolo MQTT são:

- É um protocolo de transporte de mensagens Cliente/Servidor no conceito Publicação/Assinatura (*Publish/Subscribe*);
- É leve, de código aberto e fácil implementação e administração;
- Ideal para cenários com recursos restritos, onde são exigidos pacotes de dados e/ou largura de banda mínimos;
- É um protocolo executado sobre TCP/IP ou sobre outros protocolos de rede que forneçam conexões bidirecionais ordenadas sem perdas;
- Endereçamento da mensagem para nenhum, um ou vários clientes;
- Nível de serviço adequado à relevância de cada mensagem;
- Alta Segurança, com criptografia da conexão e dos dados.

Esse protocolo trabalha com um modelo de publicação e assinatura (do inglês, *publish e subscriber*), de modo que um cliente pode ser qualquer entidade que consiga interagir com o *broker* e receber mensagens. O cliente pode assinar qualquer tópico de mensagem do *broker mqtt*. As mensagens transitadas são formadas basicamente do tópico e da carga útil, que pode ser texto, binário, JSON ou XML, dentre outros.

O fluxo de funcionamento dessa área inicia quando o dispositivo cria a mensagem no formato do protocolo LoRaWAN e a envia ao *gateway* usando a modularização LoRa. Ao receber a mensagem, o *gateway* a encaminha para o *broker mqtt*, que redireciona essa mensagem para o servidor LoRaWAN por meio de um tópico. O servidor LoRaWAN, por sua vez, descriptografa as

informações e as envia para o *broker mqtt*, que novamente as redireciona para o cliente, que irá tratar e analisar as informações.

Na arquitetura proposta, as mensagens do *broker mqtt* são encaminhadas para os servidores de aplicações, que também são clientes do *broker* e estão identificados na figura 18 na área denominada área de “Gerenciamento e análise do processamento de dados”.

5.2.1.3 Área Gerenciamento e análise de processamento de dados

Essa área é responsável por recepcionar todos os dados oriundos dos sensores, encaminhados através do *broker mqtt*. Os servidores de aplicações, que são *softwares* executados em um servidor responsável pelo processamento seguro da aplicação, executam instruções a partir de serviços de forma automática e muitas vezes em paralelo.

No contexto deste trabalho, a comunicação nessa camada foi implementada via uma API (*Application Programming Interface*), com um conjunto de rotinas e padrões de programação para acesso a um aplicativo de *software* ou alguma plataforma *Web*. Essa API é responsável por recepcionar as mensagens do *broker*, num formato JSON, interpretá-las e em seguida gravar num BD de forma estruturada. Também pode haver o caminho inverso, ou seja, essa API pode recepcionar mensagens que devam ser encaminhadas ao *broker* para serem entregues a clientes que assinam um determinado tópico.

5.2.1.4 Área Apresentação e conectividade de negócios

É nessa área onde o usuário interage com a aplicação, podendo incluir e alterar regras, podendo inclusive receber alertas, consultar o consumo do equipamento, e analisar os indicadores de gestão energética. Essa área será melhor detalhada nas próximas seções.

A seguir, serão descritos os componentes da RSSF e o funcionamento dos equipamentos (seção 5.1.2). Por fim, veremos detalhes dos *softwares* implementados na aplicação (seção 5.1.3).

5.2.2 Infraestrutura de *Hardware*

Os procedimentos para obtenção dos dados estatísticos da rede elétrica têm origem nas entradas dos sensores não invasivos de tensão e corrente. Nosso nó sensor (*endpoint*) é composto por uma unidade de sensoriamento (sensor de corrente com transformador analógico-digital), uma unidade de processamento, a unidade de comunicação (emissor e receptor) e a unidade de alimentação (Capítulo 2). O nó sensor construído neste trabalho teve sua alimentação a partir da rede elétrica, e não por meio de baterias.

O mesmo tipo de nó sensor foi utilizado nas duas aplicações (geladeira e extrusora), sofrendo pequenas alterações no circuito, a fim de fazer a conversão das leituras da rede elétrica. Isso foi necessário para adequar a operação em redes monofásicas (geladeira) e trifásicas (extrusora).

Para a leitura monofásica, foi utilizado um único sensor/transformador não invasivo, modelo SCT-013 de 100A com saída de 50mA (figura 19a). Já para a solução utilizada na extrusora, foi necessária uma alteração para comportar leituras trifásicas, utilizando-se então três transformadores de corrente de até 600A com saída de 5A (figura 19b).

Figura 19- Sensores utilizados nos casos de uso



Figura 19a. Sensor de Corrente Não Invasivo SCT-013



Figura 19b. Transformador de corrente bipartido com abertura central (janela)

Fonte: Autor

O processador utilizado no nó sensor foi o microcontrolador ESP8266⁷, para processamento dos sinais e comunicação sem fio. Para prover a

⁷ <https://pt.wikipedia.org/wiki/ESP8266>

comunicação com o *gateway*, foi instalado um módulo LoRa 915MHz RF *wireless*. Além disso, com vistas a permitir acesso remoto, foi inserido um sistema de comunicação sem fio (*Wi-Fi do inglês Wireless Fidelity*) de curto alcance, operando em frequência 2.4 GHz, utilizando o protocolo de comunicação IEEE 802.11b/g/n. Por sua vez, o *gateway*, que tem acesso *Wi-Fi*, comunica-se com *broker mqtt*.

O gateway utilizado foi o LoRaWAN da marca DRAGINO , modelo LPS8 de código aberto. Ele permite conectar a rede sem fio LoRa a uma rede IP via *Wi-Fi, Ethernet*.

5.2.3 Software desenvolvido

Por fim, veremos que as instruções de *software* estão incluídas basicamente em duas áreas (figura 18): *Área de conectividade dos dispositivos e Gerenciamento e análise do processamento de dados*.

O *firmware*, isto é, o código que está armazenado no microcontrolador ESP8266 do nó sensor, foi desenvolvido na linguagem C. Suas principais funções são a interpretação dos dados brutos obtidos a partir dos sensores, e a implementação de filtro simples para detecção de dados espúrios, ou *outliers*.

O pseudocódigo abaixo (Algoritmo 1) apresenta o algoritmo implantado na camada física da arquitetura de fusão de dados, que recebe como entrada a leitura dos dados brutos dos sensores de tensão e corrente, descartando os valores zerados e encaminhando os demais para a camada de médio nível.

Algoritmo 1: Leitura dos dados brutos dos sensores na camada física

Entrada: Dados brutos dos sensores de tensão e corrente

Saída: Dados para a camada de médio nível

início

Coletar Valor da Tensão (V)

Coletar Valor da Corrente (A)

Se Valor da Tensão ≤ 0 OU Valor da Corrente ≤ 0

Descartar valor

Senão

Encaminhar para camada de médio nível

FimSe

Fim

O Algoritmo 2 apresenta o pseudocódigo que foi implementado na camada de médio nível. Basicamente, ele efetua um filtro de detecção de *outliers* simples. O código recebe os dados brutos da camada física, armazena a média dos valores recebidos e compara a leitura recebida com a média calculada. Se a leitura diferir da média calculada em Z% (Z é um número real parametrizável), para mais ou menos, a leitura é marcada como *outlier*.

A potência é calculada em watts (W), sendo depois convertida para a unidade quilowatt (kW). A partir daí, obtém-se o consumo, que é encaminhado para a camada de alto nível. Posteriormente, o valor da nova média é atualizado.

Algoritmo 2: Leitura dos dados brutos dos sensores na camada médio nível

Entrada: Dados brutos dos sensores de tensão e corrente da camada física

Saída: Dados para a camada de alto nível

início

Coletar dados tensão (V)

Coletar dados corrente (A)

Se valor variar em Z% ou -Z% da média

 Marcar leitura como *outlier*

FimSe

Calcular Potência = tensão * corrente

Calcular Consumo em kW/h

Encaminhar Consumo à camada de Alto Nível

Atualizar Média

fim

Por fim, na área de *Gerenciamento e análise do processamento de dados*, foi implementado um *software* na linguagem .NET. Essa área contém a implementação das camadas de médio e alto nível da arquitetura de fusão de dados. A área utiliza os dados oriundos das leituras dos sensores, que são atualizados frequentemente, a partir de instruções de *software* que executam como serviços num servidor de aplicação, e atualizam um banco de dados MongoDB⁸.

⁸ <https://www.mongodb.com>

5.3 ESTUDOS DE CASO

Como já mencionado antes, a AMnFD foi implementada a partir de uma rede de sensores sem fio, tendo sido testada em dois estudos de caso, a fim de obter o consumo de energia a partir das leituras das grandezas de corrente e tensão. Nos dois estudos de caso, foram utilizados a mesma infraestrutura de rede e os mesmos sensores, apenas com pequenas alterações no nó sensor para proporcionar leituras tanto em redes monofásicas quanto trifásicas.

A principal característica desses estudos é a coleta de dados e análise, com vistas a tomar ações corretivas e/ou preventivas a partir do monitoramento dos equipamentos. Buscou-se também mostrar que, a partir da implementação da AMnFD proposta, é possível garantir uma maior acurácia dos dados, permitindo a identificação de *outliers* considerados como espúrios ou eventos, além de auxiliar na tomada de decisões.

5.3.1 Contexto Residencial (Geladeira)

Esse estudo de caso foi realizado em um contexto residencial, através do monitoramento de uma geladeira Electrolux Top Freezer 2 Portas DF52. De acordo com o manual do produto, seu consumo esperado é de 58 kW/h. É importante ressaltar que o equipamento já tem aproximadamente 10 anos e já passou por algumas manutenções corretivas, o que pode causar uma diferença entre o consumo esperado e o efetivamente ocorrido.

Aqui, foi implantado o nó sensor (*endpoint*) descrito na seção 5.1.2 para leituras monofásicas, isto é, o sensor de corrente não invasivo modelo SCT013 (figura 19a).

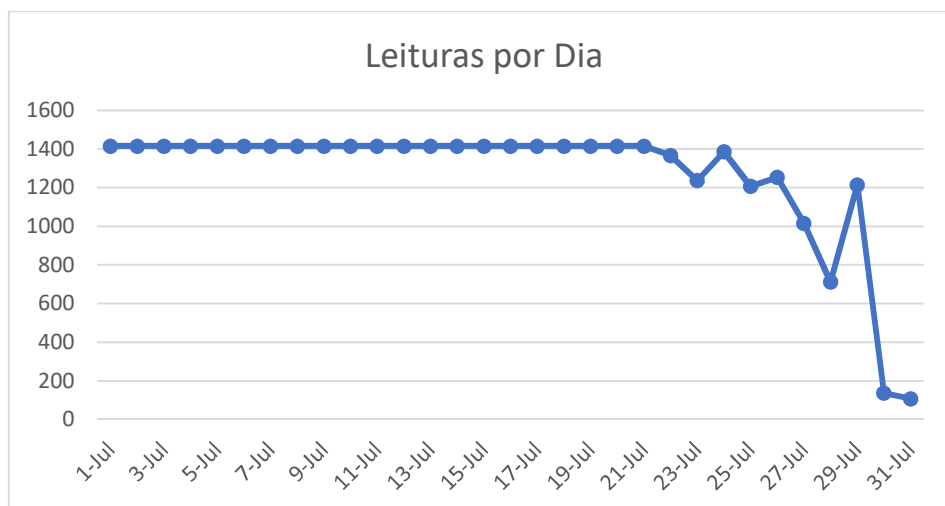
5.3.1.1 Coleta de dados

A coleta dos dados brutos foi realizada de 01 de junho a 30 de julho de 2020 (2 meses). Esses dados foram obtidos utilizando a RSSF montada na residência do autor, onde o *endpoint* conectado à geladeira coletou os dados do sensor com uma periodicidade de minuto a minuto, totalizando nesse período 42.480 leituras.

Do dia 01 de junho a 20 de julho 2020, a quantidade de leituras diárias foi igual e constante, havendo alteração a partir do dia 21 de julho de 2020, quando

as leituras passaram a sofrer variações (Figura 20). As variações aqui citadas, ocorreram em detrimento a testes (troca de equipamentos, desligamento da conectividade) que foram feitos na rede.

Figura 20- Quantidade de leituras por dia.

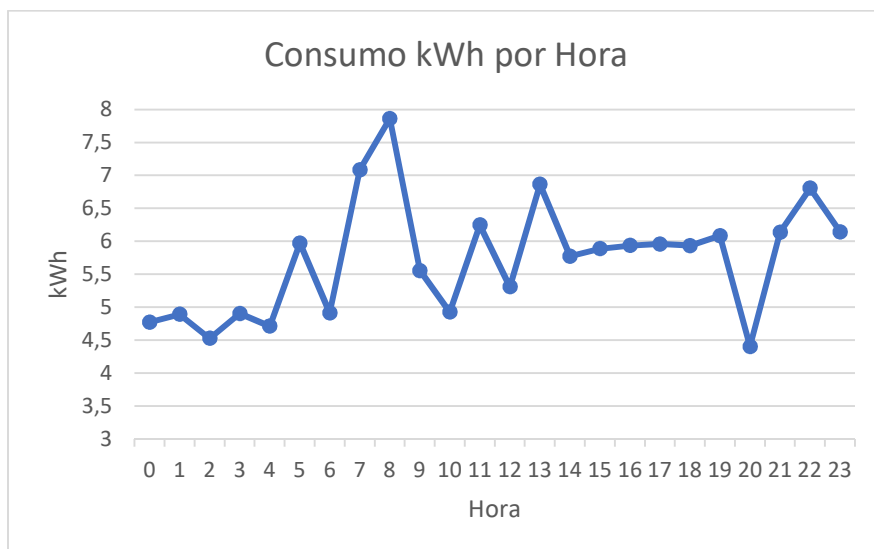


Fonte: Autor

5.3.1.2 Interpretação dos Dados Obtidos

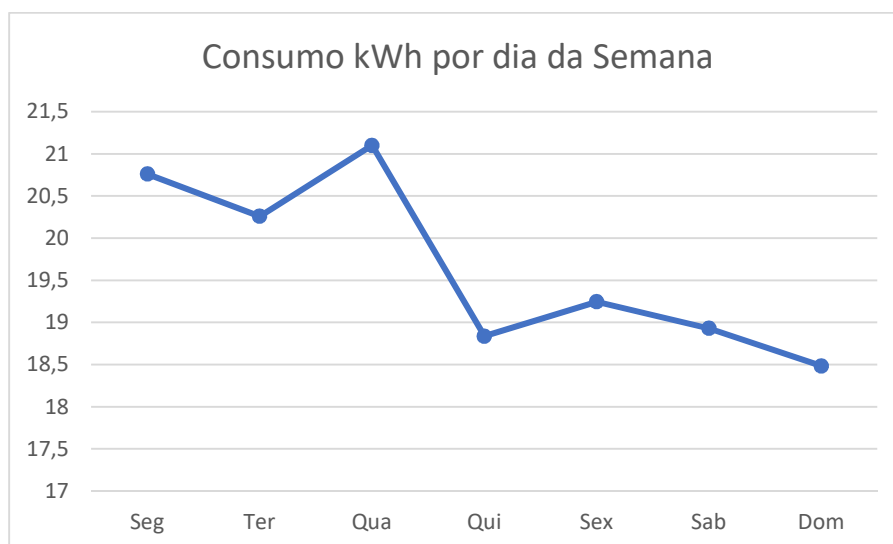
Os dados coletados dos sensores, com as grandezas de tensão e corrente, permitiram calcular o consumo da geladeira, e também efetuar algumas análises como, por exemplo, o horário do dia em que houve maior consumo. Conforme mostrado na Figura 21, a maior média de consumo ocorreu entre 07 e 08 horas da manhã. Também é possível analisar na Figura 22 que, no final de semana (sábado e domingo), houve o menor consumo do período monitorado.

Figura 21- Quantidade de leituras por hora.



Fonte: Autor

Figura 22- Quantidade de leituras por dia da semana.



Fonte: Autor

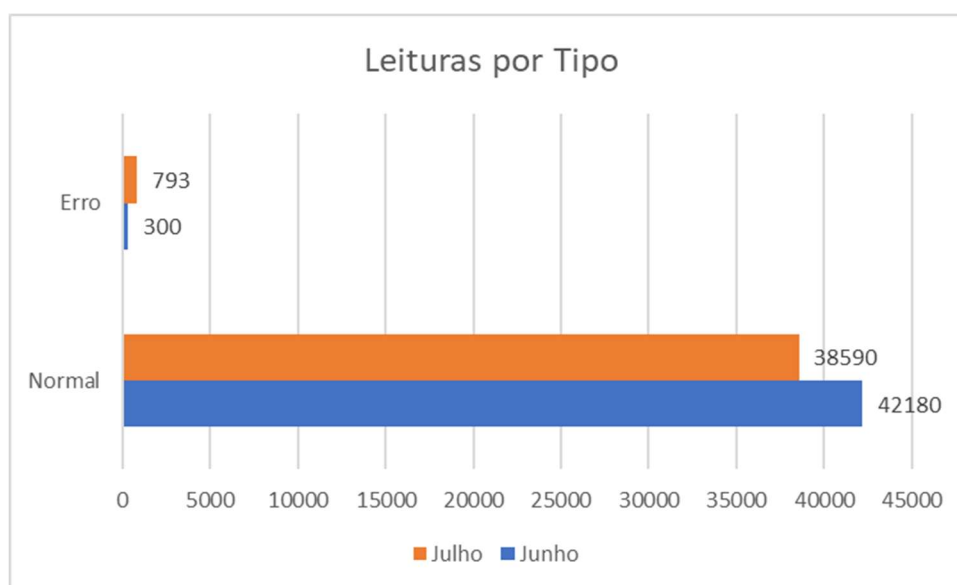
Durante a semana, houve um maior consumo de energia elétrica, pois todos os moradores estavam em casa, ocasionando uma maior frequência de manuseio (abrir e fechar) da geladeira para retirar os mantimentos, a fim de efetuar refeições. Isso ocorreu com menor frequência no final de semana, quando diminuía a quantidade de pessoas casa e a frequência de cozinhar e precisar manusear a geladeira foi muito menor.

Detecção de Outliers

Durante o período de monitoramento, foram analisadas exatas 80.770 leituras, que deveriam ter sido enviadas em intervalos de minuto a minuto. No entanto, foram observadas 1.093 situações em que o intervalo entre as leituras excedeu 1 minuto. Esses valores são indicados na Figura 23 como “Erro”.

Na tabela 3 é possível analisar que, no dia 30/06/2020, todas as leituras foram realizadas em intervalos regulares de 1 minuto, excetuando a leitura das 04:00 horas, cuja próxima leitura ocorreu 5 minutos depois, ocasionando um lapso de leitura.

Figura 23- Leituras por classificação



Fonte: Autor

Tabela 3. Leituras por intervalo de tempo

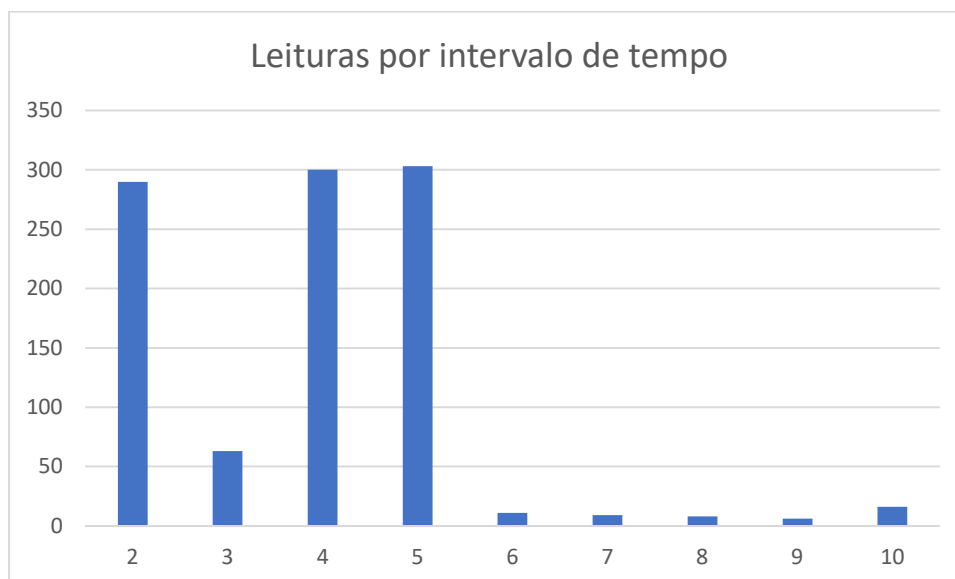
Data Hora	Leitura Kw/h	Diferença da última leitura (minutos)
30/6/2020 4:07 PM	0.00070616	1
30/6/2020 4:06 PM	0.00077503	1
30/6/2020 4:05 PM	0.00078667	1
30/6/2020 4:00 PM	0.0008982	5
30/6/2020 3:59 PM	0.00087785	1
30/6/2020 3:58 PM	0.00083905	1
30/6/2020 3:57 PM	0.000873	1
30/6/2020 3:56 PM	0.00113684	1

Também foi observado que os lapsos entre as leituras são variados, como visto na Tabela 4. Cerca de 1.006 leituras tiveram diferença no envio dos dados em intervalos de até 10 minutos. Ainda, 290 leituras sofreram diferença de envio em 2 minutos. Também foi observado que cerca de 87% se concentram com diferenças de intervalo de até 5 minutos.

Tabela 4 . Leituras por intervalo de tempo

Minutos	Leituras
<=10	1006
2	290
3	63
4	300
5	303
6	11
7	9
8	8
9	6
10	16
>11	87
Total Geral	1093

Figura 24- Gráfico de Leituras por Intervalo de tempo.



Fonte: Autor

Uma análise mais aprofundada identificou que os lapsos nas leituras, na verdade, são ausências de leituras, ou seja, o Algoritmo 1 (seção 5.1.3), de identificação de *outlier* no *firmware*, acusou leitura errada menor ou igual a zero

para as grandezas de corrente e/ou tensão, desconsiderando as mesmas. O algoritmo foi implementado na camada de baixo nível da AMnFD.

Esse tipo de *outlier* pode ter sido causado por algum problema/defeito do sensor ou na ausência de conectividade da rede. No entanto, todos os lapsos/erros ocorridos representam apenas 1,34% do total de leituras. Assim, não houve prejuízo real para a acurácia no cálculo de consumo da geladeira, nem dos outros indicadores de grandezas elétricas.

Uma possível solução para essas ausências de dados poderia ser o preenchimento de leituras com o valor da média das leituras do dia. Esse valor poderia ter sido aplicado para suprir as ausências. Note que a implementação dessa solução só poderia ser executada na camada de médio nível da AMnFD, pois nela constam todas as leituras.

No entanto, essa lógica não se aplica para casos em que a interrupção no envio de leituras perdurasse tempo suficiente para comprometer a média diária. Nesse caso, seria necessário aplicar um algoritmo mais adequado, que fosse capaz de calcular uma média ou valor mais próximo da realidade.

Tendo em vista a complexidade na detecção de *outliers*, o tema tem sido amplamente pesquisado em várias disciplinas, tais como estatística, mineração de dados, inclinação de máquina, teoria da informação e decomposição espectral dentre outras. (FREITAS,2019)

5.3.2 Contexto Industrial (Extrusora)

O segundo estudo de caso foi ambientado em uma pequena fábrica de filmes plásticos da região Metropolitana do Recife – PE. O setor de extrusão é o início do processo produtivo de filmes plásticos, onde as máquinas extrusoras recebem a matéria prima (grânulos de plásticos) através de um cilindro aquecido por resistências elétricas. Esse cilindro contém uma rosca transportadora onde ocorre a fusão dos grânulos, formando uma massa plastificada (figura 25).

A seguir, o material é transportado e passa pela Matriz aquecida, saindo pela abertura na forma de filme, como um balão. Na matriz há uma saída de ar com a finalidade de insuflar o balão. A quantidade de ar empregada irá definir a largura do balão e, conseqüentemente, a espessura do filme. A espessura do filme também é definida pela velocidade dos puxadores. Uma vez pronto, o filme é bobinado em forma de cilindro. Os filmes produzidos por esse processo são

usados em diversos tipos de embalagens. A Figura 25 apresenta a extrusora utilizada nesse estudo de caso.

Figura 25- Visão geral de uma extrusora



Fonte: Autor

A extrusora empregada nesse caso tem em suas referências, segundo o fabricante, um levantamento de carga total de 120KW, que consiste na potência dos motores e potência das resistências.

O *endpoint* utilizado neste estudo de caso tem apenas uma diferença em relação ao caso anterior. Aqui foram acrescentados 3 transformadores de corrente (figura 19b) para se adequar à rede, com cargas até 600A.

5.3.2.1 Coleta dos dados

Os dados utilizados neste estudo de caso foram coletados entre os dias 06 e 19 de dezembro de 2020. Nesse período, foram feitas leituras na subestação do equipamento extrusora a partir do *endpoint* instalado no

equipamento, com uma periodicidade de minuto a minuto, totalizando nesse período 17.302 leituras.

Do dia 06 de dezembro a 19 de dezembro 2020, a quantidade de leituras diárias foi igual e constante, havendo alteração a partir do dia 11 de dezembro de 2020, quando as leituras passaram a sofrer variações (Figura 26).

Figura 26- Quantidade de leituras por dia



Fonte: Autor

Assim como ocorreu no estudo de caso anterior, aqui também foram coletados dados com valores fora do esperado, indicando possibilidade de *outliers*. Porém, neste caso, os dados espúrios não se devem ao mau funcionamento do nó sensor ou alguma interrupção na rede. Aqui, ocorreu uma interrupção de operação inesperada da extrusora. Veremos mais detalhes sobre os dados coletados na seção a seguir.

5.3.2.2 Interpretação dos Dados Obtidos

A partir da análise das leituras realizadas, é possível fazer a extração de algumas informações sobre o funcionamento do equipamento. A tabela 5 mostra alguns dos valores obtidos/calculados a partir dos dados coletados. Essas são informações importantes para compreender o funcionamento do equipamento sendo monitorado, como a continuidade do consumo, a distância entre picos de demanda e o aproveitamento da energia consumida.

Tabela 5 Dados de funcionamento do equipamento

Data	Demanda Máx. KW	Demanda Méd. KW	Potência Instant. kW	Fator de Demanda	Fator de Carga	Média do FP
06/12/20	74,88	56,13	120	0,62	0,74	0,92
07/12/20	66,41	50,75	120	0,55	0,76	0,91
08/12/20	87,55	61,9	120	0,72	0,70	0,92
09/12/20	85,95	55,88	120	0,71	0,65	0,92
10/12/20	78,58	56,75	120	0,65	0,72	0,91
11/12/20	23,62	5,95	120	0,19	0,25	0,89
12/12/20						
13/12/20	2,66	1,96	120	0,02	0,73	0,88
14/12/20	2,68	1,96	120	0,02	0,73	0,88
15/12/20	70,11	28,15	120	0,58	0,4	0,91
16/12/20	76,41	57,12	120	0,63	0,74	0,92
17/12/20	72,02	44,12	120	0,6	0,61	0,9
18/12/20	87,78	68,9	120	0,73	0,78	0,92
19/12/20	84,87	57,9	120	0,7	0,68	0,91

Detecção de *Outliers*

Como já mencionado, durante a coleta de dados, foi observado que a máquina parou de funcionar por um breve período (início 18:14 do dia 10 e retorno efetivo das operações no dia 15/12/20 as 08:38). Esse evento foi capturado pela aplicação na camada de gerenciamento, que acusou o não recebimento de leituras por vários minutos, enviando um alerta ao responsável cadastrado no sistema, que confirmou a paralisação do equipamento.

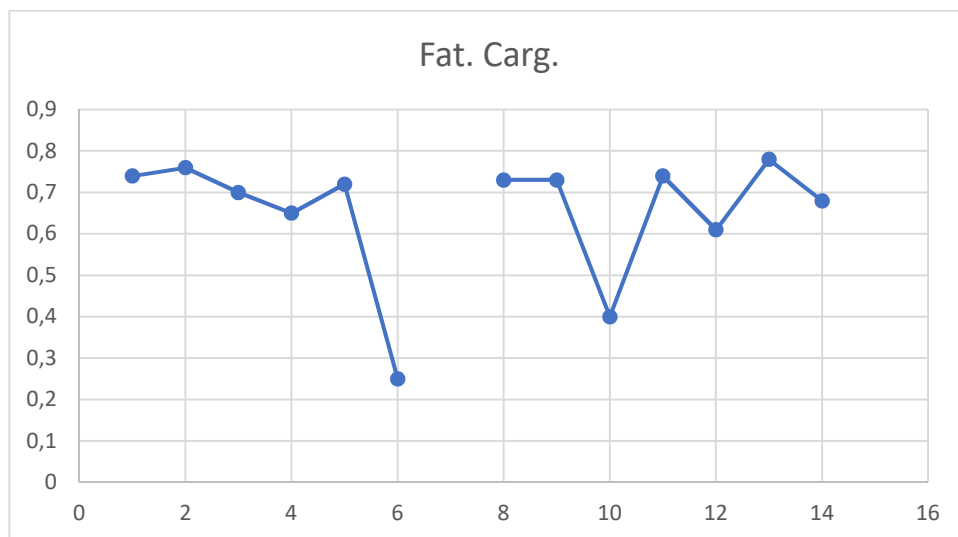
Apesar de ter capturado o evento (*outlier*), nossa aplicação não foi capaz de determinar a causa da interrupção do funcionamento da máquina. Essa extensão fica indicada como trabalhos futuros (Capítulo 6). Contudo, o alerta gerado pela aplicação foi de grande valia, agilizando a tomada de ação no sentido de corrigir o problema.

A partir do alerta enviado, o gerente de manutenção realizou as ações de correção. Porém, aparentemente, a manutenção não foi feita de forma correta. A amostra de dados coletada em seguida revelou que o equipamento sofreu alguns danos após essa parada brusca. O fator de carga e o fator de potência se mostraram instáveis após o religamento do equipamento (como ilustrado nas Figuras 27 e 28)

Os valores dos indicadores também mostram que o equipamento não está trabalhando com seu desempenho total. O fator de Demanda e o fator de Carga

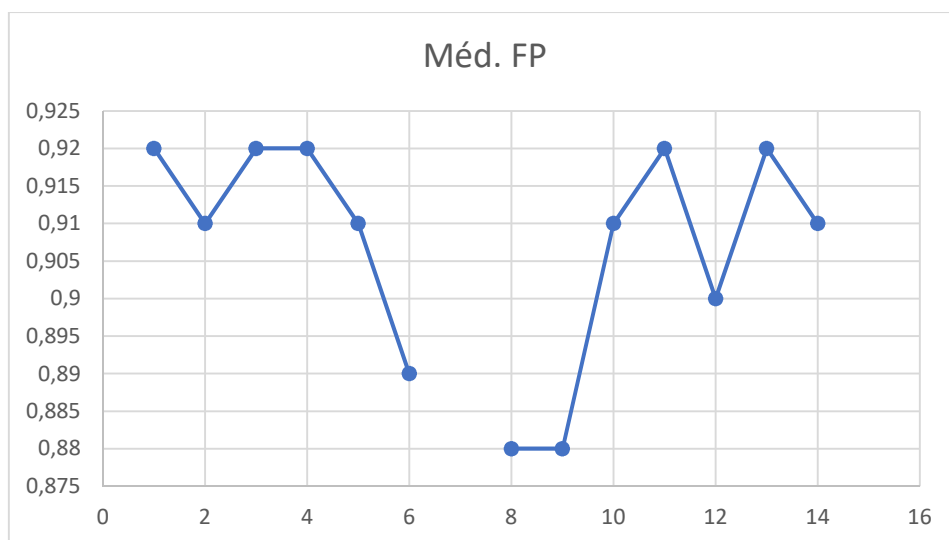
estão com valores aceitáveis (superior ou igual 0.8), porém o ideal seria ter números mais próximos do valor 1. Esses valores indicam que o equipamento poderia utilizar uma carga maior de corrente para melhor aproveitamento da energia consumida.

Figura 27- Gráfico de fator de cargas



Fonte: Autor

Figura 28- Gráfico do Fator de potência



Fonte: Autor

Como já explicado anteriormente, o fator de potência é uma das grandezas mais importantes para a indústria (Capítulo 4, seção 4.3). Como visto, esse indicador sofreu uma drástica queda durante a parada do equipamento, mostrando um desequilíbrio no consumo durante a essa ocorrência. Muito

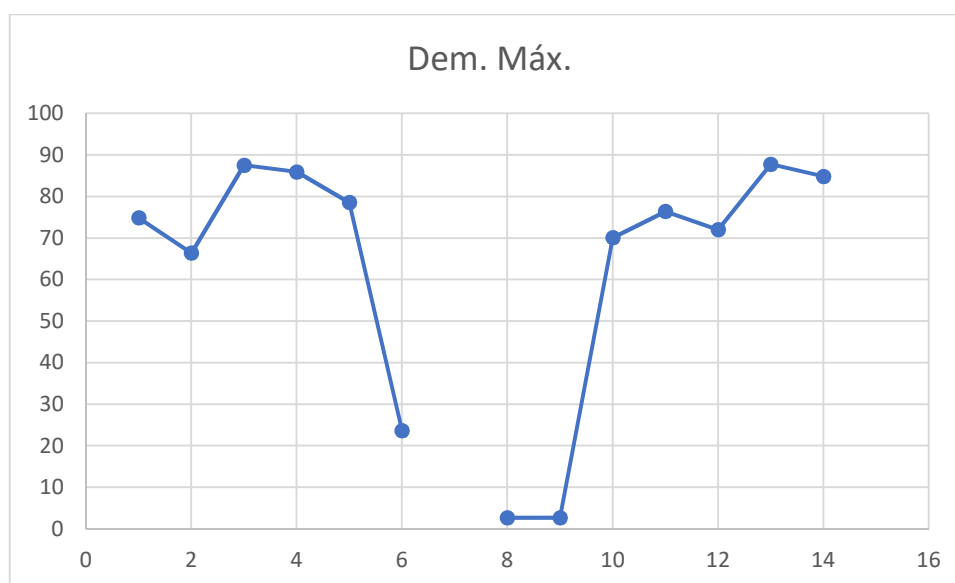
provavelmente, a ocorrência foi causada pelo mal funcionamento dos motores da máquina, afetando o equilíbrio entre produção de trabalho e perdas nos campos eletromagnéticos.

Após o religamento do equipamento, o valor desse indicador se manteve no limite mínimo aceitável, variando algumas vezes para baixo desse limite, sendo muito arriscado atuar nesse limiar. Esse cenário sugere a necessidade de realizar uma manutenção no equipamento para correção desse fator de potência, a fim de elevar a conversão de trabalho e reduzir as perdas, e assim reduzir a chance de a indústria ser multada.

A Figura 29 revela uma breve variação nos valores obtidos de Demanda Máxima. A partir daí, seria possível cogitar que o problema teria sido causado pela demanda crescente do equipamento, chegando a um ponto onde não mais conseguiu obter energia suficiente para funcionar. Mas não temos como comprovar essa hipótese via nossa aplicação.

A partir do alerta gerado pelos indicadores obtidos pela aplicação, a equipe de manutenção teria que realizar um estudo mais elaborado e mais longo para observar o comportamento do equipamento.

Figura 29- Gráfico de Demanda Máxima



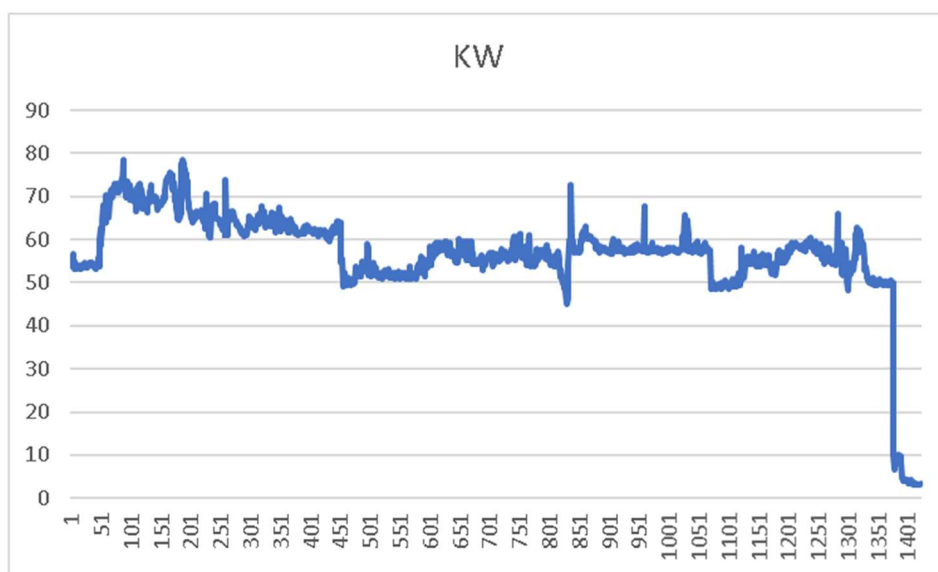
Fonte: Autor

As Figuras 30 e 31 mostram o consumo em kW do equipamento. O primeiro gráfico representa o consumo durante apenas o dia 10, pouco antes da máquina entrar em manutenção. É possível observar uma constância no

consumo antes da parada, sem muitos picos de consumo, revelando um bom funcionamento antes da parada. Isso sugere a possibilidade de a parada ter sido apenas de rotina, ou algum apagão ocasional.

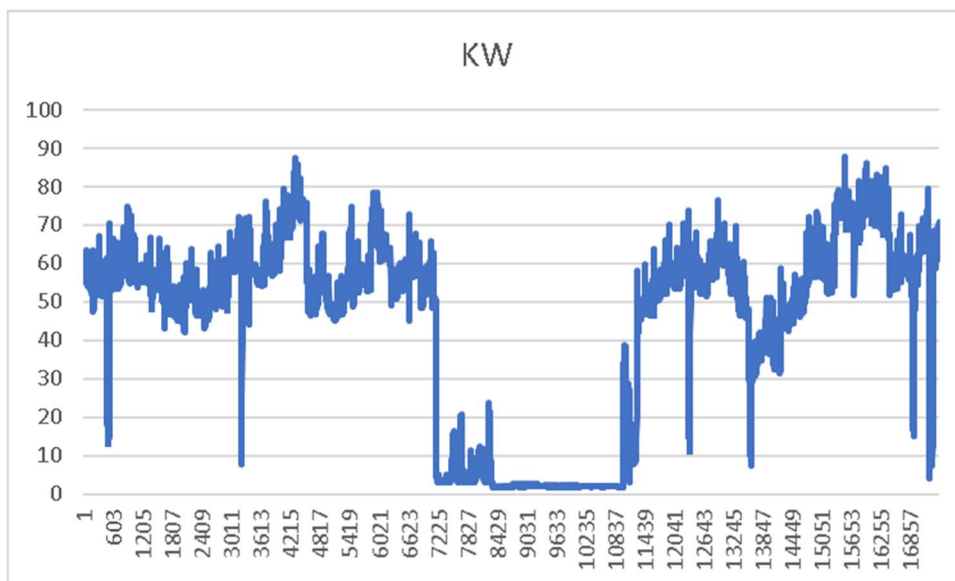
Já na Figura 31, é possível observar que a máquina apresentou algumas leves falhas durante o funcionamento, com dois picos negativos visíveis antes da parada. Após o retorno do equipamento, algumas falhas continuam sendo observadas, revelando uma possível necessidade por nova manutenção.

Figura 30- Gráfico de Consumo em KW no dia 10Dez2020.



Fonte: Autor

Figura 31- Gráfico de Consumo total em KW.



Fonte: Autor

5.3.3 Comparação com os Trabalhos relacionados

Esta seção faz uma breve comparação entre trabalho desenvolvido e apresentado aqui e os trabalhos relacionados descritos na seção 3.3. Utilizaremos as mesmas características destacadas na Tabela 2, a fim de manter os mesmos critérios já utilizados na comparação anterior.

Tabela 6. Comparação entre nossa aplicação e os trabalhos relacionados

TRABALHOS RELACIONADOS						
QP		Nesa (2017)	(Torres, A.B., (2017)	(Mohamed, I.; 2016)	(Callegaro, 2014)	Esse trabalho
QP1						
	Aplicação	Monitoramento remoto do consumo de energia em um edifício a partir da detecção ocupantes.	Hydra Agricultura inteligente (outdoor)	Harmoni Monitoramento de Saúde (indoor)	Monitoramento de ambiente (Outdoor)	Monitoramento remoto do consumo de energia Industrial & doméstico (indoor)
	Objeto do monitoramento	temperatura ,umidade, luz e CO2	Unidade do solo, sensor de chuva e temperatura	Oximetria de Pulso e frequência cardíaca	Pressão Atmosférica	Consumo de energia
	Tecnologia IoT das redes	ZigBee entre sensores e centro de fusão de dados(DC), Wifi entre o DC e a aplicação em nuvem.	Comunicação via radio wireless dos sensores para o nó concentrador e wifi do concentrador à internet.	Sensores comunicam-se com o PDA via Bluetooth PDA via wifi	Não especificada	Sensor e gateway LoraWAN
	Monitoramento remoto e em tempo real	Remoto em tempo real	Remoto em tempo real	Remoto em tempo real	Remoto em tempo real	Remoto em tempo real
	Topologia	estrela	estrela com nó concentrador	estrela	estrela com nodo concentrador	estrela
	Avaliação qualitativa da rede		++	-		++
QP2						
	Arquitetura de Fusão de dados	Não informado	Multinível em 3 camadas	Multinível em 3 camadas	Proposta multinível em 3 camadas, no entanto, implementada solução na camada de baixo nível.	Multinível em 3 camadas
	Identificação e descarte de outliers	Não informado sobre tratamento de outliers.	Descarte de ruídos e erros online no sensor; identificação de eventos off-line	Descarte online no PDA e identificação de eventos off-line	Descarte de ruídos no sensor; Não tratou identificação de eventos.	Descarte de ruídos e erros online no sensor; identificação de eventos off-line
	Melhora da acurácia dos resultados		(+) descarte inicial de ruídos e erros;	(+) descarte inicial de ruídos e erros;	(+) descarte inicial de ruídos e erros;	(+) descarte inicial de ruídos e erros;
QP3						
	Definição de regras e parâmetros relevantes para o processo de análise dos dados	Não especificado	regras definidas por parâmetros	O usuário pode definir essas regras a qualquer tempo via interface. As regras geram novos parâmetros	Não implementado	Usuário e desenvolvedor. O usuário pode definir essas regras a qualquer tempo via interface.
	Visualização dos resultados das análises	Não especificado	Gráficos			Gráficos e indicações de ações preventivas e corretivas
	O sistema realiza ações diretamente no ambiente	Não	Controla o tempo de irrigação do solo online a partir dos dados sensoriais	Não	Não	É possível, em outros estudos de caso.

diferencial da arquitetura proposta nesse trabalho é o fato dela ser multinível, onde a fusão de dados pode ser feita de forma modularizada e colaborativa, com vistas a identificação de *outliers* e eventos, além de possibilitar a fusão de decisões. Permite o acompanhamento dos indicadores através de *dashboards*

e consultas de forma remota. Embora não tenha sido implementada a infraestrutura de comunicação permite a possibilidade de o sistema atuar direto no ambiente, onde é possível por exemplo efetuar o ligar/desligar um equipamento de forma programada e sem intervenção humana.

Outro ponto relevante é que, por ser baseada na organização de abstração dos dados de entrada e saída proposta por Dasarathy (1997), torna possível a fusão não apenas para redução dos dados a serem transmitidos, mas também para identificação de eventos que necessitem a tomada de decisão.

5.4 CONSIDERAÇÕES FINAIS

Este capítulo apresentou a implementação de uma RSSF baseada na AMnFD, uma aplicação que tem por objetivo monitorar o consumo de energia em tempo real, independente do contexto aplicado. A aplicação foi testada em dois estudos de caso realizados em contextos distintos, tendo conseguido detectar *outliers* corretamente em ambos os casos.

A implementação da rede em camadas mostrou-se uma decisão acertada. O papel de cada área na identificação de *outliers* ficou evidente. A decisão de processar os dados brutos na camada física, com descarte das leituras com erros, provou ser eficiente, poupando assim recursos computacionais para transitar dados com erros por toda a rede.

No entanto, também ficou evidente que a ausência desses dados, causando lapsos entre as leituras, precisa ser tratada. Essa tarefa fica indicada como trabalho futuro, através do desenvolvimento de um algoritmo implementado na camada de médio nível, para identificar e tratar esses lapsos com vistas a preservar a coerência das leituras. Uma possível solução poderia ser o preenchimento dessas lacunas com um valor de leitura estimado, como a média das leituras válidas num período. Isso ajudaria a preservar os valores de forma temporal.

Existe uma vasta literatura no tratamento de dados faltosos em séries temporais. Essas técnicas de estimação de valores faltosos são conhecidas como imputation (ALMEIDA et al., 2010). Existem outras técnicas mais sofisticadas como ARIMA (*Autoregressive Integrated Moving Average*), onde é possível obter valores mais precisos. O contexto de atuação onde pode ser

aplicada as técnicas é bastante variado . (BRAGA *et al.*, 2012),(COUTINHO *et al.*, 2014),(FERREIRA, 2016).

O lapso das leituras por um determinado tempo também demonstrou a eficiência da camada de gestão em tomar decisões, neste caso, comunicando ao responsável (contexto industrial) sobre a ausência de leitura e a necessidade de verificação do equipamento. Isso foi fundamental para a rápida tomada de ação e a constatação da necessidade de manutenção, nesse caso, corretiva.

A seguir, o capítulo 6 apresenta as principais contribuições deste trabalho, e destaca algumas extensões (trabalhos futuros) que objetivam sanar as lacunas existentes na solução atual.

6 CONCLUSÃO

Este documento apresentou o trabalho desenvolvido durante o curso de mestrado, cujo foco principal foi de propor uma arquitetura multinível de fusão de dados para detecção de *outliers*. Trata-se de um trabalho original, e com contribuição relevante para a área de monitoramento com detecção de *outliers*.

A AMnFD foi implementada em uma RSSF para monitoramento de consumo de energia, tendo sido testada em dois estudos de caso, um em contexto residencial e outro em contexto industrial. Em ambos os casos, a aplicação foi capaz de detectar corretamente a ocorrência de *outliers* a partir dos dados monitorados, bem como forneceu indicadores de grande utilidade para os clientes industriais.

A implementação da AMnFD demonstrou que segmentar/modularizar o desenvolvimento traz benefícios para a RSSF, de modo que cada camada da arquitetura efetua seu papel de forma colaborativa, contribuindo para o bom funcionamento de toda a aplicação. Isso melhora o desempenho da rede, além de conferir maior confiabilidade aos dados. Também é possível apresentar os resultados das análises através de relatórios e/ou *dashboards*, auxiliando no monitoramento de indicadores que, quando devidamente gerenciados, contribuem para um melhor gerenciamento energético.

A seguir, destacamos algumas das principais contribuições deste trabalho. Por fim, listamos alguns dos trabalhos futuros que podem ser desenvolvidos a partir do que foi realizado neste projeto.

6.1 PRINCIPAIS CONTRIBUIÇÕES

Como principais contribuições desse trabalho, podemos relacionar:

- Foi possível testar a tecnologia de comunicação LoRa, onde se mostrou ser uma excelente opção por se manter estável e segura. Não foi detectado perda de sinal entre os sensores e o *gateway* mesmo tendo obstáculos entre os dois.
- Definição de uma arquitetura multinível para fusão de dados de monitoramento independente de domínio, com foco na detecção de *outliers* e eventos. Aqui, a fusão de dados ocorre de forma modularizada e colaborativa, auxiliando a tomada de decisões;

- A camada de comunicação da AMnFD permite o acompanhamento dos indicadores calculados através de gráficos e *dashboards* com consulta remota;
- Implementação de uma RSSF para o contexto de monitoramento de energia elétrica, objetivando eficiência energética;
- Realização de dois estudos de caso em contextos distintos, confirmando a escalabilidade da aplicação implementada;
- Os resultados obtidos através da análise dos dados coletados em ambos estudos de caso comprovam a eficácia da aplicação desenvolvida.

6.2 TRABALHOS FUTUROS

Como trabalhos futuros, enumeramos a seguir algumas melhorias relacionadas à implementação da RSSF, bem como extensões mais elaboradas da nossa proposta.

- Implementar um módulo de interpretação dos dados monitorados, com vistas a identificar a causa dos *outliers* (se são eventos normais ou falhas no equipamento);
- Implementar técnicas que consigam tratar os lapsos de dados gerados por eventuais falhas de comunicação e/ou nos nós sensores;
- Utilizar técnicas de Aprendizagem de Máquina para predição de falhas e/ou para sugerir manutenção preventiva de equipamentos;
- A arquitetura projetada possibilita a implementação de interação direto com o ambiente sendo monitorado (percepção e ação). Essa funcionalidade não foi implementada na RSSF utilizada nos estudos de caso realizados, porém a infraestrutura da AMnFD permite essa possibilidade;

Estender a RSSF atual para monitorar vários equipamentos simultaneamente (com vários nós sensores);

- Implementar outra RSSF para aplicação em um novo domínio (por exemplo, monitoramento de temperatura), confirmando sua independência de domínio.

REFERÊNCIAS

- ABES. **Mercado brasileiro de software: panorama e tendências**. Associação Brasileira das Empresas de Software, . ISBN 978-8586700-03-3, São Paulo, 2019.
- AKYILDIZ, I. F., SU, W., SANKARASUBRAMANIAM, Y., & CAYIRCI, E. **Wireless sensor networks: a survey**. Computer networks, 38(4), 393-422, 2002.
- ALEXANDER, C. K. , SADIKU, M. N. **Fundamentos de circuitos elétricos** ,5ª Edição Bookman, Porto Alegre, 2003.
- ALMEIDA, Antonia Fabiana Marques; MOREIRA, Maria Elisabeth Pinheiro; GOMES, Marcos José Timbó Lima. **Aplicação de modelos de séries temporais para estimação de dados faltosos de volume de tráfego das rodovias estaduais do Ceará**. Fortaleza, 2010.
- ANDRÉ, P. **Detecção e identificação de outliers em redes de sensores sem fio de larga escala**. Florianópolis : Qualificação de Mestrado - Universidade Federal de Santa Catarina, p.36 , 2017.
- ARANHA, C.; PASSOS, E. **A Tecnologia de Mineração de Textos**. RESI - Revista Eletrônica de Sistemas de Informação. v. 5. n. 2. Rio de Janeiro, 2006. Disponível em: <<http://www.periodicosibepes.org.br/index.php/reinfo/article/view/171/66>>. Acesso em: 06 Dez. 2019
- ARAUJO, R. B. **Computação Ubíqua: Princípios, Tecnologias e Desafios**, In: Simpósio Brasileiro de Redes de Computadores, XXI, Natal. Minicurso: Livro Texto. Natal, RN: UFRN/DIMAp: UnP, p. 45-115,2003
- ASHTON, K. That 'Internet of Things' Thing. 2009. Disponível em: <<https://www.rfidjournal.com/that-internet-of-things-thing> >. Acesso em 14 Feb 2020.
- BARNETT, V; LEWIS, T. **Outliers in statistical data**.Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics, 1984.
- BELL, B.; SANTOS, E.; BROWN, S. M. **Making adversary decision modeling tractable with intent inference and information fusion**. ADVERSARIAL INTENT INFERENCE FOR PREDICTIVE BATTLESPACE AWARENESS, p. 4, 2005.
- BEN21 **Balanco Energético Nacional**. Disponível em: <<https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/balanco-energetico-nacional-2021>> , Aceso em: 10 Ago, 2021.
- BENEFIT, APPLICATION. **The evolution of wireless sensor networks**. Silicon Lab. Inc., Austin, TX, USA, White Paper Rev, v. 1, p. 1-5, 2013.

BHOJANNAWAR, SATISH S.; BULLA, CHETAN M.; DANAWADE, VISHAL M. **Anomaly detection techniques for wireless sensor networks-a survey**. International Journal of Advanced Research in Computer and Communication Engineering, v. 2, n. 10, p. 3852-3857, 2013.

BOSTRÖM, H.; ANDLER, S. F.; BROHEDE, M.; JOHANSSON, R.; KARLSSON, A., VAN L.; J., ... & Ziemke, T. **On the definition of information fusion as a field of research**. 2007.

BOYD, J. R. **A discourse on winning and losing**: Unpublished set of briefing slides available at Air University Library. Alabama, USA: Maxwell AFB 1987.

BRAGA, Célia Campos et al. **Aplicação dos componentes principais na simulação e consistência de séries temporais**. Revista de Geografia (UFPE), v. 29, n. 1, p. 113-125, 2012.

CALLEGARO, R; MONTEZ, C., PINTO, A. R., & Moraes, R. **Uma Arquitetura para Fusão de Dados e Detecção de Outliers em Sensores de Baixo Custo de Redes de Sensores sem Fio**. In: Anais do II Workshop de Comunicação em Sistemas Embarcados Críticos-WoCCES, S. p. 3-16, 2014

CASTANEDO, F.. **A review of data fusion techniques**. The scientific world journal, v. 2013, 2013.

CHANDOLA, V.; BANERJEE, A.; KUMAR, V. **Anomaly detection: A survey**. ACM computing surveys (CSUR), v. 41, n. 3, p. 1-58, 2009.

COUTINHO, Eluã Ramos et al. **Aplicação de técnicas de inteligência computacional no preenchimento de falhas de séries temporais climatológicas**. 2014.

DASARATHY, B. V. **Sensor fusion potential exploitation-innovative architectures and illustrative applications**. Proceedings of the IEEE, v. 85, n. 1, p. 24-38, 1997.

DEMPSTER, A. P. **A generalization of Bayesian inference**. Journal of the Royal Statistical Society: Series B (Methodological), 30(2), 205-23, 1968.

DURRANT-WHYTE, H. F. **Sensor models and multisensor integration**. In: Autonomous robot vehicles. Springer, New York, NY. p. 73-89, 1990
HALL, D. L.; LLINAS, J. **An introduction to multisensor data fusion**. Proceedings of the IEEE, v. 85, n. 1, p. 6-23, 1997.

FERREIRA, J.; TAPAJÓS, R.; CONDE, G. **Redes neurais artificiais para o preenchimento de falhas em séries temporais meteorológicas**. 2016.

FREITAS, IGOR WESCLEY SILVA de et al. **Um estudo comparativo de técnicas de detecção de outliers no contexto de classificação de dados**. 2019.

HIMEUR, Yassine et al. **Data fusion strategies for energy efficiency in buildings: Overview, challenges and novel orientations**. Information Fusion, v. 64, p. 99-120, 2020.

JAMES, G.; WITTEN, D.; HASTIE, T. & TIBSHIRANI, R. **An introduction to statistical learning**. New York: springer, 2013.

KHALEGHI, Bahador et al. **Multisensor data fusion: A review of the state-of-the-art**. Information fusion, v. 14, n. 1, p. 28-44, 2013.

LEITE, J. R. E; MARTINS, P; URSINI, L. E. **A Internet das Coisas (IoT) : Tecnologias e Aplicações**. 2021. 6 f. Monografia (Especialização) - Curso de Asda, Unicamp, São Paulo, 2021. Disponível em: https://www.researchgate.net/profile/Paulo-Martins-25/publication/325807577_A_INTERNET_das_COISAS_IoT_Tecnologias_e_Aplicacoes_ISSN_2447-8326/links/5b257514458515270fd40d5d/A-INTERNET-das-COISAS-IoT-Tecnologias-e-Aplicacoes-ISSN-2447-8326.pdf. Acesso em: 31 mar. 2021.

LEITE, J. R. E; MARTINS, P; URSINI, L. E. **A Internet das Coisas (IoT) : Tecnologias e Aplicações**. In: Brazilian Technology Symposium. 2017.

LOUREIRO, A. A. F., NOGUEIRA, J. M. S., RUIZ, L. B., DE FREITAS MINI, R. A., NAKAMURA, E. F., AND FIGUEIREDO, C. M. S. **Redes de sensores sem fio**. In Simpósio Brasileiro de Redes de Computadores, pages 179 – 226, 2003

LUO, R. C.; YIH, C.; SU, K. L. **Multisensor fusion and integration: approaches, applications, and future research directions**. IEEE Sensors journal, v. 2, n. 2, p. 107-119, 2002.

MCTIC. **Identificação dos tópicos de relevância para a viabilização da Internet das Coisas no Brasil**. Ministério da Ciência, Tecnologia, Inovações e Comunicação, 2016. Disponível em: www.abinee.org.br/informac/arquivos/aiot.pdf , Acesso em 24 Jan 2022

MOHOMED, I.; MISRA, A., EBLING, M., & JEROME, W. **HARMONI: Context-aware filtering of sensor data for continuous remote health monitoring..** In: 2008 Sixth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom): Hong Kong, March. p. 17-21, 2008.

NAKAMURA, E. F.; LOUREIRO, A. A. F; FRERY, A. C. **Information fusion for wireless sensor networks: Methods, models, and classifications**. ACM Computing Surveys (CSUR), v. 39, n. 3, p. 9-es, 2007.

NESA, N.; BANERJEE, I. **IoT-based sensor data fusion for occupancy sensing using Dempster–Shafer evidence theory for smart buildings**. IEEE Internet of Things Journal, v. 4, n. 5, p. 1563-1570, 2017.

OASIS Standard, 2019. Disponível em: <https://docs.oasis-open.org/mqtt/mqtt/v5.0/os/mqtt-v5.0-os.html>. Acesso em Dez/2021

RANGEL, R. **Internet das coisas: nova revolução da conectividade**, Revista Inovação da Pauta, n. 18, p. 6-9, 2014

RASSAM, M. A.; ZAINAL, A.; MAAROF, M. A. Advancements of data anomaly detection research in Wireless Sensor Networks: A survey and open issues. *Sensors*, v. 13, n. 8, p. 10087–10122, 2013.

RUIZ, L. B., CORREIRA, L. H. A., VIEIRA, L. F. M., MACEDO, D. F., NAKAMURA, E. F., FIGUEIREDO, C. M., ... & FERNANDES, A. O. **Arquiteturas para redes de sensores sem fio**. Tutorial of the Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC), 2004.

SEMTECH Semtech LoRa and LoRaWAN: **A Technical Overview**. 2019 Technical Paper, Disponível em: https://loradevelopers.semtech.com/uploads/documents/files/LoRa_and_LoRa_WAN-A_Tech_Overview-Downloadable.pdf > Acessado em Dez/2021.

SEVERINO, V.; VASCONCELOS, D.; MILFONT, R.; ANDRADE, R.; SANTOS, I., & ANDRADE, C. **Produto Mínimo Viável para Monitoramento Elétrico em Smart Homes**. In: Anais do II Workshop Brasileiro de Cidades Inteligentes. SBC, 2019.

SHAHID, N.; NAQVI, I. H.; QAISAR, S. B. **Characteristics and classification of outlier detection techniques for wireless sensor networks in harsh environments: a survey**. *Artificial Intelligence Review*, v. 43, n. 2, p. 193-228, 2015.

SHARDA, R., DELEN, D., TURBAN, E. **Business Intelligence e Análise de Dados para Gestão do Negócio**, 4ª Edição Bookman, Porto Alegre, 2019.

SICSU, A. L.; DANA, S. **Estatística aplicada**. Saraiva Educação SA, 2017.

SILVA, L. H. de S. **Desenvolvimento de uma rede de sensores sem fio utilizando zigbee para aplicações diversas**. In: . [S.l.]: Trabalho de Conclusão de Curso Engenharia da Computação, 2011. Citado 6 vezes nas páginas 13, 15, 16, 17, 18 e 21.

SLAB. Silicon Laboratories. The Evolution of Wireless Sensor Networks. 2013. 1–5 p.

TORRES, A. B. B. **Fusão de dados multinível para sistemas de internet das coisas em agricultura inteligente**. 2017.

WANG, H.; BAH, M. J.; HAMMAD, M.. **Progress in outlier detection techniques: A survey**. *Ieee Access*, v. 7, p. 107964-108000, 2019.

WHITE, F. E. **Data fusion lexicon**. Joint Directors of Labs Washington DC, 1991.

YICK, JENNIFER; MUKHERJEE, BISWANATH; GHOSAL, DIPAK **WIRELESS SENSOR NETWORK SURVEY**. *Computer networks*, v. 52, n. 12, p. 2292-2330, 2008.

YINBIAO ,S. ; LEE, K.; LANCTOT, P.; JUANBIN, F.; HAO, H.; CHOW, B.; QUI, W.

Internet of Things: wireless sensor networks Int. Electron. Commision, pp. 1-78, 2014.

YOKOGAWA, C. S; DA SILVA, B. R. **Dashboard para monitoramento de energia elétrica. Revista Eletrônica eF@ tec**, v. 10, n. 1, p. 13-13, 2020.

ZHANG, Y.; MERATNIA, N; HAVINGA, Paul. **Outlier detection techniques for wireless sensor networks: A survey.** IEEE communications surveys & tutorials, v. 12, n. 2, p. 159-170, 2010.