



Universidade Federal de Pernambuco  
Centro de Informática

Graduação em Engenharia da Computação

**Desenvolvimento de Sistema de Localização  
Interna Baseado no Processador Edison e Redes  
Locais sem Fio**

Lucas Felix Lima Barbosa

Trabalho de Graduação

Recife  
2017

Lucas Felix Lima Barbosa

**Desenvolvimento de Sistema de Localização Interna Baseado no  
Processador Edison e Redes Locais sem Fio**

*Trabalho apresentado ao Programa de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.*

Orientadora: *Profa. Dra. Edna Natividade da Silva Barros*

Recife  
2017

*A Deus, por ter me ajudado até aqui.*

# Agradecimentos

Agradeço primeiramente a Deus, pois sem Ele eu nada seria, o criador de tudo o que há e que sempre me ajudou e cuidou de mim, dando sabedoria e ajudando desde o primeiro dia até a conclusão deste trabalho, assim como em toda minha vida.

Agradeço a Valdirene e Leonilda, minha mãe e minha tia, que são as duas pessoas mais importantes da minha vida, que me criaram e sempre cuidaram de mim, me puseram nos melhores caminhos para que eu pudesse chegar até aqui com a graça de Deus, sempre lutaram por mim e muitas vezes sacrificaram-se para que tudo hoje fosse realizado e concretizado.

Agradeço a Livia, minha noiva, que me acompanha nessa trajetória a muito tempo e está sempre ao meu lado, me incentivando, me ajudando, dando sugestões, e principalmente acreditando em mim e no meu trabalho.

Agradeço a Matheus e Edicleison, meu irmão e meu pai, que sempre me deram forças e acreditaram em mim e no meu trabalho.

Agradeço especialmente a Profa. Edna Barros, que ao longo de toda minha trajetória acadêmica foi mais que minha orientadora, foi uma grande amiga e até mesmo uma mãe, que me acolheu, me incentivou, abriu portas, me ajudou e me guiou por um caminho no qual não me arrependo, sempre paciente e dedicada aos meus trabalhos, sempre me ouvia como também criticava, melhorando assim não só os meus trabalhos como também meu desempenho acadêmico, agradeço por neste último trabalho da graduação estar tão presente e ter ajudado na conclusão do mesmo, acreditando e sempre ajudando em tudo que eu precisasse.

Agradeço em especial ao sr. Antônio Barros, que foi praticamente meu orientador também, onde além de mentor e idealizador, foi e é um grande amigo, sempre disponível para ajudar e pela sua dedicação ao desempenho do meu trabalho, além da paciência que teve em tirar todas minhas dúvidas para realização deste trabalho.

Por fim agradeço a todos os familiares e amigos, que estiveram torcendo, incentivando e orando para o sucesso deste trabalho.

# Resumo

Atualmente as redes locais sem fio (WLANs) em ambientes internos estão presentes na maioria dos prédios públicos. Estas redes, além da sua função principal, podem ser utilizadas para localização de pessoas e objetos, salientando que nestes ambientes não é adequada a utilização do sinal de GPS para esta finalidade. Diversos estudos e pesquisas nesta área têm sido realizados. Serviços baseados na localização interna possuem inúmeras aplicações nas áreas de segurança, médica, monitoramento, navegação, auxílio a deficientes, gerenciamento de pessoas, entre outras e hoje já movimentam um mercado de US\$1 Bilhão. Com a proliferação da Internet das Coisas (IoT), estes valores serão ainda maiores.

Os sistemas de localização interna utilizam tecnologias como GPS, RFID, Bluetooth, Zigbee e WLAN, variando conforme a precisão, exatidão, custo, velocidade de resposta, infraestrutura e aplicação.

O presente trabalho propõe uma melhoria na técnica de localização interna do trabalho de Antônio Barros [de Assunção Barros, 2016] que utiliza a intensidade do sinal recebido (*RSS* — *Received Signal Strength*) das redes WLAN presentes como medida para localização. Na técnica proposta, é feito inicialmente um mapeamento das intensidades dos sinais da WLANs existentes. Estes valores são classificados através de um Algoritmo de Agrupamento (*K-Means clustering*) e, posteriormente, é aplicado, a cada agrupamento, um algoritmo de regressão não linear através de uma rede neural artificial (*MLP* — *Multilayer Perceptron*) para o cálculo da localização. Associada a estas técnicas são aplicados filtros visando minimizar as variações do sinal medido decorrentes de interferências do meio. Esta técnica não necessita de grandes esforços de calibração nem alterações na estrutura existente, apenas utilizando a rede WLAN já instalada, obtendo assim uma precisão compatível com aplicações de localização de pessoas e objetos e auxílio em navegação em ambientes internos.

Na implementação e testes da técnica proposta, foi construído um sistema embarcado com o SoC Edison da Intel para a coleta das intensidades dos sinais — RSS fazendo assim o mapeamento do local, como também para a localização. Como plataforma para processamento dos dados foi utilizado um servidor Local no Centro de Informática, onde são executados os algoritmos e a comunicação com o sistema embarcado. As medições para composição dos conjuntos de testes e validação foram realizadas no prédio do Centro de Informática da UFPE, demonstrando que apesar do baixo esforço de calibração, sem alteração da estrutura existente, atendem aos requisitos necessários. Resultados preliminares mostram que 50% das amostras estavam com erro inferior a 7,88 metros.

**Palavras-chave:** Redes sem fio. Localização Indoor. WLAN, Wi-Fi. RSS. IPS. Clustering. Sistemas Embarcados.

# Lista de Figuras

2.1	Arquitetura das camadas MAC e Física no padrão 802.11[Akin and Geier, 2004].	12
2.2	Faixa de medição do valor RSS no quadro PPDU [IEEE, 2012].	13
2.3	Sistema utilizando Agrupamento de Dados e Redes Neurais [de Assunção Barros, 2016].	15
2.4	Representação do Processo de Regressão.	18
2.5	Processo de Localização [de Assunção Barros, 2016].	20
3.1	Método proposto automatizado, Parte Off-line e On-line.	22
3.2	Vista do Usuário, Vista de Cima e Vista Frontal do capacete.	23
3.3	Diagrama de blocos do Processador Edison[Intel, 2015].	24
3.4	Arquitetura de Software da Placa de Rede do Edison[Intel, 2015].	25
3.5	Estrutura do Servidor Local.	27
3.6	Fluxograma Básico do Método K-Means.	31
3.7	Modelo Neurônio Artificial [Neural Network in a Nutshell, 2016].	32
3.8	Processo de Localização.	34
4.1	Planta Baixa do Centro de Informática - UFPE.	36
4.2	Regiões Virtuais do CIn.	44
4.3	Agrupamento dos pontos formando as Regiões Virtuais do CIn.	45
4.4	Resultado das Redes Neurais	46

## Lista de Tabelas

4.1	Parâmetros K-Means.	37
4.2	Configuração das Redes Neurais.	38
4.3	Uma amostra dos resultado dos valores RSS fornecido pelo SoC Edison.	39
4.4	Quantidade de APs por SSID no CIN-UFPE.	41
4.5	Uma amostra do mapeamento dos sinais RSS, com as posições X,Y e número da Varredura.	42
4.6	Métrica do Agrupamento conforme o método "silhueta".	43
4.7	Resultado Regiões Virtuais.	45
4.8	Resultado Redes Neurais com Regiões Virtuais.	45
4.9	Comparativo entre sistemas.	47

# Sumário

<b>1</b>	<b>Introdução</b>	<b>5</b>
1.1	Motivação	5
1.2	Tecnologias de Localização baseadas em Rádio Frequência	6
1.2.1	GPS — <i>Global Positioning System</i>	6
1.2.2	RFID ( <i>Radio Frequency Identification</i> )	6
1.2.3	WLAN	7
1.2.4	Bluetooth	7
1.2.5	Zigbee	7
1.3	Sistemas de Localização indoor utilizando redes WLAN	8
1.3.1	Radar	8
1.3.2	Sistema Horus	8
1.3.3	O Sistema Ekahau	9
1.3.4	Sistema Cisco de Localização	9
1.3.5	Sistema utilizando Aprendizagem de Máquina - Antônio Barros	9
1.4	Objetivos	9
1.5	Estrutura do Trabalho	10
<b>2</b>	<b>Técnica de Localização Indoor Proposta por Antônio Barros</b>	<b>11</b>
2.1	RSS — <i>Received Signal Strenght</i>	11
2.2	Técnica de Localização Indoor baseada no Sina RSS	13
2.2.1	Parte Off-line	14
2.2.1.1	Análise do Ambiente	14
2.2.1.2	A Medição da intensidade do sinal — RSS	14
2.2.1.3	Regiões Virtuais	16
2.2.1.4	Treinamento com Redes Neurais	17
2.2.2	Parte On-line	18
2.2.2.1	Localização	18
2.3	Principais Problemas da Implementação de Antônio Barros	19
<b>3</b>	<b>Proposta de Implementação Automatizada e Local para o Método de Barros</b>	<b>21</b>
3.1	Visão Geral da Técnica Implementada	21
3.1.1	Sistema Embarcado Móvel	23
3.1.2	Servidor Local	26
3.2	Módulo de Medição da Intensidade do Sinal RSS	26
3.2.1	Módulo de Leitura do Sinal	29

3.2.2	Módulo de Tratamento dos Dados	29
3.2.3	Módulo de Comunicação com o Servidor	30
3.3	Módulo de Treinamento Off-line	30
3.3.1	Algoritmo K-means	31
3.3.2	Rede Neural MLP — <i>Multilayer Perceptron</i>	32
3.4	Módulo de Localização On-line	33
<b>4</b>	<b>Experimentos e Resultados</b>	<b>35</b>
4.1	Ambiente e Equipamentos Utilizados	35
4.1.1	Descrição do Ambiente do CIn	35
4.1.2	Escolha do parâmetros	35
4.1.2.1	Medição das intensidades dos Sinais — RSS no ambiente.	37
4.1.2.2	Parametrização para definição das regiões virtuais	37
4.1.2.3	Parametrização da etapa de localização	38
4.2	Medições Realizadas	38
4.2.1	Medições com o Sistema Embarcado	39
4.2.1.1	Valores RSS	40
4.2.1.2	Resultados RSS	40
4.2.1.3	Mapeamento RSS	40
4.2.2	Criação das Regiões Virtuais e Redes Neurais	43
4.3	Resultados - Localização	43
4.3.1	Alocação nas Regiões Virtuais	43
4.3.2	Saídas das Redes Neurais	45
4.4	Comparativo com o Sistema de Barros	45
4.5	Implementação	48
<b>5</b>	<b>Conclusões e trabalhos futuros</b>	<b>49</b>
5.1	Conclusões	49
5.2	Trabalhos Futuros	51
	<b>Referências Bibliográficas</b>	<b>53</b>

# 1 Introdução

## 1.1 Motivação

As pessoas sempre precisaram se localizar. Desde a antiguidade a necessidade de localização foi imprescindível para o ser humano e esta necessidade favoreceu o desenvolvimentos de várias ciências. Hoje os avanços nas tecnologias de comunicação e computação estão permitindo o desenvolvimento de sistemas de localização cada vez mais avançados e precisos. O mercado de sistemas de localização interiores movimenta anualmente mais de US\$ 1 Bilhão, com um crescimento anual de 36% e previsto para atingir US\$ 4 Bilhões em 2022 segundo a empresa Ekahau [Ekahau, 2016]. Estudos da empresa IDtechEx estimam um mercado de US\$ 10 Bilhões para 2024 [Peter Harrop, 2016].

Um dos sistemas de localização mais utilizados hoje é o GPS (*Global Positioning System*), que iniciou sua operação em 1995, e hoje está bastante difundido. O GPS, porém, possui a limitação de seu uso ser eficaz somente em áreas externas. Assim surgiram os sistemas de localização para áreas interiores ou Sistemas de Localização Indoor.

No caso dos sistemas de localização interna é utilizado o termo IPS (*Indoor Positioning System*) para a designação de sistemas que visam localizar, rastrear pessoas ou objetos em áreas internas, cobertas. Pode-se dizer que sistemas IPS são utilizados onde não se pode utilizar o GPS, ou seja, é o GPS interno. O termo RTLS (*Real-time locating systems*) foi utilizado por Tim Harrington (WhereNet) em 1995 para descrever a tecnologia de identificação da localização de um determinado objeto em uma tela de computador.

A diferença entre estes dois termos é muito sutil. A empresa IdtechEx define sistemas IPS como a tecnologia que utiliza os sensores e sinais Wi-fi (WLAN) de uma maneira ubíqua, enquanto, segundo a mesma empresa, sistemas RTLS suportam áreas maiores e, normalmente, são baseados na tecnologia de celulares bem como de alguns outros sensores como, por exemplo, o RFID. Apesar de ambos os termos serem encontrados na literatura, este trabalho usará o termo IPS para designar sistemas de localização em áreas interiores.

Assim o presente trabalho tem como objetivo contribuir na área de sistemas IPS automatizando as etapas de um método para localização de pessoas e objetos em ambientes interiores proposto por [de Assunção Barros, 2016], que consiste num método rápido, confiável e que utiliza a infraestrutura já existente, de maneira que possa ser aplicado em ambientes interiores, tais como: áreas públicas, escolas, hospitais, museus, aeroportos, estações e shopping centers.

## 1.2 Tecnologias de Localização baseadas em Rádio Frequência

A principal vantagem de se utilizar tecnologia de rádio frequência reside no fato de ondas de rádio poderem atravessar paredes, corpos humano e outros objetos, permitindo, assim, a cobertura de uma área bem maior. Nas subseções a seguir serão detalhadas as tecnologias baseadas em rádio frequência mais usadas em sistemas IPS.

### 1.2.1 GPS — *Global Positioning System*

Em 1973 iniciou-se o desenvolvimento do GPS, um sistema de localização baseado em 24 satélites, que emitem sinais de rádio. Estes satélites são posicionados de forma que pelo menos 4 deles fiquem visíveis de qualquer ponto da terra. De acordo com o momento em que estes sinais chegam a cada receptor, é calculada a posição geográfica (latitude e longitude) independentemente das condições meteorológicas.

O GPS foi projetado, inicialmente, para uso militar, mas foi lentamente sendo liberado para uso civil. As restrições para uso civil deixaram de existir, melhorando em aproximadamente dez vezes o nível de precisão. Os receptores atuais são capazes de trabalhar com os dois códigos de correção e são chamados de dupla frequência. Atualmente a precisão de um GPS é da ordem de 15 metros, porém existem GPS geodésicos que trabalham com as duas portadoras atingindo uma precisão da ordem de 1mm na horizontal e de 5mm na vertical e tal precisão é extremamente importante para estudos geodésicos e espaciais em micro e macro escala [Bernardi and Landín, 2002]. Salientando que estes são GPS geodésicos, com antenas especiais, não o que é utilizado nos celulares ou automóveis, mas que utilizam os mesmos satélites.

O GPS foi um sistema que desenvolveu-se rapidamente e, hoje, é raro uma pessoa que não possua um receptor incorporado ao *smartphone* e até mesmo nos sistemas embarcados.

### 1.2.2 RFID (*Radio Frequency Identification*)

A técnica de RFID consiste na emissão de um código identificador para um sensor através de rádio frequência, portanto, é uma tecnologia sem fio. Esta tecnologia possui dois tipos de sensores; o ativo e o passivo. O RFID passivo é de baixo custo, os dispositivos não possuem alimentação interna, operam sem bateria, sendo que sua alimentação é fornecida pelo próprio leitor através das ondas eletromagnéticas, e estes sensores são relativamente baratos. O RFID ativo diferencia do passivo, pois utiliza uma fonte de energia interna. Mais caro e menos popular, o RFID ativo consegue atingir distâncias maiores até cem de metros.

Devido ao baixo custo, o RFID passivo é mais utilizado atualmente. No entanto, possui a limitação de distância máxima de 3 metros. Apesar de ter um custo baixo, é limitado na distância máxima de alcance e normalmente identifica apenas se um determinado objeto se encontra em um determinado local ou não.

O sistema IPS baseado em RFID mais conhecido comercialmente é o sistema Wherenet da Zebra Technology Company [Gu et al., 2009], que possui uma precisão de 2-3 metros.

### 1.2.3 WLAN

Uma rede WLAN (*Wireless Local Area Network*), também denominada rede Wi-Fi (*Wireless Fidelity*), é um padrão de comunicação em redes sem fio baseados no padrão IEEE 802.11. Uma rede Wi-Fi opera na faixa de frequência de 2,4 GHz e/ou 5 GHz com alta taxa de transferência de dados. Pode ser considerada como uma tecnologia ubíqua pois está em toda parte e em todos os momentos. O número de áreas públicas que possuem redes Wi-Fi tem crescido muito nos últimos anos e, atualmente, praticamente não temos nenhuma área pública tais como hospitais, escolas, shopping centers, aeroportos que não possuam uma rede WLAN.

A principal motivação para o crescimento no número de redes Wi-Fi é o crescente número de dispositivos móveis que possuem sensores Wi-Fi que atualmente é estimado em 2 bilhões de dispositivos móveis no mundo. Devido à existência de uma estrutura instalada em um grande número de ambientes públicos e comerciais, bem como ao crescente número de dispositivos com sensores Wi-Fi, esta tecnologia é considerada como bastante promissora para ser utilizada nos sistemas IPS. Com a sua utilização é possível abranger o maior número de pessoas com um custo mais baixo. Várias pesquisas métodos e técnicas para localização se encontram em desenvolvimento como será visto posteriormente.

No caso da tecnologia WLAN é possível a utilização de uma infraestrutura já existente com bons resultados. Isto é, utilizar a rede de APs que estão presentes, para desenvolver um sistema de localização. Utilizando-se a WLAN, através do sinal de Rádio Frequência, utiliza-se dois tipos de técnicas para se obter a localização: Uma é a da triangulação e a outra é do mapeamento dos sinais existentes (*Fingerprinting*).

### 1.2.4 Bluetooth

O sistema Bluetooth é um padrão de comunicação sem fio através de ondas de rádio, que surgiu em 1994 e pode ser encontrado na maioria dos smartphones, laptops e tablets. O sensor possui um baixo consumo de energia, alcança distâncias de até 100 metros e possui um baixo custo. Quando comparado com a tecnologia Wi-Fi, possui uma menor taxa de transferência de dados. Este sensor foi projetado para substituir o sensor infravermelho, e opera em uma rede “piconet” com até oito elementos simultâneos. O Bluetooth é regulamentado pela norma IEEE 802.15, operando também na frequência de 2,4 GHz. Devido a estas características também é utilizado em sistemas IPS. Alguns sistemas comerciais que utilizam sensores bluetooth, em alguns casos combinados com outras tecnologias são o sistema Igate [Igate, 2016], o sistema Zonith [Zonith, 2016], o sistema Teldio [Teldio, 2016], entre outros.

### 1.2.5 Zigbee

A rede conhecida como ZigBee, criada pelo IEEE em conjunto com a ZigBee Alliance, foi criada com o intuito de disponibilizar uma rede de dados baseada em RF, com sinal sem fio, operando aqui no Brasil na frequência de 2,4 GHz, com extrema baixa potência de operação, ocasionando um baixo consumo de energia nos dispositivos, estendendo a vida útil de suas baterias, podendo as mesmas durarem anos. Dessa forma, a rede tem como principais casos de uso dispositivos que não necessitem de taxas de transmissão de dados tão altas quanto as per-

mitidas pelo Bluetooth, e querem se aproveitar das características de baixo consumo [Vasques, 2010].

Esta técnica utiliza um emissor de rádio de baixa potência e necessita de uma rede já ZigBee, definida pelo padrão IEEE 802.15.4. Utilizando-se a tecnologia ZigBee para IPS, experimentos mostram que esta tecnologia possui uma boa precisão, com erros menores que 1 metro, baixo consumo de energia, apesar da baixa potência e possui alcance de até 100 metros com linha de visada.

Apesar desta tecnologia ter um baixo custo de implantação, necessita da implantação de uma infraestrutura específica, uma vez que os ambientes públicos normalmente não possuem este tipo de rede disponível. Além disto, possui uma baixa taxa de transmissão em comparação com outras tecnologias, operando em três frequências com diferentes taxas de transmissão: 868 MHz e 20kbps; 915 MHz e 40kbps; 2,4 GHz e 250 kbps.

É bastante utilizado em hospitais, pois além do sistema de localização, possui aplicações já desenvolvidas comercialmente para monitoramento de sinais vitais de pacientes [Alliance, 2009].

### **1.3 Sistemas de Localização indoor utilizando redes WLAN**

Os sistemas que suportam a localização em ambientes interiores envolvem áreas geográficas menores, a maioria das aplicações necessita de uma maior precisão e rapidez em comparação com os sistemas que suportam a localização em ambientes exteriores, devido aos objetos e locais a serem localizados, possuem uma menor dimensão, como por exemplo localizar uma sala, exige uma precisão maior que localizar uma casa, localizar um corredor em um edifício, exige uma precisão maior que localizar uma estrada. Diversas técnicas e métodos têm sido desenvolvidos para a solução do problema [Adler et al., 2015].

Nesta seção serão detalhados alguns sistemas de localização propostos por diferentes grupos de pesquisa. Desde o surgimento e popularização das redes WLAN, foram propostos trabalhos analisando diversos métodos de localização.

#### **1.3.1 Radar**

O sistema Radar [Bahl and Padmanabhan, 2000] foi o primeiro sistema proposto e citado em toda a literatura referente a sistemas IPS. Foi o primeiro estudo a provar que é possível, através da intensidade do sinal RSS, determinar a posição de um objeto em um ambiente interior. Este trabalho observa o decaimento da intensidade do sinal RSS em dispositivos móveis e faz a comparação de dois métodos para obter a localização.

#### **1.3.2 Sistema Horus**

O sistema Horus [Youssef and Agrawala, 2005] foi proposto por Moustafa Youssef e Ashok Agrawala. Este sistema utilizou pela primeira vez a técnica de "clustering" e também buscou identificar as causas da variação da intensidade do sinal WLAN. Com isto, conseguiu minimizar o custo computacional e obter alta precisão, e ainda suporta um alto número de usuários, já que roda localmente no dispositivo móvel e não em um servidor.

### 1.3.3 O Sistema Ekahau

O sistema Ekahau - RTLS *Real-Time Location System*, foi desenvolvido pela empresa finlandesa do mesmo nome e possui 10 patentes relacionadas ao seu sistema de localização [Myllymäki and Tirri, 2007]. O sistema é caracterizado por usar além da tecnologia Wi-Fi, também as tecnologias RFID e Infravermelho, como apoio para conseguir uma melhor precisão.

### 1.3.4 Sistema Cisco de Localização

A empresa Cisco também desenvolve pesquisas na área de localização, e cujo elemento principal é a *Cisco Location Appliance*. A solução proposta pela Cisco é toda baseada em uma infraestrutura de equipamentos de sua fabricação, utiliza a intensidade de sinal como elemento de localização e, como a maioria dos sistemas, também possui uma fase de treinamento bem detalhada, com vários levantamentos, para que o sistema opere bem.

### 1.3.5 Sistema utilizando Aprendizagem de Máquina - Antônio Barros

Este sistema foi proposto por Antônio Barros, neste trabalho são usadas técnicas de inteligência computacional, onde foram usados algoritmos de agrupamentos e redes neurais, de forma a suportar a localização baseada pela semelhança de grupos dos AP's e intensidade do sinal RSS - *Received Signal Strength* [de Assunção Barros, 2016]. O nosso método proposto foi baseado neste trabalho, que será explicado mais detalhadamente no capítulo a seguir.

## 1.4 Objetivos

O principal objetivo deste trabalho é propor uma implementação automatizada baseada em servidor local e sistema embarcado da técnica proposta por Barros [de Assunção Barros, 2016] (que será detalhada no capítulo seguinte) para localização de pessoas ou objetos em ambientes interiores que utilize a infraestrutura de comunicação Wi-Fi existente, disponível em grande parte de prédios comerciais e públicos, incluindo um sistema eletrônico embarcado utilizando o SoC Intel Edison que possui um módulo Wi-Fi que obedece ao padrão IEEE 802.11 para WLAN, para a localização.

Na implementação proposta, a localização é obtida a partir de uma medição prévia da intensidade dos sinais Wi-Fi RSS (*Received Signal Strength*) através do sistema eletrônico embarcado. Além do sistema embarcado, a utilização de um servidor local para execução dos algoritmos de aprendizagem de máquina que foram implementados é um das principais contribuições do trabalho pois permite a automatização de todo o processo de localização proposto no trabalho de Barros [de Assunção Barros, 2016], além de tornar o sistema mais eficiente e preciso.

## **1.5 Estrutura do Trabalho**

O presente trabalho está estruturado como descrito a seguir.

No Capítulo 2 é apresentada a técnica de localização indoor proposta por Antônio Barros [de Assunção Barros, 2016].

No Capítulo 3 é apresentada a proposta de implementação baseada em eletrônica embarcada e servidor local.

No Capítulo 4 são descritos os experimentos e resultados obtidos com a implementação proposta.

O Capítulo 5 apresenta uma análise da implementação incluindo vantagens e desvantagens. Melhorias a serem implementadas em versões futuras e alguns novos trabalhos decorrentes da técnica proposta.

## 2 Técnica de Localização Indoor Proposta por Antônio Barros

Este Capítulo apresenta inicialmente uma introdução ao sinal RSS, enfatizando as vantagens e desvantagens de usar este tipo de sinal na localização indoor, e em seguida será mostrado uma visão geral do trabalho feito por Antônio Barros, no qual o método proposto deste trabalho foi baseado.

### 2.1 RSS — *Received Signal Strength*

Diversos padrões que especificam redes sem fios surgiram na década de 1990. O que mais se difundiu e passou a ser o mais usado foi o padrão IEEE 802.11, também conhecido como Wi-Fi ou WLAN [Kurose and Ross, 2010]. Dentro do padrão 802.11 existem diferentes tecnologias, sendo os padrões mais conhecidos os 802.11a, 802.11b, 802.11g, 802.11n, que se diferenciam em termos de faixa de frequência e também taxa de transferência de dados, acompanhando a evolução tecnológica.

Diferente de outros sistemas de comunicação móvel, o padrão IEEE 802.11, não possui em sua especificação, técnicas para medição com precisão do tempo de chegada do sinal. Por isto, a característica mais utilizada do sinal de Rádio Frequência para estimação da localização é a intensidade do sinal — RSS. O padrão IEEE 802.11 fornece uma maneira de obtermos este sinal através da interface da placa de rede. Isto porque, esta intensidade pode ser utilizada para a escolha da rede em que o dispositivo deve se conectar, na maioria das vezes a de maior intensidade.

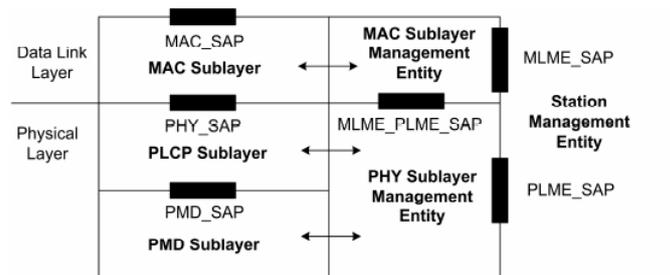
O valor da intensidade do sinal — RSS é gerado na camada física da rede, durante a recepção do sinais e descrito a seguir.

As camadas MAC (*Media Access Control*) e Física são representados na Figura 2.1.

De acordo com a Figura existem duas subcamadas relacionadas ao RSS e que fazem parte da camada física (PHY). A camada de Protocolo de Convergência da Camada Física de alto nível (*Higher Physical Layer Convergence Protocol* — PLCP) e a camada dependente do meio físico (*Lower Physical Media Dependent* — PMD). Estas camadas se comunicam através de SAPs (*Service Access Point*, Ponto de Acesso a Serviços).

A subcamada PLCP recebe da camada MAC o dado ou MPDUs (*MAC Protocol Data Units*) para transmissão, e prepara as MPDUs para entregar à sub camada PMD os frames para serem transmitidas pela antena. A estes são acrescentados ainda um cabeçalho que contém informações referentes ao meio físico, gerando assim as PPDU (*PLCP protocol data unit*).

Sob a direção da subcamada PLCP, a subcamada PMD (*Physical Medium Dependent*) rea-



**Figura 2.1** Arquitetura das camadas MAC e Física no padrão 802.11[Akin and Geier, 2004].

liza a transmissão e recepção da PPDUs entre duas estações via sinais de rádio, enviando para o circuito transmissor para transmissão ou recebendo no caso da recepção.

Cada *Access Point* envia periodicamente uma mensagem de *beacon*. Cada estação móvel recebe este sinal, de todos os APs, independente se está conectado a ele. O *beacon* é um quadro de sincronismo enviado periodicamente pelo ponto de acesso. Ele tem a função de avisar os clientes de que a rede está presente, avisar sobre quadros gravados no *buffer* do *access point* (aguardando transmissão) e também sincronizar a transmissão dos dados. Por default, o *beacon* é transmitido a cada 100 milisegundos, mas na maioria dos pontos de acesso é possível especificar qualquer valor entre 10 e 1000 milisegundos [Morimoto, 2008]. Através do *beacon* recebido, a camada PMD mede a intensidade RSS.

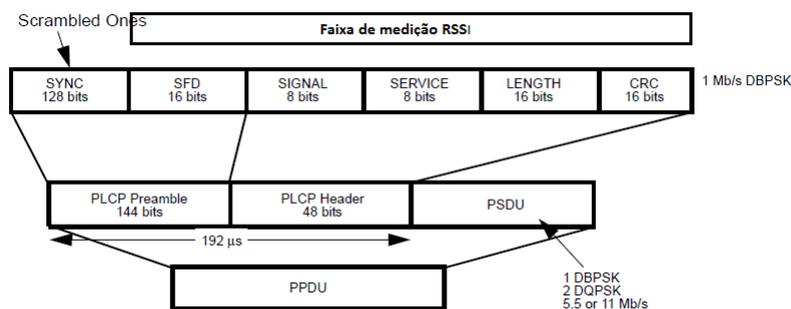
O RSS é um dos parâmetros disponíveis na interface de rede sem fio. Segundo a norma 802.11, este parâmetro pode ser consultado pelas camadas de níveis superiores. O RSS é a medida na camada PMD da quantidade de energia que a antena utiliza para receber a unidade de dados PLCP, segundo a própria norma 802.11 [IEEE, 2012]:

*O RSS é um parâmetro opcional que possui um valor inicial de 0 até RSS\_Max. Esse parâmetro consiste na medida da subcamada PHY da energia que a antena utiliza para receber a PPDU atual. A medida RSS deve ser iniciada entre o início do frame delimitador (SFD) e o fim do cabeçalho da verificação de erro (HEC). O valor RSS deve ser utilizado de forma relativa, uma vez que a precisão absoluta não é especificada.*

A Figura 2.2 mostra a região do quadro onde o sinal é medido.

O valor RSS é representado por 8 bits, sendo 0 o menor nível de energia RF para o circuito receptor. Este nível é chamado de "sensibilidade do receptor" e, normalmente, vem especificado nas interfaces de redes em dBm. O valor de -96dBm é muito frequente em placas de rede. Este valor é obtido pelo conversor analógico-digital integrado ao módulo transmissor - receptor. O valor RSS, medido na subcamada PMD, pode ser obtido pelas camadas de níveis superiores, que utilizamos no trabalho proposto.

Uma das desvantagens de se utilizar esse tipo de sinal é a grande passividade a interferências. Como muitos aparelhos hoje em dia utilizam cada vez mais sistemas de altas frequências, como forno micro-ondas, os sinais RSS provenientes dos AP's podem sofrer interferência de



**Figura 2.2** Faixa de medição do valor RSS no quadro PPDU [IEEE, 2012].

radiofrequência, desse modo influenciando de pequenas a grandes variações nos valores medidos do sinal RSS. Outros problemas causadores de interferência também são fontes de água, pessoas (já que as pessoas tem grande composição de água no corpo), paredes, objetos e outros sistemas que utilizem radiofrequencia.

## 2.2 Técnica de Localização Indoor baseada no Sina RSS

O sistema em questão foi proposto para o projeto de localização interna para dispositivos móveis utilizando rede locais sem fio, visando a localização de equipamentos e pessoas. A proposta é que o método pudesse ser aplicado em diversas áreas e que a sua implementação fosse simples e rápida, não usando de instrumentos especiais, nem análise detalhada do ambiente em termos de sua infraestrutura existente.

A estratégia utilizada foi primeiro fazer um levantamento das intensidades do sinal no ambiente. Este levantamento foi realizado preferencialmente nas áreas de maior incidência de pessoas ou objetos, sugerindo que não tivesse área com raio maior que 5 metros sem existir uma medição. O valor de 5 metros foi escolhido em virtude de ser o objetivo do erro médio que seria alcançado.

A partir da medição, é acrescentado a cada ponto medido, a posição em que se encontra. Esta posição é um valor  $x,y$  que é a distância, em metros, relativa a um ponto de referência a ser escolhido no local a ser monitorado. Assim, é obtido um mapa com a localização de cada ponto bidimensional e a intensidade de sinal de todas as redes existentes. A partir deste conjunto de valores, a estratégia é dividir este conjunto em parte menores, que contenham alguma similaridade com a localização para que, a partir de grupos menores pudesse ser calculado a posição. Para efetuar esta solução foi utilizada a técnica de agrupamento (*clustering*) baseada no algoritmo *K-means*.

Após o processamento dos valores RSS, foi aplicada a técnica de clustering baseada no algoritmo *K-means clustering*. Os valores de RSS são divididos em agrupamentos que possuem uma correlação entre as intensidades dos sinais naquela região. Após os dados serem agrupados da melhor maneira, foi utilizado uma técnica de regressão, para encontrar os valores da localização  $x,y$ . As técnicas de regressão visam estimar funções de valores reais, e para solução desse problema foi utilizado uma rede neural.

Após o treino de cada rede em cada agrupamento, as redes treinadas são armazenadas para serem aplicadas na etapa de localização. Elas fornecem a localização  $x,y$  a partir da entrada dos valores de intensidades dos sinais, em uma área que foi inicialmente treinada.

A Figura 2.3 fornece uma visão de cada etapa deste processo.

### 2.2.1 Parte Off-line

Todo processo da parte Off-line é feito manualmente, desde a preparação e tratamento dos dados, até a execução dos algoritmos sobre os dados. Na técnica em questão todos os itens a seguir (que explicam as técnicas utilizadas) foram executados em uma plataforma na nuvem, o Studio Azure Machine Learning (Ambiente de Aprendizado de Máquina do Azure), e logo após foi salvo todas as configurações para serem utilizadas para consulta da parte On-line.

#### 2.2.1.1 Análise do Ambiente

Primeiramente é utilizado o mapa do local a ser mapeado, enumerando pontos ao longo de todo local, sugerindo que não tenhamos uma área com raio maior que 5 metros sem termos uma medição. Em seguida foi feito um levantamento das intensidades dos sinais do ambiente em relação a cada ponto, este levantamento é realizado preferencialmente nas áreas de maior incidência de pessoas ou objetos. O valor de 5 metros foi escolhido em virtude de ser o objetivo de erro médio a ser alcançado, e que durante a fase de medições gerou um resultado satisfatório. A medição foi feita com dispositivos móveis e com a placa WLAN do SoC Edison da Intel.

A partir desta medição, é acrescentada a cada ponto medido, a posição em que se encontra. Esta posição é um valor  $x,y$  que é a distância, em metros, relativa a um ponto de referência a ser escolhido no local a ser monitorado.

Com a obtenção deste mapeamento, temos o vetor  $p_i$ , que representa as intensidades dos sinais em cada ponto  $i$  e a posição relativa deste ponto que pode ser representada por:

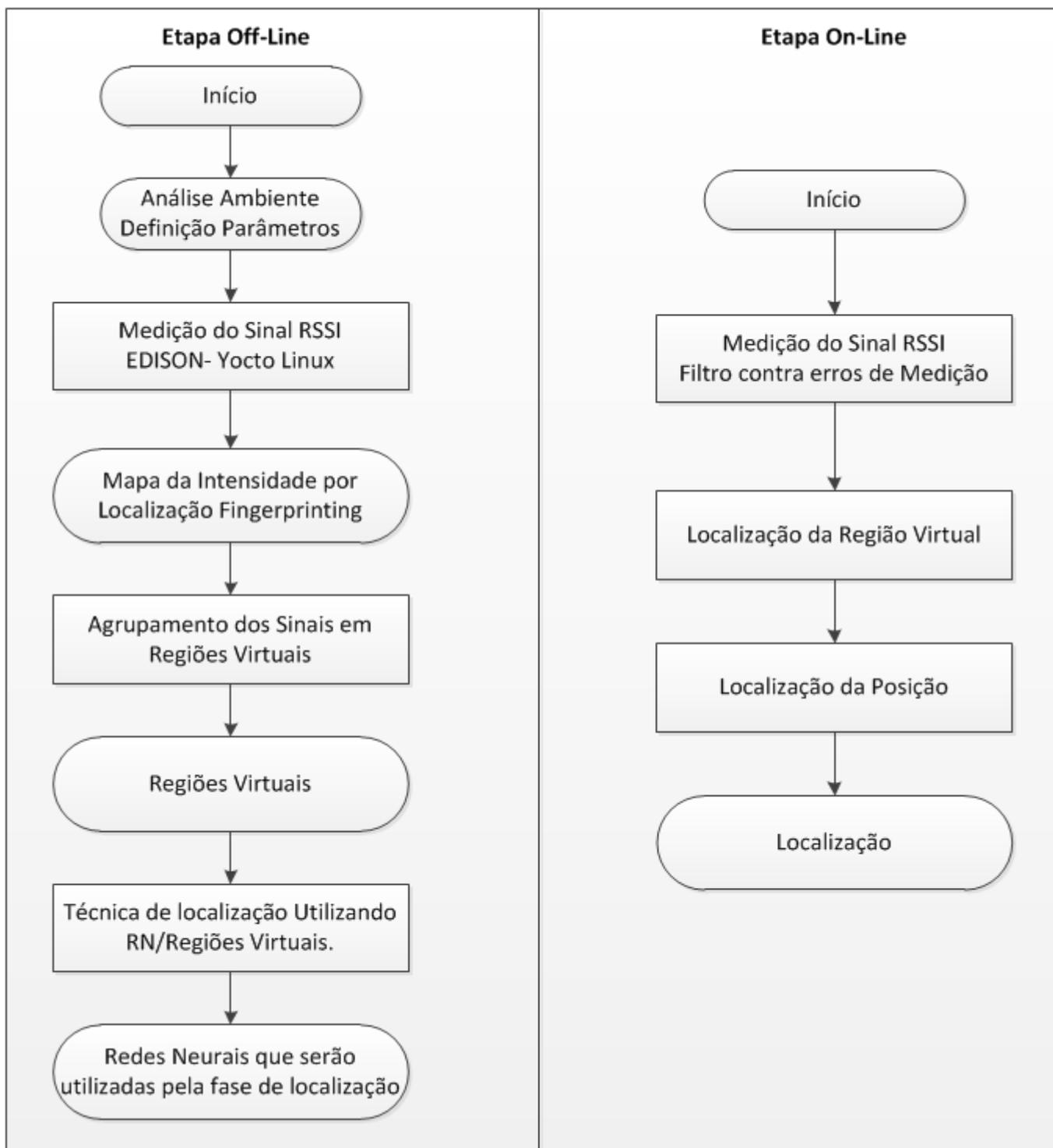
$$p_i = [x_1, x_2, x_3, \dots, x_n, x, y]; \quad (2.1)$$

Onde  $x_n$  é a intensidade do sinal de cada rede  $n$  presente no local, e  $x,y$  é a posição relativa ao ponto de referência escolhido.

Assim, é obtido um mapa com a localização de cada ponto bidimensional e a intensidade de sinal de todas as redes existentes. Salientando que não é necessário conhecer a localização de cada *access point*, pois o objetivo é conhecer as características dos sinais naquele ponto, a “impressão digital”, esta técnica é conhecida por *Fingerprinting*. A técnica *Fingerprinting* vem da palavra *fingerprint* traduzida para o português *impressão digital*, é utilizada em computação, áudio e vídeo. Consiste no procedimento de mapear uma quantidade grande de dados, para um conjunto menor, que identifique um conjunto dos dados originais para algum propósito da aplicação, da mesma maneira que a impressão digital é utilizada para o reconhecimento de uma pessoa.

#### 2.2.1.2 A Medição da intensidade do sinal — RSS

As intensidades dos sinais recebidos — RSS são a base da localização. Se eles forem bem medidos, sem erros, serão obtidos melhores resultados. Vários são os fatores que influenciam a



**Figura 2.3** Sistema utilizando Agrupamento de Dados e Redes Neurais [de Assunção Barros, 2016].

propagação do sinal, como também a sua medição, sendo difícil quantificar os fatores que influenciam na sua propagação e medição [Wysocki and Zepernick, 2000]. Como a técnica proposta no trabalho de Antônio Barros é baseada no mapeamento do ambiente (*Fingerprinting*), serão mostrados os principais fatores dinâmicos que influenciam na medição da intensidade, os quais foram levados em consideração para uma boa medição.

- **Orientação da Antena:** a antena de um equipamento móvel é omnidirecional. Mas, dependendo do equipamento e da posição dela e da posição do sistema embarcado, pode ser que existam obstáculos além da própria composição do edifício.
- **Temperatura, Chuva e Umidade:** esta variável também irá influenciar na propagação do sinal. Porém, como a técnica é utilizada em partes internas, estes fatores não influenciam de modo significativos.
- **Tráfego de Pessoas:** esta é uma das maiores causas de variação do sinal medido. A diferente concentração de pessoas em um ambiente, muito comum em corredores, proporciona uma variação na intensidade do sinal medido.
- **Placa de Rede e Sistema operacional:** a mudança da placa de rede, do fabricante, e até mesmo do sistema operacional irão proporcionar medidas diferentes do valor RSS. Assim, o ideal seria a realização de uma calibração para cada conjunto de equipamento e sistema operacional.
- **Interferências Eletromagnéticas:** os sinais WLAN operam na faixa de frequência de 2.400MHz a 2.500MHz. Os sinais de Wi-Fi que seguem a norma 802.11a e 802.11n operam também na faixa de frequência de 5.000MHz. Diversos equipamentos operam em valores de frequências similares e como seu uso é não controlado, as técnicas baseadas em Wi-Fi sofrem influências de equipamentos, entre os quais se destacam forno de microondas, telefone sem fio, e equipamentos com bluetooth que operam na mesma faixa de frequência de 2.400MHz.

### 2.2.1.3 Regiões Virtuais

Depois da criação do mapa com as intensidades dos sinais e seu pontos de localização, o próximo passo da técnica em questão é dividir este número total de sinais em grupos menores, que serão denominados *Regiões Virtuais*. Através da utilização da técnica de Clustering com o algoritmo K-Means foram obtidos bons resultados, com mais de 90% dos pontos corretamente alocados e os outros 10% dos pontos localizados em área limítrofes entre uma região e outra. A seguir será explicado o funcionamento dos algoritmos de clustering e o K-Means.

Como técnica de agrupamento de dados para gerar as regiões virtuais foi utilizado o algoritmo K-means, que é um dos mais conhecidos e utilizados algoritmos de Clustering. A ideia básica do algoritmo K-Means é apresentado no fluxograma 3.6 apresentado a seguir.

1. Primeiro é determinado o número de centroides.
2. Estes centroides são inicialmente atribuídos de forma aleatória entre k elementos do conjunto.

3. A distância de cada objeto a cada centroide é calculada.
4. Cada objeto é atribuído ao agrupamento que possui o centroide mais próximo.
5. Houve algum elemento que mudou de agrupamento? Se sim, os centroides de cada agrupamento são recalculados e retorna ao passo 3. Senão, termina.

Este algoritmo é extremamente veloz, geralmente convergindo em poucas iterações para uma configuração estável, na qual nenhum elemento está designado para um cluster cujo centro não lhe seja o mais próximo.

O algoritmo K-means depende do parâmetro  $K$  que define o número de agrupamentos sendo um parâmetro de entrada. Isto costuma ser um problema, tendo em vista que normalmente não se sabe quantos agrupamentos existem ou seriam ideais para o conjunto de sinais amostrados. Na técnica em questão foi utilizada uma faixa de valores, e o algoritmo K-Means busca o valor de  $K$  dentro da faixa, para o qual foi obtido o melhor resultado.

Esta faixa de valores para o número de agrupamentos representa os setores onde existe uma similaridade maior entre os valores RSS's e uma separação maior dos outros valores. Para ser encontrado o melhor valor de  $K$ , dentro da faixa estabelecida, foi utilizada a técnica denominada **silhueta** (*silhouette*).

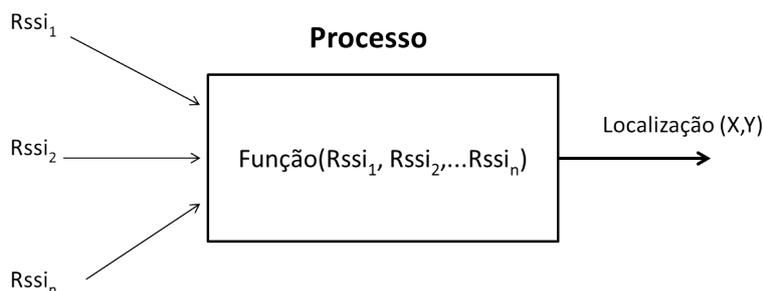
*Silhouette* é um método de validação e interpretação de um agrupamento de dados. É um método que representa graficamente cada agrupamento, ou como cada objeto se localiza dentro de um agrupamento. Este método foi proposto por Rousseeuw em 1987 [Rousseeuw, 1987]. A média das larguras das *silhouettes* fornece a qualidade de um agrupamento e pode ser usada para avaliar a quantidade de agrupamentos que é mais adequada a um conjunto de amostras.

Esta técnica foi usada no método de localização proposto a fim de determinar qual a melhor quantidade de agrupamentos de acordo com o conjunto de dados. Quais valores dos sinais RSS têm maiores semelhanças entre si, e maiores dissimilaridades com os valores de outros agrupamentos. A estes agrupamentos chamaremos de *Regiões Virtuais*, que indicam os agrupamentos onde estão as medidas similares. Estas regiões facilitam a transformação de uma área de grandes dimensões em várias áreas menores de modo a ser possível utilizar o método de regressão para determinar com melhor precisão onde uma medida RSS foi realizada, ou onde está o objeto ou pessoa a ser localizada.

#### 2.2.1.4 Treinamento com Redes Neurais

Uma vez que as *Regiões Virtuais* foram encontradas, pode-se afirmar que aqueles elementos de uma mesma região possuem uma semelhança entre os seus sinais RSS's e alguma diferença entre os sinais RSS's das outras regiões. Pela análise dos resultados e pelo o conhecimento das propriedades de propagação de sinais, pode-se afirmar que estes pontos estão próximos fisicamente. O próximo problema a ser resolvido consiste em localizar um ponto a partir dos pontos existentes e com base nas características da intensidade do sinal — RSS.

A técnica em questão baseou-se em técnicas de regressão, que tentam modelar o comportamento da localização em função das intensidades dos sinais RSS's medidos inicialmente. A técnica em questão fez a predição da localização em função dos novos valores RSS's medidos, a partir da base inicial de mapeamento.



**Figura 2.4** Representação do Processo de Regressão.

Este processo está representado na Figura 2.4.

A utilização de redes neurais em um sistema de localização foi proposto pela primeira vez por Battiti et al. [Battiti et al., 2002]. Eles identificaram que devido à não linearidade do comportamento do sinal em relação à distância e também devido ao grande número de parâmetros que influenciam o sinal, uma rede neural seria uma ótima ferramenta para esta aplicação. Eles citam que, neste caso, a distância e a localização do access point não tem importância e que utilizando uma rede neural, a partir de um conjunto de exemplos rotulados, pode-se construir um modelo que vai generalizar de forma adequada.

Na técnica de localização em questão foi utilizada uma rede neural para regressão em cada cluster, com uma camada oculta, o número de nós de entrada corresponde à quantidade de APs presentes no ambiente, número de nós na saída são dois e corresponde à posição  $x, y$ . Quanto ao número de nós na camada oculta, foram realizadas diversas simulações, para determinar a quantidade ideal, aquela que produz um menor erro, fixando o máximo em 100 nós.

Na técnica em questão, após as redes de cada região virtual serem treinadas, isto é, os pesos de cada neurônio terem sido ajustados e os erros minimizados, estas são armazenadas para serem utilizadas pela técnica de localização.

### 2.2.2 Parte On-line

Na parte On-line foram utilizados novamente alguns dispositivos móveis e a placa WLAN do SoC Edison da Intel para medir o sinal RSS de um ponto qualquer e assim manualmente tratar os valores medidos e logo após fazer a consulta no banco de dados que foi salvo na parte Off-line no Azure Machine Learning.

#### 2.2.2.1 Localização

Através de um aplicativo que foi chamado de *Localiza*, foi feita a consulta ao banco de dados das regiões virtuais, que foi gerado pela técnica baseada no K-means, identificando a que região este conjunto de valores estava situado.

A informação da região virtual e os valores medidos são repassados para o método de localização baseado em Rede Neural que consulta a rede neural correspondente a região virtual retornando a posição do objeto. Esta posição é informada em valores de coordenada  $x, y$  nor-

malmente em metros, e é relativa a algum pondo do edifício em que foi treinado e baseado no mesmo modo do treinamento.

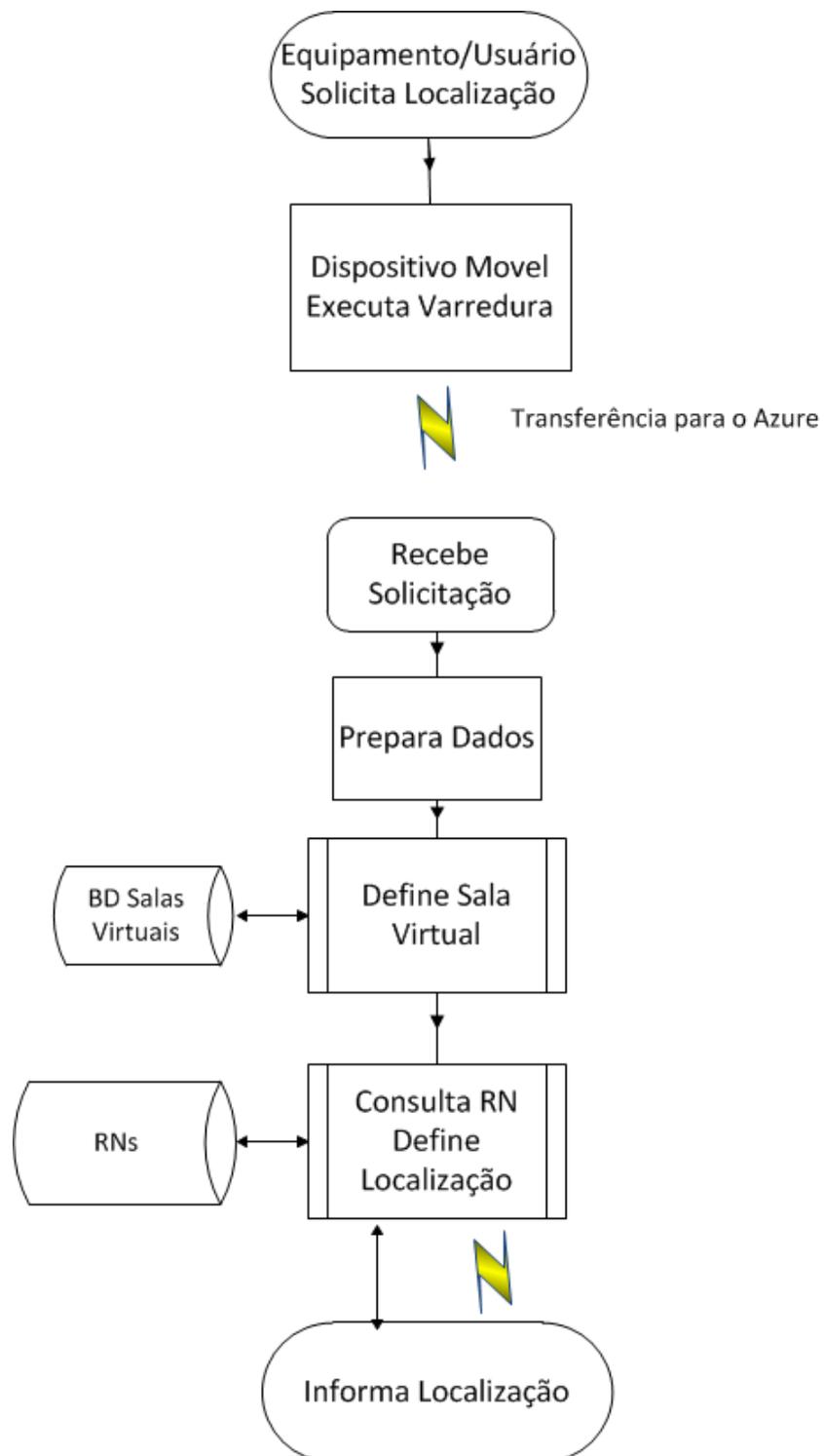
A Figura 2.5 fornece uma visão de cada etapa deste processo. Onde grande parte do processo para identificar a localização foi feita de forma manual.

## 2.3 Principais Problemas da Implementação de Antônio Barros

Mediante todo o trabalho de Antônio Barros, foi identificado alguns problemas em sua implementação:

- **Processo manual de preparação dos dados:** O processo de preparação dos dados é um processo que requer bastante atenção para que não haja problemas em relação a identificação dos dados, erro de valores, inconsistência na integração e até mesmo na transformação dos dados, fazendo com que assim esses problemas passem despercebidos e venham alterar o resultado do experimento.
- **Ambiente de Aprendizagem de Máquina em Nuvem:** Toda a parte Off-line foi estruturada no ambiente Azure Machine Learning, então para poder realizar o experimento era necessário uma conta nesse ambiente da Microsoft, o qual não suporta várias requisições simultâneas em uma mesma conta pra consultas, dificultando assim o processo de localização de vários objetos.
- **Tempo de execução:** Desde a preparação dos dados até a forma de coloca-los como entrada no ambiente Azure, requer um tempo elevado, que pode ser medido na faixa de 30 minutos a 1 hora para a parte Off-line e de 10 a 15 minutos para ter um resultado para localização. Isso é muito significativo para um tempo de resposta que se aproxima de um sistema de tempo real, e para implementação do sistema no mundo real.
- **Conexão com a internet:** Todo o sistema proposto para comunicação da parte Off-line com a parte On-line depende de uma conexão com a internet, para que se possa levar os dados até suas respectivas entradas. Isso dificulta o processo de localização, pois caso a internet vem a falhar o sistema de localização irá parar e o objeto não conseguirá se localizar, podendo até falhar todo o sistema por inconsistência e perda dos dados.
- **Interface de comunicação com o usuário:** O sistema de localização não possui uma interface de comunicação direta com o usuário, sendo todas as informações sendo acessíveis apenas ao operador do sistema, que faz todo o processo de forma manual.

Dessa forma, o presente trabalho propôs uma melhoria significativa nos problemas identificados acima. No próximo capítulo mostraremos o nosso sistema proposto com o intuito de automatizar todo o processo até aqui explicado, aplicando as melhorias nos problemas identificados no trabalho de Antônio Barros.



**Figura 2.5** Processo de Localização [de Assunção Barros, 2016].

## 3 Proposta de Implementação Automatizada e Local para o Método de Barros

Neste capítulo será feita a descrição da proposta de implementação em servidor local e baseada em um sistema embarcado que visa uma maior automação e precisão do sistema de localização proposto por Barros [de Assunção Barros, 2016]. No início será apresentada uma visão geral da implementação proposta, em seguida serão detalhadas as suas etapas, a implementação da eletrônica embarcada para captura dos sinais, seguida pela descrição da implementação dos módulos de treinamento Off-line e de localização On-line no servidor local.

### 3.1 Visão Geral da Técnica Implementada

A implementação realizada neste trabalho teve como principal objetivo uma maior automação com implementação baseada em servidor local e um sistema embarcado do método proposto no trabalho de Antônio Barros [de Assunção Barros, 2016]. A implementação realizada neste trabalho inclui duas partes fundamentais para automatização do sistema de localização em questão:

- **Sistema Embarcado Móvel para localização;**
- **Servidor local para execução dos algoritmos de aprendizagem de máquina e consultas para localização.**

A implementação proposta também utilizou um protocolo de "comunicação" entre o sistema embarcado e o servidor local mais seguro, onde ambos trocam informações através de requisições SSH (*Secure Shell*), que é um protocolo de rede criptográfico para operação de serviços de rede de forma segura sobre uma rede insegura [Systems, 2006]. A melhor aplicação de exemplo conhecida é para login remoto a sistemas de computadores pelos usuários. O SSH fornece um canal seguro sobre uma rede insegura em uma arquitetura cliente-servidor, conectando uma aplicação cliente SSH com um servidor SSH.

Dessa o protocolo baseado em ssh permite o sistema embarcado se comunicar com o servidor local durante todo o processo, seja na parte Off-line como na parte On-line como será detalhado mais adiante.

A figura 3.1 mostra uma visão geral da implementação proposta.

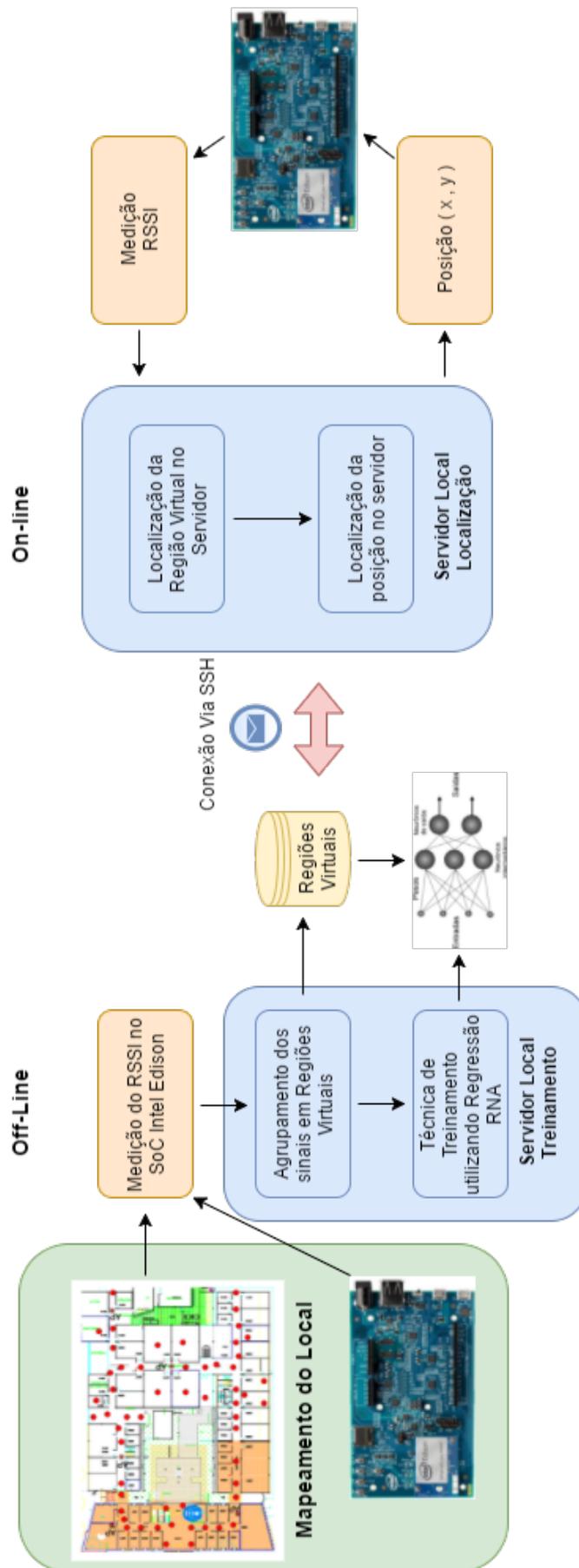


Figura 3.1 Método proposto automatizado, Parte Off-line e On-line.



**Figura 3.2** Vista do Usuário, Vista de Cima e Vista Frontal do capacete.

### 3.1.1 Sistema Embarcado Móvel

Para que o método de localização em questão fosse automatizado de forma a se ter uma medição dos sinais RSS mais precisa e eficiente, bem como uma melhor comunicação com o servidor local e uma boa integração dos dados, foi desenvolvido um dispositivo móvel baseado no SoC Edison da Intel.

Esse sistema embarcado é composto pelos seguintes componentes:

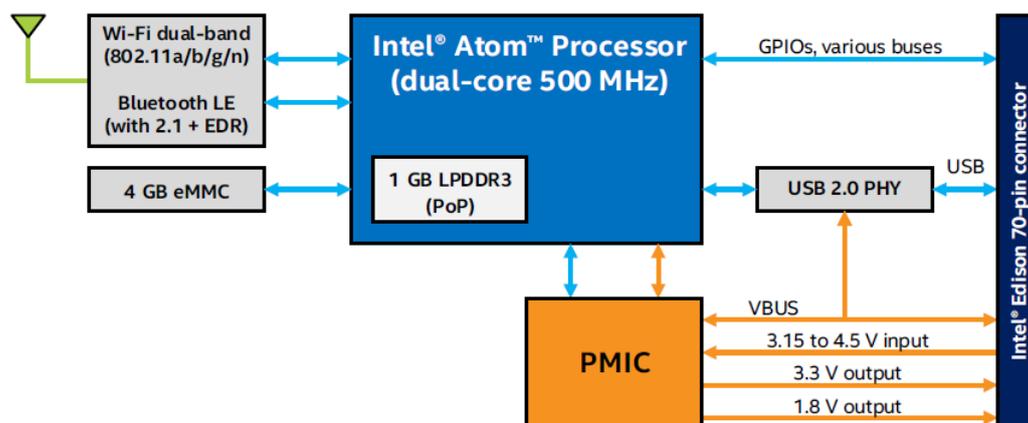
- **Um capacete onde foi montado a estrutura do sistema embarcado;**
- **Um SoC Intel Edison com interface para arduino.**
- **Um Led verde.**
- **Um Display LCD.**
- **Um Botão.**
- **Uma Antena externa.**

Foi utilizado um capacete para facilitar as medições dos sinais RSS, pois enquanto era feito o mapeamento do local as mãos estariam livres para que se pudesse marcar os pontos de medição no mapa do local. Assim garantindo mais precisão da distribuição dos pontos em todo ambiente.

No capacete foram integrados a placa com interface para arduino já com o SoC Edison fixado na placa, o led verde, o display LCD, o botão e uma antena externa. Conforme mostra a figura 3.2.

Os sinais foram medidos com uma placa baseada no processador Edison. O processador Edison da Intel® é um processador embarcado desenvolvido para dispositivos portáteis apropriado para aplicações em "Internet das coisas".

O processador Edison é um SoC (*System on Chip*) da Intel fabricado na tecnologia de 22 nm que inclui um processador Atom™ dual-core, dual-threaded que opera a 500 MHz e um micro-controlador Intel® Quark™ de 32-bits de 100 MHz. A memória RAM é de 1 GB LPDDR3, e



**Figura 3.3** Diagrama de blocos do Processador Edison[Intel, 2015].

possui 4GB de memória flash. O processador possui uma interface WLAN Broadcom\* 43340 802.11 a/b/g/n com Dual-band (2,4 e 5 GHz) e a sua antena é integrada. A interface Bluetooth também presente é 4.0. Este processador foi lançado durante a International Consumer Electronics Show (CES) em 2014.

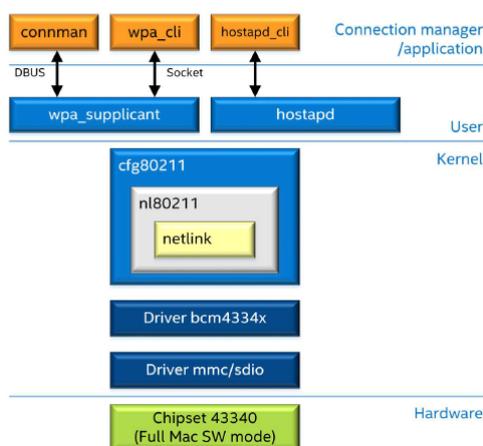
A interface do processador está descrita a seguir.

- 20 pinos de entrada ou saída digital, sendo 4 deles com recurso PWM (*Pulse Width Modulation*) Modulação por largura de pulso.
- 6 entradas digitais.
- 1 UART- Rx/Tx (*Universal Asynchronous Receiver/Transmitter*) Transmissão e Recepção Assíncrona Universal.
- 1 I2C - Inter-Integrated Circuit - Circuito Inter-integrado.
- 1 ICSP 6-pin header (SPI - *Serial Peripheral Interface*).
- 1 Conector Micro USB.
- Conector *SD Card*.
- Conector para fonte de alimentação 7 - 15V.

O sistema operacional utilizado foi o Yocto Linux 1.6. Seu ambiente de desenvolvimento é compatível com várias IDEs incluindo o Arduino\* IDE; o framework Eclipse: C, C++, e Python; a IDE Intel XDK: Node.JS e HTML5.

As dimensões do processador são bastante reduzidas dadas por 35,5 × 25,0 × 3,9 mm. A Figura 3.3 mostra o diagrama de blocos do processador Edison, onde se observa estas características, com destaque à placa WLAN que foi utilizada para a coleta de dados.

Para o acesso às informações dos valores RSS, o processador Edison possui o módulo wpa-suplicant que permite a conexão direta com o driver BCN4334x da Broadcom, sua placa de



**Figura 3.4** Arquitetura de Software da Placa de Rede do Edison[Intel, 2015].

rede embarcada e que fornece os valores RSS. Devido ao fato de ser um processador destinado a aplicações embarcadas, também são disponibilizados pela Intel os arquivos fontes dos drivers e softwares da placa de rede Broadcom. Conforme a documentação do Edison as localizações dos arquivos estão nos diretórios indicados a seguir [Intel, 2015].

- `broadcom_cws/wlan/driver_bcm43x`: Arquivos fonte para Broadcom Linux Driver.
- `wpa_supplicant / wpa_cli`: arquivos fontes para `wpa_supplicant / wpa_cli 2.1`
- `linux-kernel/net/wireless`: Arquivos fontes para modelo Mac definido em `mac80211`

Na Figura 3.4 são mostrados os módulos dos componentes da placa de rede Broadcom do Edison. Pelo módulo Wpa-suplicant pode ser feito o acesso a diversos parâmetros e valores da placa de rede. Uma lista destes comandos é disponibilizada na documentação [Die.net, 2016].

O SoC faz o controle de todo sistema embarcado, onde cada um tem uma funcionalidade, que é descrita a seguir:

- **Led verde**: É acionado toda vez que acaba de fazer uma medição, tanto na parte Off-line na hora do mapeamento como na parte On-line, quando é necessário localizar um ponto qualquer.
- **Display LCD**: É utilizado na parte On-line, informando ao usuário a posição  $x,y$  em que se encontra.
- **Botão**: É utilizado para informar ao SoC quando o usuário quer fazer a medição em um determinado ponto. Basta que o usuário pressione uma vez o botão e assim o SoC faz a medição.
- **Antena externa**: O próprio SoC tem uma antena interna WLAN, mas foi acrescentada uma antena externa para melhorar a qualidade da medição do sinal.

O sistema embarcado é usado na fase de mapeamento, medição dos sinais, localização e interface com o usuário. Os valores medidos pelo sistema de embarcado são enviadas ao servidor local através de requisições SSH.

### 3.1.2 Servidor Local

No servidor local são executadas as partes offline e online da técnica proposta, para isto foram implementados os algoritmos de aprendizagem, os algoritmos de treinamento e todo o sistema de controle da implementação proposta.

Foi alocado um servidor que se encontra em funcionamento na central de suporte do Centro de Informática e que possui a seguinte configuração:

- **Processador:** Intel(R) Xeon(R) CPU E5620 @2.40GHz - Quad Core - 64 bits
- **Memória Ram:** 6 Gigabytes
- **Armazenamento:** 90 Gigabytes
- **Sistema Operacional:** Ubuntu 16.04.1 LTS 64 bits

Foi implementado um sistema central no servidor que faz o controle de todo o processo. Na parte Off-line depois que o mapeamento é feito usando o sistema embarcado, os dados são enviados para o servidor via requisição SSH. Quando o sistema central recebe os dados ele os envia para um subsistema que faz o tratamento dos dados. Após esta fase, os dados enviados para outro subsistema que tem o algoritmo K-means implementado, onde agora o processo de agrupamento dos dados é realizado resultando os dados agrupados da melhor forma possível, através da técnica de validação silhueta. Assim são montadas a estrutura de regiões virtuais que fica armazenada em um banco de dados.

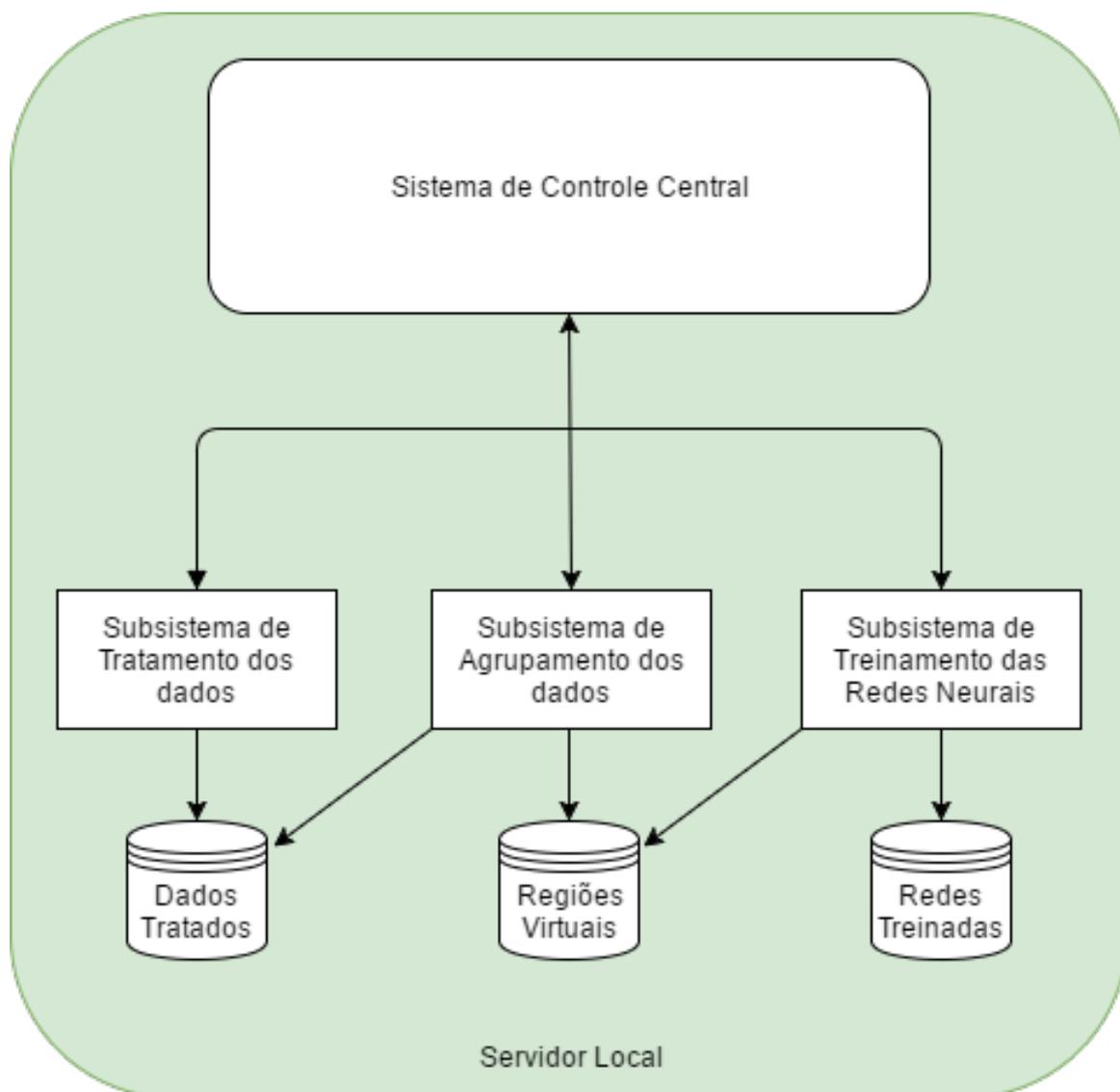
No passo a seguir, o sistema central associa para cada região virtual uma rede neural do tipo MLP (*Multilayer Perceptron*), que são treinadas com seus respectivos pontos e parâmetros. Depois do treinamento as informações de cada rede são salvas em um banco de dados. O sistema central fica, então, a espera agora de alguma solicitação de localização.

A figura 3.5 mostra a configuração do servidor local.

## 3.2 Módulo de Medição da Intensidade do Sinal RSS

O primeiro passo para otimização da técnica de localização é a análise do ambiente no qual será feita a medição dos sinais. Para isso foram verificados os seguintes itens:

- **Cobertura Wi-Fi:** a área em que o sistema de localização será utilizado deve ser coberta com pelo menos três sinais de APs diferentes em cada ponto a ser monitorado. Cada ponto da medição tem sinais com intensidade maior que -75dBm, a fim de melhorar a exatidão da medida.



**Figura 3.5** Estrutura do Servidor Local.

- **Mapa do Local:** com uma planta baixa do local a ser monitorado será possível determinar a área a ser abrangida e melhor determinar os pontos de medição.
- **Fatores de Interferência:** como citado anteriormente, alguns fatores têm grande interferência nas medições e devem ser analisados ou pelo menos conhecidos, pois são dinâmicos e irão afetar bastante os valores das amostras medidos no mapeamento e a utilização no dia a dia.
- **Potência do sinal constante:** é muito importante que a potência do sinal emitido pelo AP seja constante.
- **Alterações no fluxo de pessoas:** este fator merece também ser analisado, já que este é o principal propósito da melhoria da técnica proposta por Antônio Barros, pois mediante o fluxo de pessoas os dados terão tratamentos diferentes.

Como o método proposto é baseado no trabalho de Barros, foram considerados os principais fatores dinâmicos que influenciam na medição da intensidade dos sinais, os quais serão descritos a seguir.

- **Orientação da Antena:** a antena externa utilizada no capacete tem uma precisão melhor do que antena do SoC Edison. A antena é um equipamento omnidirecional, mas dependendo do equipamento e da posição dele e da posição de quem está operando, pode ser que existam obstáculos além da própria composição do edifício. Deve ser salientado que o corpo humano é praticamente feito de água e este é um grande obstáculo à propagação do sinal. Por isso que a antena foi colocada em cima do capacete, isto é, na parte mais alta como mostra a figura 3.2, driblando assim possíveis interferências como o fluxo de pessoas.
- **Temperatura, Chuva e Umidade:** Como houveram algumas medições na parte externa do local, foram levados em consideração esses fatores no tratamento dos dados.
- **Tráfego de Pessoas:** esta é uma das maiores causas de variação do sinal medido. Por isso durante as medições foi levado em consideração no subsistema de tratamento dos dados a quantidade de pessoas no local.
- **Placa de Rede e Sistema operacional:** Foram verificados todos os AP's que realmente funcionam no local medido, desconsiderando AP's dinâmicos que poderiam atrapalhar as medições.
- **Interferências Eletromagnéticas:** Não foram feitas medições próximas a aparelhos que pudessem causar interferências eletromagnéticas, prejudicando o sinal medido.

Todos os itens mencionados foram verificados durante as medições, bem como no ajuste de alguns parâmetros do módulo de medição de Intensidade do sinal RSS e nas variáveis no subsistema de tratamento de dados no servidor.

### 3.2.1 Módulo de Leitura do Sinal

O módulo de leitura do sinal é executado no sistema embarcado e foi implementado conforme o pseudo código descrito a seguir:

```

inicio ;
variaveis :
    Porta de entrada do botao ;
    Porta de entrada do display ;
    Porta de saida do LED ;

Faca sempre :

    Cria arquivo que gravara valores da medicaao ;
    Espera botao ser apertado ;
    Se o Botao foi apertado :
        Faz a medicaao de todos os sinais daquele ponto ;
        Acende LED indicando medicaao em andamento ;
        Espera 8 segundos ; // Tempo pra varredura completa ;

    Apaga o LED para informar que a medicaao acabou ;
    Fecha o arquivo ;

    Faz uma requisicao SSH enviando arquivo para o Servidor Local ;

```

### 3.2.2 Módulo de Tratamento dos Dados

Como informado anteriormente, o módulo de tratamento dos dados é um subsistema que aguarda ordem do sistema de controle do servidor para tratar dos dados recebidos. Após o tratamento, os dados são armazenados em um banco de dados. O subsistema de tratamento dos dados foi implementado em linguagem C conforme o pseudo código descrito a seguir:

```

inicio ;
variaveis :
    Solicitacao do Sistema de Controle para tratar 1 sinal ;
    Solicitacao do Sistema de Controle para gerar entrada K-means ;

Se recebeu requisicao para tratar 1 sinal , faca :

    Pegue apenas os macs adress e intensidades do sinais ;
    Poe numa lista encadeada os novos macs adress ;
    Se for o primeiro sinal do mapeamento , faca :
        Cria tabela de pontos por mac adress ;
        Insere primeiro ponto na tabela ;
        Coloca as intensidades do sinal para cada mac adress ;
    Se n o for , faca :
        Insere ponto na tabela ;

```

```

Coloca as intensidades do sinal para cada mac adress;
Varre tabela e bota zero para os macs sem intensidades;

```

Se nao recebeu , faça :

```

Espera solicitacao do Sistema de Controle para tratar 1 sinal;

```

Se recebeu requisicao para tratar entrada K-means , faça :

```

Varra a tabela e exclua mac adress que nao tiveram mais
de 3 pontos;
Normalize as intensidades dos sinais;
Coloque zero nos mac adress em que as intensidades
foram abaixo de 35;
Envia tabela para entrada do K-means para o sistema
central;

```

### 3.2.3 Módulo de Comunicação com o Servidor

Para viabilizar a comunicação de dados tanto do lado do servidor como do lado do sistema embarcado, foi implementado o subsistema de comunicação baseado em requisições SSH e arquivos. Cada vez que o SoC Edison quer enviar algo para o servidor, o sistema do lado do servidor verifica um arquivo que fica salvo em uma pasta, notificando que o SoC Edison enviou algo pra ele. Do mesmo modo acontece quando o servidor quer enviar algo para o SoC Edison.

As requisições SSH são chamadas pelos sistemas operacionais executando tanto do lado do servidor como do lado do sistema embarcado. Elas são chamadas por linhas de comandos como detalhado a seguir:

```
sshpass -p 'senha' scp 'caminho saida arquivo' 'endereço ip': 'caminho destino arquivo'
```

A senha que vai na requisição é a senha do destinatário, pois ambos os sistemas operacionais tanto do lado servidor como do SoC Edison são protegidos por senha.

O caminho do arquivo de saída tem que ser bem especificado, senão na hora da requisição a comunicação falha, assim como o caminho de destino do arquivo.

O endereço IP é o endereço do destinatário, para onde o arquivo deve ir.

Todas essas configurações foram consideradas na implementação do subsistema de comunicação no servidor como no sistema embarcado. Assim todas os parametros são bem definidos em ambos os sistemas.

## 3.3 Módulo de Treinamento Off-line

Depois que o mapeamento do local foi feito com a medição da intensidade dos sinais em seus pontos de localização, o próximo passo é dividir este número total de sinais em grupos menores, que serão denominados *Regiões Virtuais*. A *Regiões Virtuais* são obtidas através da implementação do algoritmo K-means que será descrito a seguir.

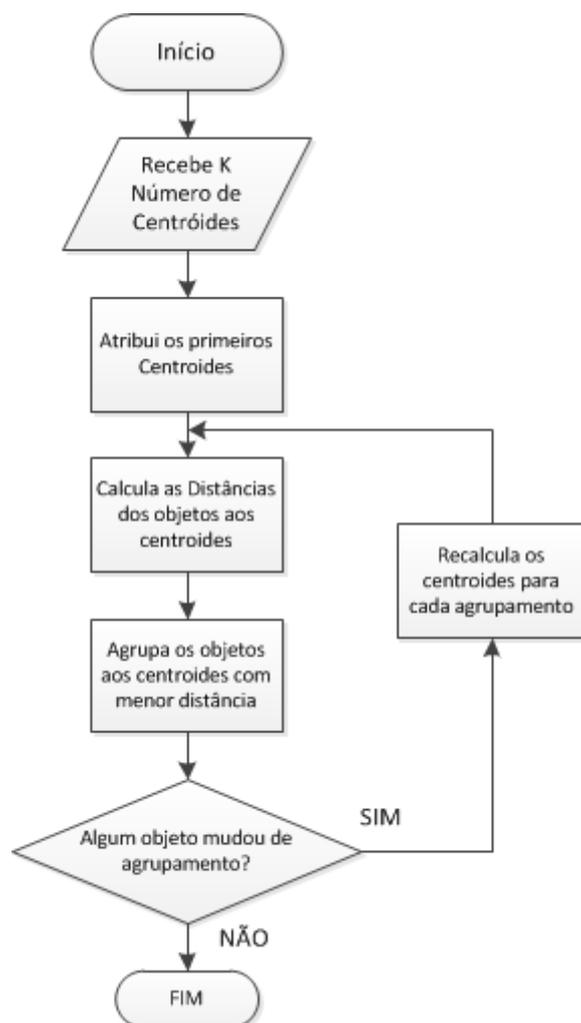


Figura 3.6 Fluxograma Básico do Método K-Means.

### 3.3.1 Algoritmo K-means

O algoritmo K-means é um dos mais conhecidos e utilizados algoritmos de Clustering. Em português este algoritmo é chamado de K-média, e consiste em minimizar a distância de cada ponto de um agrupamento ao centro do agrupamento e estes centros, denominado de centróides de forma iterativa. O algoritmo K-Means foi implementado de acordo com fluxograma 3.6.

O algoritmo K-means depende do parâmetro  $K$  que define o número de agrupamentos sendo um parâmetro de entrada. Isto costuma ser um problema, tendo em vista que normalmente não se sabe quantos agrupamentos existem ou seriam ideais para o conjunto de sinais amostrados. Na técnica proposta foi utilizada uma faixa de valores, e o algoritmo K-Means busca o valor de  $K$  dentro da faixa, para o qual foi obtido o melhor resultado.

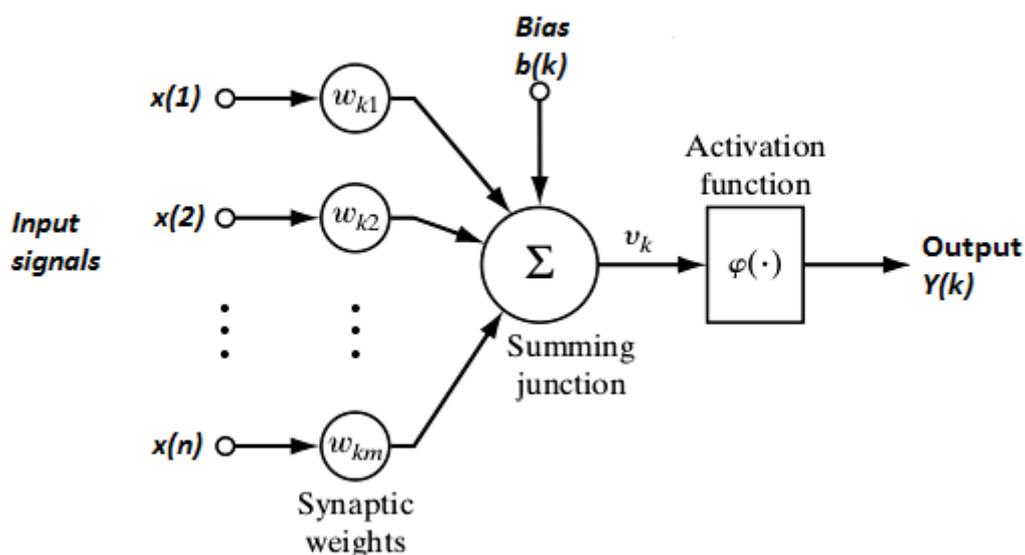
Para validação dos agrupamentos foi utilizado e implementado também o método de validação silhueta. Esta técnica também foi usada para determinar qual a melhor quantidade de agrupamentos de acordo com o conjunto de dados. Foi criado no servidor local um subsistema

com a implementação do K-means em linguagem C.

Uma vez que as *Regiões Virtuais* foram encontradas, pode-se afirmar que aqueles elementos de uma mesma região possuem uma semelhança entre os seus sinais RSS's e alguma diferença entre os sinais RSS's das outras regiões. Pela análise dos resultados e pelo conhecimento das propriedades de propagação de sinais, pode-se afirmar que estes pontos estão próximos fisicamente. O próximo problema a ser resolvido consiste em localizar um ponto a partir dos pontos existentes e com base nas características da intensidade do sinal RSS. Por isso foi criado uma rede neural para cada agrupamento de dados.

### 3.3.2 Rede Neural MLP — *Multilayer Perceptron*

As redes neurais artificiais são compostas por um conjunto de neurônio artificiais. O neurônio artificial é uma aproximação elementar do neurônio biológico [Vieira and Roisenberg, 2003] como mostrado na Figura 3.7.



**Figura 3.7** Modelo Neurônio Artificial [Neural Network in a Nutshell, 2016].

Neste modelo de rede, as entradas do neurônio  $x_k$  podem ser saídas de outros neurônios ou entradas externas ou ainda um bias, que é um parâmetro externo ou uma constante, podendo ser também qualquer combinação destes elementos. Estas entradas são ponderadas pelos pesos  $w_{ik}$  que são inspirados na força da conexão sináptica. Desta forma, temos a saída do neurônio que é geralmente o somatório de todas as entradas multiplicadas pelos seus respectivos pesos somado ao valor do bias e aplicada a função de saída,  $\varphi$ . A equação da função de saída está descrita abaixo.

$$Y(k) = \varphi \left( \sum_{k=1}^n x_k W_{ik} + b_k \right) \quad (3.1)$$

Onde  $w_{ik}$  representa o peso de cada conexão sináptica no neurônio  $k$ . A função de saída pode ser uma função de ativação  $\varphi(\cdot)$  associada também a outras funções e normalmente tem forma contínua e crescente, de tal sorte que seu domínio geralmente se encontra no âmbito dos números reais. Geralmente utiliza-se como função de saída as funções lineares, sigmoideal ou logística e a função tangente hiperbólica. No nosso modelo de rede foi utilizada a função tangente hiperbólica, por ter uma convergência e um resultado melhor do que as outras funções.

Na implementação proposta, após as redes de cada região virtual serem treinadas, isto é, os pesos de cada neurônio terem sido ajustados e o erros minimizados, estas são armazenadas para serem utilizadas pela técnica de localização na parte On-line.

Neste trabalho foi utilizada a biblioteca de Redes Neurais chamada Scikit Learn [Scikit-Learn, 2016] conhecida e bem difundida e escrita em linguagem de programação Python. Assim foi feita a integração do subsistema central com as bibliotecas em Python para criação e treinamento das redes neurais MLP's.

### 3.4 Módulo de Localização On-line

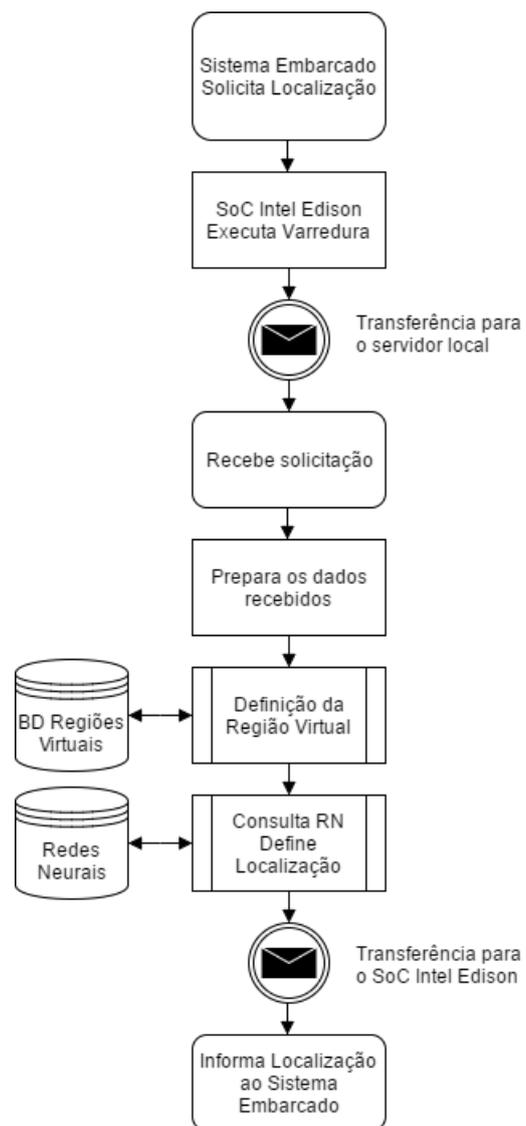
O processo de localização é descrito no fluxograma apresentado na Figura 3.8. Ele tem início com a aquisição dos dados de intensidade do sinal Wi-Fi na posição a ser localizada.

O usuário do sistema embarcado pode solicitar ao sistema a localização de um ponto (local em que ele está) simplesmente apertando o botão. Assim após ser concluída a varredura das intensidades dos sinais, o sistema embarcado faz uma requisição SSH para o servidor enviando um arquivo com as intensidades do ponto que deseja se localizar.

Os valores das intensidades dos sinais RSS são transferidos, para o servidor local. O sistema de controle recebe a informação, e faz a preparação dos dados. Esta preparação consiste no envio das intensidades dos sinais para o subsistema de tratamentos dos dados, que vai fazer todo o tratamento da mesma forma que na fase de mapeamento, para que os sinais da localização tenha o mesmo tratamento dos sinais de treinamento, assim não havendo divergência nos parâmetros utilizados e garantindo uma boa localização.

Em seguida o sistema de controle faz uma consulta ao banco de dados das regiões virtuais e verifica em qual o agrupamento melhor se encaixa o ponto da localização em questão. Feito isso, o sistema de controle envia os valores das intensidades do ponto para rede neural MLP do agrupamento no qual ele foi alocado. A rede neural retorna, então, como saída a posição  $x,y$  do ponto em questão.

Dessa forma, novamente o sistema de controle do servidor faz uma requisição SSH para o SoC Edison, enviando um arquivo com as posições  $x,y$  do ponto solicitado. Então é mostrado ao usuário no display LCD a localização  $x,y$  que foram calculados na rede neural.



**Figura 3.8** Processo de Localização.

## 4 Experimentos e Resultados

Este capítulo descreve o ambiente, os equipamentos e as ferramentas utilizadas para realizar os experimentos a fim de validar o método implementado neste trabalho e apresentado no capítulo anterior.

### 4.1 Ambiente e Equipamentos Utilizados

O objetivo desta seção é descrever os detalhes do ambiente do CIn — Centro de Informática da UFPE, do sistema embarcado com o SoC Edison e do Servidor Local.

#### 4.1.1 Descrição do Ambiente do CIn

O ambiente, onde foram realizados os testes, foi o Centro de Informática da UFPE. Este local foi escolhido por ser uma área pública e típica na qual o sistema poderia ser implantado, bem como onde a técnica proposta por Barros [de Assunção Barros, 2016] foi desenvolvida.

O Centro de Informática ocupa uma área de 3.522 m<sup>2</sup> com um perímetro de 240 metros. O Centro possui um corredor central que dá acesso às salas de aula, às sala de professores, aos laboratórios e ao auditório; bem como um corredor secundário. As medições foram realizadas com maior frequência nos corredores, e em algumas salas e laboratórios. As paredes são de alvenaria ou concreto e os APs estão instalados quase na sua totalidade no teto a uma altura aproximada de 3 metros. Vários APs possuem mais de uma rede, sendo todos os sinais de um mesmo AP, o qual foi levado em consideração na etapa de medição. Também foram utilizadas as redes de 2,4GHz e de 5GHz.

Na coleta de dados para o mapeamento das regiões virtuais, os sinais RSS foram medidos em 100 pontos diferentes e distribuídos em todos os corredores, nas duas entradas e em 8 salas.

A Figura 4.1 mostra a planta baixa do Centro de Informática. Na planta estão assinalados os pontos nas áreas e salas onde foram realizadas as medidas para o mapeamento das regiões virtuais.

#### 4.1.2 Escolha do parâmetros

Os parâmetros utilizados para a implementação do sistema de localização serão descritos nesta seção. De acordo com a Figura 3.1, existem três grupos de parâmetros: os parâmetros de medição da intensidade do sinal — RSS na etapa de mapeamento do local; os parâmetros para definição das regiões virtuais e, por último, os parâmetros de configuração e treinamento das redes neurais.

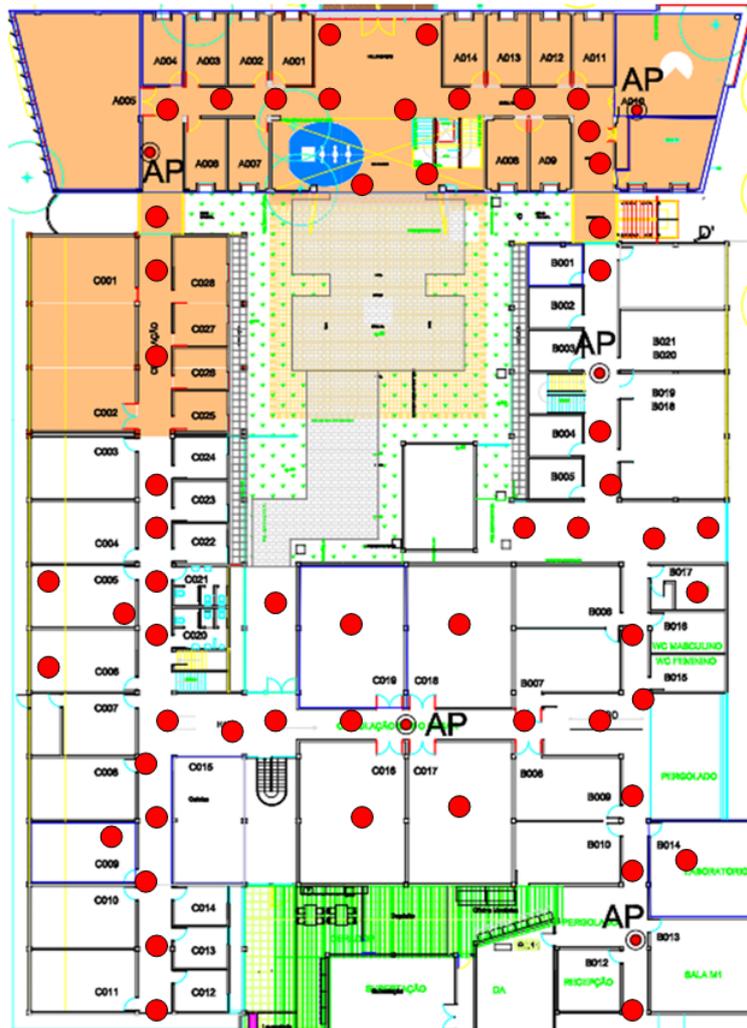


Figura 4.1 Planta Baixa do Centro de Informática - UFPE.

**Tabela 4.1** Parâmetros K-Means.  
Configurações K-Means

Parâmetros	Intervalo de Parâmetros
Intervalo dos Agrupamentos	5-7
Quantidade de Repetições	100
Número de Sementes Aleatórias	5
Métrica	Euclidiana
Iterações	300
Método para Medição do resultado do Cluster	Silhueta Simplificada

#### 4.1.2.1 Medição das intensidades dos Sinais — RSS no ambiente.

O primeiro ponto a ser definido é a área que será coberta pelo sistema. Esta área servirá de dado de entrada para outras etapas do processo. Além da área de interesse, a faixa das regiões virtuais e o número de medições a serem realizadas durante a fase de treinamento devem ser definidos e serão explicados na próxima seção.

Em seguida, é necessário verificar a quantidade de APs presentes no ambiente a fim de garantir que o sistema seja exequível, foi definido que se deve verificar se a cada 25 metros existem pelo menos 3 APs com sinais acima de -75dBm, estes valores garantem a possibilidade de localizar um dispositivo com a precisão desejada. Estes são também valores de distância e ponto de corte indicados pela Cisco [Cisco Systems, 2008].

Na coleta de dados realizada no CIN-UFPE os pontos ficaram a uma distância entre 3 a 5 metros em uma área de  $3.522 m^2$ .

#### 4.1.2.2 Parametrização para definição das regiões virtuais

Os dados coletados, intensidade do sinal e MacAddress em cada ponto  $x, y$ , são então organizados por MacAddress e os valores de intensidade, de cujos APs que não tiveram em toda a extensão do ambiente, pelo menos um sinal superior a -75dBm, são eliminados. Assim são eliminados os sinais dos APs, que não obtiveram pelo menos em um ponto, um sinal acima de -75dBm.

A estes dados coletados foi aplicado o algoritmo de agrupamento K-Means. A aplicação deste algoritmo foi realizada com os parâmetros descritos na Tabela 4.1.

Na primeira linha da Tabela foi definido que o sistema deve avaliar dentro de uma faixa de valores de números de agrupamentos qual o melhor valor, na segunda linha foi definida esta faixa, no caso, de 5-7 agrupamentos. O sistema irá buscar o que tiver melhor resultado, este resultado é encontrado através da métrica *silhueta* apresentada no Capítulo 2 e que está definido na última linha da tabela. Os resultados apresentados foram obtidos após experimentos que resultaram os seguintes valores. A quantidade de pontos por agrupamento teve que ser a maior possível, pois com poucos pontos a rede neural ficava mal treinada.

Para a escolha das sementes de inicialização do *K-means* foi escolhido o método K-Means++, que garante uma forma de convergência mais rápida que uma escolha aleatória dos valores das sementes de inicialização [Arthur and Vassilvitskii, 2007]. E o número de sementes randô-

**Tabela 4.2** Configuração das Redes Neurais.

Modo de Treinamento da Rede	Intervalo de Parâmetros
Especificação da Camada Oculta	Totalmente conectada
Número de Nós Ocultos	32
Função de Ativação	Tangente Hiperbólica
Taxa de Aprendizado	0,0001
Número Máximo de Iterações	1000000
Tipo de Solucionador	lbfgs

micas escolhido foi de 5, este valor de 5 grupos de sementes iniciais, possibilita uma melhor escolha do melhor agrupamento, pois irá processar de 5 diferentes grupos de sementes iniciais, buscando o melhor valor.

Um outro parâmetro é a métrica de similaridade. Utilizamos a distância euclidiana, por ser a que mais se aproxima da variação do sinal real em função da distância.

#### 4.1.2.3 Parametrização da etapa de localização

Após terem sido obtidas as identificações das regiões virtuais pela técnica baseada no algoritmo K-Means, o próximo passo foi a definição da localização, a partir destas regiões. Para esta etapa, foi usada uma técnica baseada em redes neurais, conforme citado no Capítulo anterior. A rede neural utilizada foi uma rede MLP — *Multilayer Perceptron* com os parâmetros mostrados na Tabela 4.2, de acordo com a biblioteca Scikit Learn.

A primeira linha indica que para o treinamento deve ser pesquisada qual das taxas descritas apresenta um erro menor durante a fase de treinamento. O segundo parâmetro define uma rede completamente conectada. Isto é, todos os nós de entrada estão ligados a todos os nós da camada intermediária, e estes à saída.

A terceira linha indica que o número de nós da camada intermediária foi igual a 32. A definição deste valor se deu após a realização de testes com 100, 60 e 32 nós, sendo que com 32 nós foram obtidos melhores resultados. Esta configuração também é sugerida por [Youssef and Agrawala, 2005].

Na quarta linha é definido a taxa de aprendizado da rede, foram adotados os valores *default* de 0,0001.

O número de iterações foi definido como 1000000.

A função de ativação escolhida foi a tangente hiperbólica, pois os resultados com essa função foram os melhores. O tipo de solucionador escolhido foi o 'lbfgs', que é um solucionador default da biblioteca. Na documentação da Scikit Learn diz que esse solucionador é o ideal quando precisa treinar uma rede com poucos dados.

## 4.2 Medições Realizadas

Nesta seção serão apresentados os dados obtidos em cada fase do processo, bem como o resultado do sistema de localização para todos os pontos testados. Ao final será apresentada uma

comparação entre o método implementado e a implementação de Barros.

### 4.2.1 Medições com o Sistema Embarcado

O primeiro dado obtido é o valor da intensidade do sinal — RSS medido pela placa de medição baseada no processador Edison. A Tabela 4.3 mostra a forma como os dados são fornecidos pela placa de medição das intensidades dos sinais — RSS, esta tabela é apenas uma amostra dos resultados.

**Tabela 4.3** Uma amostra dos resultado dos valores RSS fornecido pelo SoC Edison.

Ponto <sub>1,3.txt</sub>	BSID	Freq	RSS	Flags	SSID
	00:19:77:74:de:1b	2462	-68	[ESS]	testCin
	c8:d7:19:fb:bd:27	2432	-57	[WPA2-PSK-CCMP][ESS]	/x00/x00/x00/x00/x00/x00/x00
	00:19:77:78:1b:96	2412	-43	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	CINUFPE2
	9c:5d:12:6f:f4:56	2472	-38	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	CINUFPE2
	68:7f:74:e6:23:11	2452	-66	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	wifimac2
	9c:5d:12:6f:e0:a8	5220	-85	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	CIn-GUESTS2
	9c:5d:12:6f:e0:96	2412	-76	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	CINUFPE2
	00:19:77:74:de:18	2462	-70	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	APTtest
	00:19:77:74:de:28	5785	-81	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	APTtest
	48:f8:b3:ca:7f:54	2452	-47	[WPA2-PSK-CCMP][ESS]	/x00/x00/x00/x00/x00/x00/x00/x00/x00/x00
	00:19:77:74:de:2b	5785	-69	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	CINUFPE2
	9c:5d:12:6f:e0:9b	2412	-77	[ESS]	testCin
	00:19:77:74:de:2f	5785	-80	[ESS]	testCin
	00:0f:cb:ab:52:be	2462	-82	[WEP][ESS]	Oficina-Mecanica
	00:19:77:74:de:17	2462	-75	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	CIn-GUESTS2
	9c:5d:12:6f:e0:a7	5220	-82	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	CINUFPE2
	40:18:b1:59:11:19	2452	-74	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	CINUFPE2
	9c:5d:12:6f:e0:ab	5220	-84	[ESS]	testCin
	06:15:6d:da:fd:b0	2412	-85	[ESS]	INCT-INAMI
	0a:15:6d:da:fd:b0	2412	-80	[WPA-PSK-TKIP][ESS]	mySecure
	00:19:77:78:1b:97	2412	-42	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	CIn-GUESTS2
	9c:5d:12:6f:f4:57	2472	-40	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	CIn-GUESTS2
	9c:5d:12:6f:f4:69	5180	-62	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	CINUFPE2
	9c:5d:12:6f:f4:6a	5180	-62	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	CIn-GUESTS2
	00:19:77:78:1b:ab	5745	-59	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	CINUFPE2
	2e:a4:3c:ed:f9:53	2437	-87	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	CINUFPE
	40:18:b1:59:11:1a	2452	-80	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	CIn-GUESTS2
	9c:5d:12:6f:e0:97	2412	-74	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	CIn-GUESTS2
	e8:de:27:ee:0a:2c	2417	-91	[WPA2-PSK-CCMP][WPS][ESS]	DMAT
	00:19:77:74:de:16	2462	-70	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	CINUFPE2
	08:ea:44:77:4f:19	2412	-79	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	MotoEmbratel
	08:ea:44:77:4f:1c	2412	-86	[WPA-PSK-CCMP][ESS]	
	40:18:b1:e8:9e:16	2437	-80	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	CINUFPE2
	9c:5d:12:6f:f5:17	2452	-88	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	MotoRNP
	9c:5d:12:6f:f5:16	2452	-89	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	MotoEmbratel
	08:ea:44:77:4f:16	2412	-76	[WPA-EAP-CCMP+TKIP][WPA2-EAP-CCMP+TKIP][ESS]	CINUFPE2
	08:ea:44:77:4f:17	2412	-75	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	CIn-GUESTS2
	08:ea:44:77:4f:1a	2412	-80	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	MotoRNP
	6c:70:9f:e5:2e:02	2412	-86	[WPA2-PSK-CCMP][ESS]	RMAC
	a0:f3:c1:f8:6b:0b	2437	-87	[WPA-PSK-CCMP+TKIP][WPS][ESS]	OBAMA
	a0:f3:c1:f8:6c:71	2472	-91	[WPA2-PSK-CCMP+TKIP][WPS][ESS]	R17
	00:19:77:78:1b:ac	5745	-57	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP][ESS]	CIn-GUESTS2
	48:f8:b3:ca:7f:56	5805	-63	[WPA2-PSK-CCMP][ESS]	/x00/x00/x00/x00/x00/x00/x00/x00/x00
	00:1a:3f:4a:d2:02	2437	-88	[WPA-PSK-CCMP+TKIP][WPA2-PSK-CCMP+TKIP-preauth][ESS]	ppgmtr

#### 4.2.1.1 Valores RSS

Para cada grupo de medição está associado um arquivo em formato 'txt', com o objetivo de identificar o local onde aqueles sinais foram medidos e, assim, gerar o mapeamento inicial e durante os testes para a verificação do erro.

Na primeira coluna da Tabela 4.3 é o BSID (*Basic service set identification*), que é identificado pelo *MacAddress* de cada placa de rede, de cada AP. A segunda coluna é a frequência do sinal, onde identifica-se a frequência em kHz e o canal de cada AP. A terceira coluna tem o valor RSS em dBm, a coluna flags, mostra os protocolos de segurança transmitidos e a última coluna lista o SSID (*Service set identification*) que é o nome da rede.

#### 4.2.1.2 Resultados RSS

Durante as medições foram identificados um total de 224 sinais de diferentes AP's, incluso nestes todos os pontos de acesso encontrados. Porém depois do tratamento dos dados só foi trabalhado com 157 sinais de diferentes AP's, pois foi verificado que 67 sinais eram ruidosos e dinâmicos, assim eles foram excluídos para não atrapalharem o fluxo do processo.

A Tabela 4.4 é um pequeno exemplo e mostra a quantidade de APs por SSID (*Service Set Identifier*), bem como os nomes de algumas das redes sem fio encontradas no ambiente do CIN.

#### 4.2.1.3 Mapeamento RSS

Depois que os dados foram coletados, eles são agrupados por SSID e ordenados por pontos, e são acrescentadas as referências x, y. A Tabela 4.5 mostra os resultados depois da normatização. Esta normatização consiste em preencher os valores inexistentes com -100 dBm, e acrescentar + 100 a todos os valores, de modo que fiquem positivos. Aqui está apresentado apenas um fragmento da Tabela completa para fins de visualização dos formatos dos dados. Os sinais RSS serão transformados em positivos para alimentação do algoritmo de geração das *Regiões Virtuais*.

**Tabela 4.4** Quantidade de APs por SSID no CIN-UFPE.

<b>SSID</b>	<b>Quantidade Aps</b>
CIn-GUESTS2	34
CINUFPE2	34
MotoEmbratel	23
MotoRNP	22
AutoTest	12
CINGUESTS	7
CINUFPE	6
testCin	6
/x00/x00/x00/x00/x00/x00/x00/x00	5
INCT_INAMI	5
mySecure	5
DMAT	4
/x00/x00/x00/x00/x00/x00/x00/x00/x00	3
mesh	3
APTest	2
/x00/x00/x00/x00/x00/x00	1
DE-PG	1
DE-PG2	1
DIRECT-ppML-2160 Bruno	1
DoutorMarcio	1
eapsim	1
Eduroam	1
FamiliaFeliz	1
ppgmtr	1
R02	1
R17	1
R175GHZ	1
RMAC	1
RMAC-5GHz	1
STP-LINK	1
TCMauroWifi	1
wifi UFPE	1
wifimac2	1
Redes com SSID ocultos	11

**Tabela 4.5** Uma amostra do mapeamento dos sinais RSS, com as posições X, Y e número da Varredura.

Pos: X	Pos: Y	00:19:77:78:23:96	00:19:77:78:23:ab	00:19:77:7c:1b:96	00:19:77:7c:1b:ab	08:ca:44:77:4f:16	08:ca:44:77:4f:2b	08:ca:44:77:4f:53	08:ca:44:77:4f:6b	08:ca:44:77:80:d5	
1	405	209	0	0	0	0	0	0	18	0	21
2	405	209	0	0	0	0	0	0	18	0	21
3	405	209	0	0	0	0	0	0	18	0	21
4	405	216	0	0	0	0	0	0	0	0	23
5	405	216	0	0	0	0	0	0	0	0	30
6	410	216	0	0	0	0	0	35	0	0	28
7	410	216	0	0	0	0	0	31	0	0	28
8	411	212	28	0	0	0	0	35	0	0	28
9	411	209	28	0	0	0	0	35	0	0	28
10	411	209	50	0	34	30	48	42	0	0	0
11	412	209	42	0	43	33	50	42	0	0	28
12	413	207	55	29	54	37	64	29	0	0	30
13	414	206	42	19	48	37	65	49	26	0	22
14	414	203	45	29	46	34	49	49	26	0	25
15	413	202	40	29	50	30	63	36	27	0	25
16	414	201	33	29	43	36	69	36	27	0	25
17	414	198	34	29	49	25	72	36	27	0	25
18	413	196	43	0	44	31	68	61	0	0	0
19	413	194	43	0	49	32	65	61	0	0	0
20	413	190	43	0	44	30	62	61	0	0	0
21	413	186	31	0	47	26	60	56	0	0	0
22	412	187	31	0	51	32	55	56	0	0	0
23	412	190	34	0	56	34	61	55	0	0	0
24	412	191	36	0	59	34	58	55	0	0	0
25	412	193	40	0	53	40	76	69	0	0	0
26	412	194	40	0	48	34	67	65	0	0	0
27	412	196	40	0	51	39	61	65	0	0	0
28	412	199	43	0	45	35	69	53	0	0	0
29	412	201	43	0	52	36	60	39	0	0	27
30	412	202	39	0	49	36	58	39	0	0	27
31	412	205	31	0	34	36	58	39	0	0	19
32	415	204	31	0	31	0	52	37	0	0	19
33	416	204	31	0	29	0	36	37	0	0	19
34	418	204	19	0	29	0	36	23	0	0	19
35	420	204	19	0	29	0	37	24	0	0	27
36	421	207	28	0	30	0	50	29	0	0	27
37	417	207	28	0	30	0	43	39	0	0	20
38	414	207	45	0	41	0	55	55	24	0	20
39	413	208	38	0	35	23	52	24	0	0	35
40	413	209	44	31	36	0	50	24	0	0	33
41	413	212	46	30	36	0	50	41	0	0	33
42	413	214	45	32	36	0	49	44	0	0	39
43	413	220	42	27	37	0	49	39	0	0	29
44	413	222	51	32	37	28	45	48	0	17	33
45	413	223	47	33	49	28	51	42	0	24	35
46	413	225	52	36	49	28	44	33	0	24	34
47	413	226	51	33	49	28	43	42	0	24	39
48	413	226	45	40	49	0	43	35	0	24	27
49	413	227	40	25	0	0	44	30	0	0	27
50	413	229	48	24	0	0	44	39	0	0	37
51	414	231	43	24	0	0	43	36	0	0	37
52	413	232	49	31	0	0	43	32	0	0	27
53	413	235	33	23	0	0	40	29	0	0	15
54	413	236	41	23	0	0	46	35	0	0	15

**Tabela 4.6** Métrica do Agrupamento conforme o método "silhueta".

Métrica	Nr. Agrupamentos
0.332338	5
0.364363	6
0.384862	7

#### 4.2.2 Criação das Regiões Virtuais e Redes Neurais

Os dados normatizados foram entradas para o método de geração das *Regiões Virtuais*, que foi baseado no algoritmo K-Means com um número de iterações de no máximo 300 ou até a convergência do algoritmo.

As métricas dos agrupamentos testados pela ferramenta, para encontrar a melhor distribuição são apresentadas na Tabela 4.6.

Observa-se que, segundo o método silhueta, a pior distribuição seria com 5 agrupamentos e a melhor com 7.

O mapa da Figura 4.2 mostra o resultado da área de interesse no CIn classificado por região virtual.

A Figura 4.3 mostra os agrupamentos que formaram as Regiões Virtuais do CIn. Para obter estes valores, foram testadas várias combinações sendo estas regiões apresentadas, as que melhores representavam para os dados RSS medidos. Observa-se pelo mapa da distribuição que os elementos foram agrupados de acordo com a localidade, apesar das medições terem sido propositalmente trocadas em alguns pontos para verificação da eficiência do algoritmo.

A informação das regiões virtuais e respectivos valores de sinais por região são usados no treinamento das redes neurais para a determinação da posição de um objeto a partir das medições das intensidades dos sinais — RSS.

### 4.3 Resultados - Localização

Para validação do sistema proposto, foram executadas 236 medições em diferentes localizações do CIn e em dias e horários diferentes. Onde 200 pontos foram usados para treinamento da parte Off-line e 36 pontos para localização na parte On-line. Para cada medição foi aplicada a técnica de localização. Os resultados são apresentados nas subseções a seguir.

#### 4.3.1 Alocação nas Regiões Virtuais

Para cada ponto foi realizada a medição dos valores RSS, com estes valores a técnica alocou cada ponto a uma região virtual. Esta região virtual foi comparada à região virtual estabelecida na fase offline de acordo com a posição real em que cada ponto foi medido. O resultado é apresentado na Tabela 4.7.

Observa-se uma taxa de acerto muito grande, os pontos de fronteira são aqueles que ficam na transição de um agrupamento para outro, podendo ser alocado em ambos, que não irão induzir em erros.

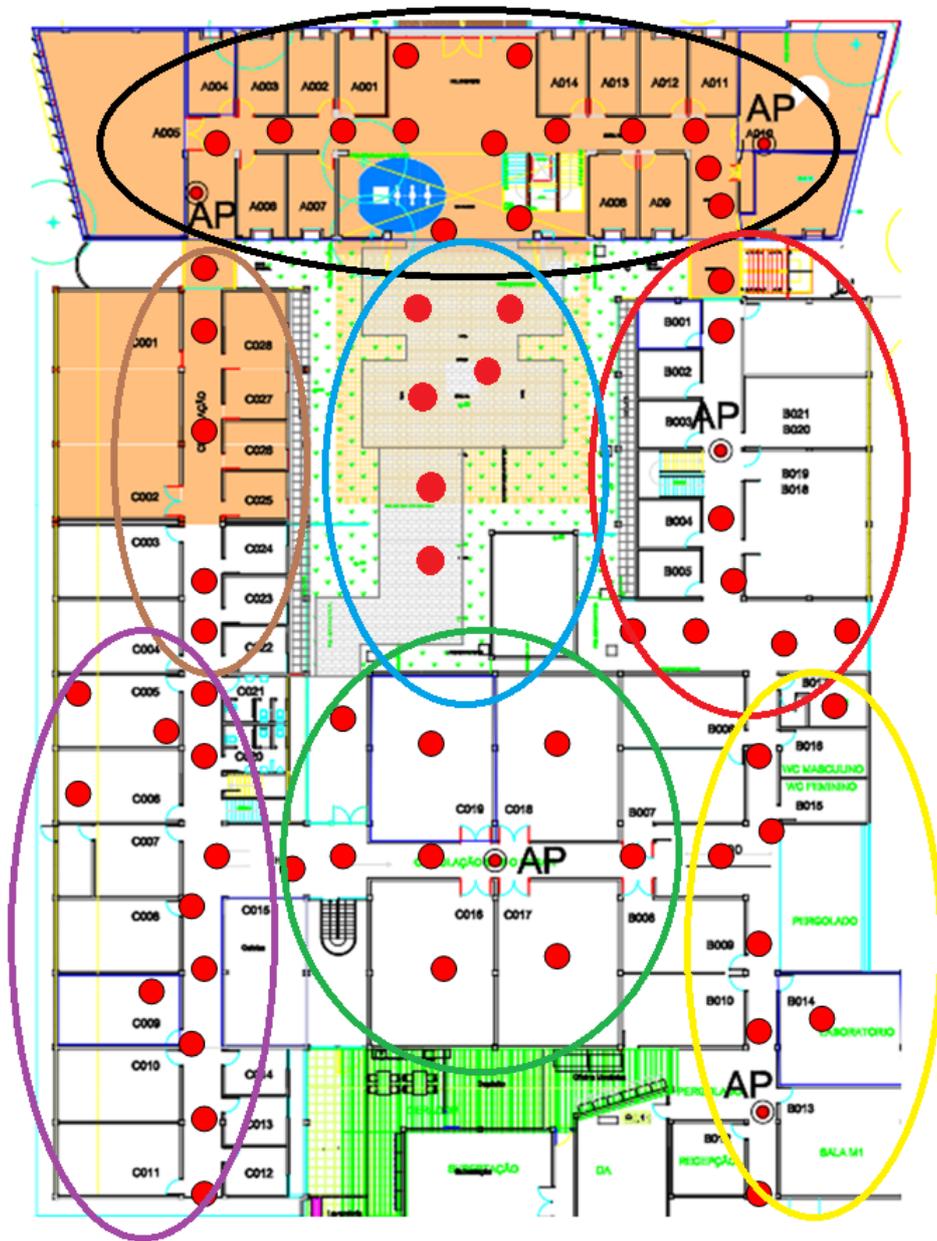


Figura 4.2 Regiões Virtuais do Cln.

```

Grupo 1:
10 11 12 13 14 15 16 17 18 19 20 21 92

Grupo 2:
23 24 25 26 27 28 29 30 31 32 33 34 35 36 37

Grupo 3:
1 2 3 4 5 6 7 8 9 60 81 82 83 84 85 86 87 88 89 90 91

Grupo 4:
72 73 74 75 76 77 78 79 80

Grupo 5:
61 62 63 64 65 66 67 68 69 70 71

Grupo 6:
38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57

Grupo 7:
22 58 59 93 94 95 96 97 98 99 100

```

**Figura 4.3** Agrupamento dos pontos formando as Regiões Virtuais do CIn.

**Tabela 4.7** Resultado Regiões Virtuais.

	Pontos	%
Acertos	33	91,68
Pontos de Fronteira	1	2,77
Erros	2	5,55
Pontos Totais	36	

### 4.3.2 Saídas das Redes Neurais

Para os 36 pontos que foram utilizados para localização, eles foram alocados para suas regiões virtuais e em seguida o seus sinais foram as entradas para a rede neural do seu grupo, assim pudemos fazer uma média geral dos erros como visto na figura 4.4 e na tabela 4.8, e também a estimativa de suas posições que foram as saídas de suas respectivas redes neurais.

**Tabela 4.8** Resultado Redes Neurais com Regiões Virtuais.

Média dos Erros	7,88m
Desvio Padrão	4,67
Erro de medição	1m pra menos

## 4.4 Comparativo com o Sistema de Barros

A tabela 4.9 mostra a comparação entre o Sistema proposto e o Sistema de Antônio Barros. Como o sistema proposto é uma automatização do sistema de Barros, podemos notar que a implementação do sistema embarcado mais o servidor local diminui drasticamente o tempo de execução para as duas fases do processo.

	Pontos	Posição Real		Posição Predicta Pela Rede		Distância (m)	Distância sem o erro de Medição (m)
		X	Y	X	Y		
Grupo 1:	112	445.2381	239.0133	453.0411	230.9823	11.79	10.79
	113	444.3496	233.0298	451.4468	232.9265	7.09	6.09
	114	443.3859	219.4196	438.6170	230.6603	12.21	11.21
Grupo 2:	115	445.7975	206.1529	446.8279	198.3870	7.83	6.83
Grupo 3:	101	431.6963	252.6810	442.3551	247.9833	11.64	10.64
	102	426.2748	252.3920	429.2700	254.8943	3.90	2.90
	103	421.1781	249.7908	421.2171	249.6471	0.14	0
	104	412.2619	249.7970	423.9403	250.4331	11.69	10.69
	105	413.7275	246.3486	428.2697	247.4089	14.58	13.58
	106	424.3760	249.3756	424.7087	255.3036	5.93	4.93
	107	425.4099	243.9368	430.7637	250.9396	8.81	7.81
	108	432.1043	243.8996	427.9290	250.0938	7.47	6.47
	109	434.4410	249.7330	423.3208	252.6624	11.49	10.49
	110	437.3033	249.8503	440.3047	245.8175	5.02	4.02
	111	445.0521	249.7929	443.2083	248.9075	2.04	1.04
	133	427.0826	243.5697	430.9443	254.4241	11.52	10.52
134	431.1797	249.0651	444.3808	252.6653	13.68	12.68	
Grupo 4:	130	413.3035	228.7157	410.7649	234.9421	6.72	5.72
	131	413.4898	233.0756	418.9365	240.4422	9.16	8.16
	132	413.3667	237.7738	416.7909	239.4440	3.80	2.80
Grupo 5:	126	417.0589	206.0414	423.1511	215.4174	11.18	10.18
	127	413.4652	210.8134	414.1231	230.9439	20.14	19.14
	128	413.3421	216.0036	413.5815	211.8007	4.20	3.20
	129	413.4159	220.6604	413.8009	205.6195	15.04	14.04
Grupo 6:	116	431.9351	203.7854	434.6332	206.7541	4.01	3.01
	117	434.5873	201.6489	426.8925	200.1333	7.84	6.84
	118	431.8917	208.3061	447.4719	210.4927	15.73	14.73
	119	434.5775	210.6878	433.7990	194.8711	15.83	14.83
	120	429.6534	203.8037	431.0072	199.3652	4.64	3.64
	121	427.1416	201.6415	425.5497	212.1135	10.59	9.59
	122	429.5788	208.3766	429.7125	212.0499	3.67	2.67
	123	427.0670	210.9115	428.7983	211.5137	1.83	1.83
	124	429.6534	206.0901	434.7759	197.0248	10.41	9.41
	125	421.5142	206.0905	428.6189	212.0630	9.28	8.28
Grupo 7:	135	428.6703	240.3111	434.3482	227.0958	14.38	13.38
	136	428.9959	233.7299	426.3154	233.6819	2.68	1.68

**Figura 4.4** Resultado das Redes Neurais

**Tabela 4.9** Comparativo entre sistemas.

Sistema	Tecnologia	Exatidão	Precisão	Complexidade	Robustez	Custo	Execução Off-line (min)	Parte	Execução Pare On-line (min)
Antônio Barros	RSS	4,7 m	60% em 5 m	Baixa	Alta	Baixo	120		30
Sistema Proposto	RSS	7,88 m	50% em 7,8 m	Baixa	Alta	Baixo	60		0,25

## 4.5 Implementação

No total fizemos dois tipos de implementações, uma foi o sistema embarcado e a outra foi a introdução do ambiente de aprendizagem de máquina no servidor local, fazendo assim a comunicação direta do servidor com o sistema embarcado.

Foram implementados os seguintes códigos no sistema embarcado:

- **Sistema de controle e comunicação:** Foi utilizada a linguagem de programação C e escrita 200 linhas de código. O tempo de controle mais o tempo de comunicação foi aproximadamente 9 segundos.

Foram implementados os seguintes códigos no servidor local:

- **Sistema de controle e comunicação:** Foi utilizada a linguagem de programação C e foram escritas 800 linhas de código. O tempo de resposta para comunicação e para controle dos subsistemas foi de aproximadamente 2 segundos.
- **Subsistema de Tratamento dos dados:** Foi utilizada a linguagem de programação C e foram escritas 400 linhas de código. O tempo de resposta para tratar os dados foi abaixo de 1 segundo.
- **Subsistema de Agrupamento dos dados:** Foi utilizada a linguagem de programação C e foram escritas 712 linhas de código. O tempo de resposta para fazer o agrupamento dos dados e gerar o banco de dados foi de aproximadamente 10 segundos.
- **Subsistema de Redes Neurais:** Foi utilizada a linguagem de programação C, Linguagem de Programação Python e a Biblioteca da Scikit Learn. Em linguagem de programação C foram escritas 200 linhas de código e em Python foram escritas 420 linhas de código. O tempo de resposta para treinar a rede neural e gerar o banco de dados foi de aproximadamente 3 segundos para cada rede neural de cada grupo, e o tempo de consulta abaixo de 1 segundo.

# 5 Conclusões e trabalhos futuros

## 5.1 Conclusões

Neste trabalho foi desenvolvido um sistema que buscou automatizar e incrementar algumas melhoras no processo de localização proposto na dissertação de Antônio Barros. A estratégia foi baseada na utilização de um sistema embarcado que capturava a intensidade dos sinais Wi-Fi das redes presentes no Ambiente. Como a intensidade de um sinal Wi-Fi varia com a distância até o ponto onde o sinal está sendo emitido, o método proposto utiliza estes valores de intensidade a partir dos diversos AP's presentes no ambiente a fim de determinar a localização.

Dessa forma o sistema embarcado está apto a enviar as informações capturadas e enviá-las para um servidor local que tem um ambiente de aprendizagem de máquina preparado. O sistema que executa no servidor recebe os dados do sistema embarcado e ativa 3 subsistemas de forma que executem em ordem.

Primeiro o subsistema de tratamento dos dados vai receber os dados brutos e começar a fazer o tratamento de forma que os dados fiquem aptos a entrada do próximo subsistema, obedecendo algumas regras e parametrizações.

Segundo o subsistema de agrupamento dos dados recebe os dados já tratados e começa a executar o algoritmo de clusterin, K-means. O algoritmo tem como saída todos os dados agrupados formando assim o que chamamos de Regiões virtuais. Cada Região Virtual está baseada na similaridade das intensidades dos sinais de seus respectivos pontos.

Terceiro o subsistema de redes neurais irá buscar no banco de dados todos os grupos que foram formados anteriormente. Dessa maneira o sistema central aloca para cada grupo uma rede neural MLP, e começa a treinar a rede neural de cada Região virtual.

Assim o sistema central está apto a receber solicitações do sistema embarcado que pode a qualquer momento solicitar uma localização de um determinado ponto aleatório que está contido dentro do mapeamento.

Com base nos experimentos realizados, as seguintes conclusões podem ser destacadas.

- O SoC Edison é bastante adequado para este tipo de aplicação, uma vez que apresenta baixo consumo de energia, é um processador de baixo custo e de tamanho pequeno, que foi bem adaptado ao capacete pra montagem do sistema embarcado. Os resultados apresentados por sua placa de rede WLAN integrada ao processador, forneceu bons resultados para o sistema de localização.
- Com a inserção do sistema embarcado no sistema de localização, diminuiu muito o tempo de mapeamento do local assim como o tempo de localização. O sistema embarcado consegue ajudar mais o usuário que está fazendo o mapeamento a ter mais liberdade

para fazer outras coisas como verificar o local do ponto de medição, localizar o ponto no mapa, determinar o tempo exato de começar a colher as intensidades dos sinais, ter um feedback do que está acontecendo durante o mapeamento, ter uma resposta do servidor em relação a receber os dados com sucesso, tudo isso enquanto o mapeamento está sendo feito.

- O sistema embarcado junto com o sistema de localização no servidor pode ser aplicado em qualquer ambiente interno, e a depender da qualidade dos materiais e aplicações, eles podem se tornar um sistema de localização de tempo real.
- O sistema embarcado além de auxiliar todos os tipos de usuários pode ser bem empregado em aplicações que auxiliem deficientes visuais a se localizarem em ambientes internos, onde seriam feitas poucas mudanças no sistema proposto, no lugar de comunicação visual, poderia ser implementado a comunicação auditiva. Além de que o sistema proposto é altamente recomendável para aplicações como robôs, pois pode ser inserido facilmente nos mesmos.
- As medidas da intensidade do sinal Wi-Fi têm uma influência muito grande nos resultados. É importante ressaltar que durante a fase de mapeamento estas medidas devem ser executadas com precisão tanto quanto à localização, como ao valor. Caso a análise dos valores medidos apresentem variações inesperadas, as medições devem ser refeitas. Como variações inesperadas podem ser citadas: uma determinada rede desaparece, mesmo tendo um valor significativo em um ponto adjacente; ou a diminuição muito grande do valor do sinal seguido do crescimento repentino entre pontos também adjacentes.
- O servidor local mostrou-se bem apropriado para a realização da tarefa, pois possibilita receber dados direto do sistema embarcado, e podemos implementar qualquer sistema no servidor pois o mesmo tem alta capacidade de processamento.
- A comunicação do sistema embarcado com o servidor local melhorou significativamente o tempo de mapeamento como o tempo de localização. Pois a comunicação feita de forma local está menos propícia a erros do que comunicações externas, como por exemplo ter que existir acesso a internet para poder se comunicar com um servidor em nuvem, então dependeria de fatores externos para que houvesse uma boa comunicação com o servidor em nuvem, diferentemente quando a comunicação é feita de forma local, que é simples e mais rápida.
- O sistema de localização é sensível a variações de potências de emissão dos APs. Caso o administrador do sistema não tenha controle total sobre estes parâmetros, é aconselhável uma verificação periódica sobre a precisão do sistema a fim de detectar erros acima dos previstos.
- O algoritmo proposto para definir as Regiões Virtuais, isto é, gerar os agrupamentos com intensidades de sinais semelhantes, foi o algoritmo K-Means, que mostrou bons resulta-

dos. O algoritmo tem uma convergência rápida e agrupa da melhor maneira possível os dados.

- A técnica *silhueta* é um método de validação e interpretação dos resultados de um agrupamento. O uso de técnica *silhueta* também foi de suma importância, pois esta é uma das principais dificuldades das técnicas de agrupamento a definição da melhor quantidade e qualidade dos agrupamentos.
- Com a utilização de técnicas de *clustering* para a definição das regiões virtuais, as redes neurais ficaram com poucos dados para treinamento, isso dificulta muito o processo de aprendizagem, pois ai tivemos que parametrizar as redes de uma forma que elas conseguissem convergir e poder dar um bom resultado. A quantidade de pontos influenciou muito na qualidade dos algoritmos de aprendizagem de máquina. Precisariamos de muito mais pontos para que a qualidade do treinamento das redes neurais aumentassem de forma a diminuir o erro da localização.
- A quantidade de pontos pode ser mais alta e bem mais distribuídas, pois foi verificado que o sistema tanto na parte Off-line quanto On-line teve um ótimo tempo de resposta. Assim com a quantidade de pontos bem maior, o sistema vai continuar respondendo rápido e a precisão da localização será melhor.

## 5.2 Trabalhos Futuros

O presente trabalho mostrou que a implementação do sistema embarcado junto com um servidor local melhorou muito o tempo de comunicação, e a flexibilidade de melhorar o sistema de localização na medida de cada vez mais diminuir o erro de localização. Mas alguns pontos podem ser objeto de futuras pesquisas e desenvolvimento.

- **Realimentação dos Dados:** Como o ambiente pode mudar, com a inclusão ou exclusão de APs, ou mesmo a variação de sua potência, no estado atual seria necessário um novo levantamento das intensidades. Poderia ser estudada uma maneira de verificar o comportamento e uma verificação se o sistema está respondendo corretamente e, se possível, uma realimentação dos dados automaticamente.
- **Comportamento dos Sinais:** Como a intensidade dos sinais oriundos dos AP's variam de acordo com fatores naturais, como a quantidade de pessoas no ambiente, horário de ligamento de equipamentos que gerem ruídos na rede, equipamentos que gerem interferência nos sinais, temperatura, chuva e umidade, poderíamos estudar a implementação de vários mapeamentos nesses horários e levantar um banco de dados para cada horário, assim ativando o sistema de localização para se adequar ao ambiente dependendo do horário em que o usuário estiver usando.
- **Filtro de dados com erros de medição:** Um dos maiores problemas do sistema são os erros de medição, sejam por interferências externas, ou mesmo do dispositivo. Poderiam ser estudados filtros de medição, baseados no estágio anterior de modo que grandes

variações fossem detectadas e eliminadas, que um histórico fosse mantido das últimas medições de cada dispositivo, que não permitisse variações maiores que um determinado limite.

- **Ambiente 3D:** Todos os testes foram efetuados no plano, pois destina-se mais à localização de pessoas. Esta técnica pode facilmente ser adaptada a ambientes com três dimensões e analisados os resultados.
- **Análise com outras técnicas de aprendizado de máquinas:** Outros tipos de algoritmos de agrupamentos, e redes neurais mais complexas podem ser testados e comparados os resultados. Os algoritmos escolhidos foram buscados por serem simples e rápidos. Porém, para uma necessidade de exatidão maior, ou de uma tolerância maior a falhas, novos métodos podem apresentar melhores resultados.

## Referências Bibliográficas

- [Adler et al., 2015] Adler, S., Schmitt, S., Wolter, K., and Kyas, M. (2015). A survey of experimental evaluation in indoor localization research. In *Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on*, pages 1–10. IEEE.
- [Akin and Geier, 2004] Akin, D. and Geier, J. (2004). 802.11 phy layers. *CWAP-certified wireless analysis professional official study guide*, pages 353–355.
- [Alliance, 2009] Alliance, Z. (2009). Zigbee wireless sensor applications for health, wellness and fitness. Zigbee whitepaper, Zigbee Alliance.
- [Arthur and Vassilvitskii, 2007] Arthur, D. and Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics.
- [Bahl and Padmanabhan, 2000] Bahl, P. and Padmanabhan, V. N. (2000). Radar: An in-building rf-based user location and tracking system. *IEEE INFOCOM 2000 19th. Annual Joint Conference of the IEEE Computer and Communications Societies*, pages 775–784.
- [Battiti et al., 2002] Battiti, R., Le, N. T., and Villani, A. (2002). Location-aware computing: a neural network model for determining location in wireless LANs. University of Trento.
- [Bernardi and Landin, 2002] Bernardi, J. V. E. and Landin, P. M. B. (2002). Aplicação do sistema de posicionamento global (gps) na coleta de dados. *DGA, IGCE, UNESP/Rio Claro, Lab. Geomatémática, Texto Didático*, 10(31):2002.
- [Cisco Systems, 2008] Cisco Systems, I. (2008). “wi-fi location-based services 4.1 design guide,” cisco systems, may 2008. <http://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Mobility/WiFiLBS-DG.pdf>.
- [de Assunção Barros, 2016] de Assunção Barros, A. C. G. (2016). Proposta de técnica de localização interna para dispositivos móveis utilizando redes locais sem fio.
- [Die.net, 2016] Die.net (2016). die.net wpa cli(8) - linux man page. <http://linux.die.net/man/8/wpacli>.
- [Ekahau, 2016] Ekahau (2016). Ekahau - real-time location system. <http://www.ekahau.com/real-time-location-system>. Acessado em Janeiro 2016.

- [Gu et al., 2009] Gu, Y., Lo, A., and Niemegeers, I. (2009). A survey of indoor positioning systems for wireless personal networks. *Communications Surveys & Tutorials, IEEE*, 11(1):13–32.
- [IEEE, 2012] IEEE (2012). *IEEE Standard — Telecommunications and information exchange between systems Local and metropolitan area networks— Specific requirements*. Number Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications.
- [Igate, 2016] Igate (2016). Igate indoor positioning system. <http://www.igate.com/documents/11041/0/IndoorpositioningsystemusingBeacons.pdf/a1814507-8f2e-4a20-bb7a-f682b034cedf>. Acessado em Janeiro 2016.
- [Intel, 2015] Intel (2015). *Intel Edison Wi-Fi Guide*. Intel, revision 1 edition. Document Number: 331438-001.
- [Kurose and Ross, 2010] Kurose, J. F. and Ross, K. W. (2010). *Redes de Computadores ea Internet: uma abordagem top-down*. Pearson.
- [Morimoto, 2008] Morimoto, C. E. (2008). Redes wireless, parte 3: Configurando o ponto de acesso. Disponível em: <http://www.hardware.com.br/tutoriais/configurando-ponto-acesso/pagina5.html>.
- [Myllymäki and Tirri, 2007] Myllymäki, P. and Tirri (2007). Location estimation in wireless telecommunication networks. US Patent 7,228,136.
- [Neural Network in a Nutshell, 2016] Neural Network in a Nutshell (2016). The Men Who Stare at Codes - ABC: Always Be Coding . Disponível em: <https://themenwhostareatcodes.wordpress.com/2014/03/02/neural-networks-in-a-nutshell/comment-page-1/>.
- [Peter Harrop, 2016] Peter Harrop, R. D. (2016). Mobile phone indoor positioning systems (IPS) and Real Time Locating Systems (RTLS) 2014-2024 forecasts, players, opportunities. [http://www.idtechex.com/research/reports/mobile-phone-indoor-positioning-systems-ips-and-real-time\\_locating-systems-rtls-2014-2024-000359.asp](http://www.idtechex.com/research/reports/mobile-phone-indoor-positioning-systems-ips-and-real-time_locating-systems-rtls-2014-2024-000359.asp).
- [Rousseeuw, 1987] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.
- [Scikit-Learn, 2016] Scikit-Learn (2016). Scikit-learn machine learning in python. <http://scikit-learn.org/stable/>. Acessado em Novembro 2016.
- [Systems, 2006] Systems, C. (2006). “the secure shell (ssh) protocol architecture”, cisco systems, january 2006. <https://tools.ietf.org/html/rfc4251>.

- [Teldio, 2016] Teldio (2016). Teldio - indoor positioning system. <http://teldio.com/products/ips/>. Accessed: 02/01/2016.
- [Vasques, 2010] Vasques, B. L. R. P. (2010). Zigbee - introdução. Disponível em: <http://www.gta.ufrj.br/grad/101/zigbee/referencias.html>.
- [Vieira and Roisenberg, 2003] Vieira, R. C. and Roisenberg, M. (2003). Redes neurais artificiais: um breve tutorial. *Laboratório de Conexão e Ciências Cognitivas, Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil*.
- [Wysocki and Zepernick, 2000] Wysocki, T. A. and Zepernick, H.-J. (2000). Characterization of the indoor radio propagation channel at 2.4 ghz. *Journal of telecommunications and information technology*, pages 84–90.
- [Youssef and Agrawala, 2005] Youssef, M. and Agrawala, A. (2005). The Horus WLAN location determination system. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services, MobiSys '05*, pages 205–218, New York, NY, USA. ACM.
- [Zonith, 2016] Zonith (2016). Zonith - staff safety solutions. <http://zonith.com/products/staffsafety/>. Acessado em Janeiro 2016.