

Utilizando Características do Ambiente Monitorado Para Aumentar a Precisão na Localização de Dispositivos Wi-Fi Evitando a Calibragem

Luís F. M. de Moraes, Bruno A. A. Nunes, Rafael de M. S. Fernandes

¹RAVEL - Laboratório de Redes de Alta Velocidade
Programa de Engenharia de Sistemas e Computação – COPPE
Universidade Federal do Rio de Janeiro (UFRJ) - Rio de Janeiro, Brasil

{moraes,bruno,rafael}@ravel.ufrj.br

Abstract. *A system for locating wireless network devices in a local area, based upon radio-frequency characteristics was proposed and analyzed in [1]. In the present work, the construction of the model considered in [1] is modified to consider the characteristics of the monitored environment, such as walls, windows, closets, and other types of obstacles. Taking in consideration these characteristics, the precision of the system grew up to 13%, without any increase in the complexity or effort of implementation.*

Resumo. *Um sistema para localização de dispositivos em redes locais sem fio, baseado em características de rádio-frequência foi proposto e analisado em [1]. No presente trabalho, a construção do modelo proposto em [1] foi modificada para considerar as características do ambiente monitorado, tais como paredes, janelas, armários e outros tipos de obstáculos. Levando em consideração essas características, a precisão do sistema aumentou em até 13%, sem qualquer aumento na complexidade ou esforço de implementação.*

1. Introdução

As redes locais sem fio baseadas no padrão IEEE 802.11 [2] (também conhecidas como redes Wi-Fi), se tornaram uma alternativa popular para a interconexão de estações em um ambiente local. Como consequência do potencial de mobilidade dos usuários, a proliferação de dispositivos móveis e os aumentos das taxas de transmissão, diversas aplicações para esses modelos de rede também tem sido desenvolvidas.

Nesse contexto, os Serviços Baseados em Localização - SBLs (*Location Based Services* - LBS) têm despertado grande interesse, sendo ultimamente alvo de várias pesquisas. O objetivo deste trabalho é melhorar a precisão de sistemas de localização que utilizem modelos de propagação levando em consideração a atenuação do sinal devido a obstáculos para estimar o valor de nível de sinal esperado em um determinado local.

O foco deste trabalho não é construir um novo modelo de propagação de sinal de rádio para estimar a localização ou novos algoritmos, como a triangularização, e sim propor um novo método de construção do modelo de localização, onde algoritmos clássicos de estimativa de localização podem ser agregados inteligentemente e modificados para aumentar a precisão e reduzir custos. Neste trabalho, utilizou-se o algoritmo descrito em [1] para estimar a localização de dispositivos Wi-Fi, modificando-se apenas a construção do modelo. Em [1], os autores não consideraram características do ambiente monitorado. No presente artigo, é mostrado que estimar a atenuação do sinal de RF, levando em consideração as características físicas do ambiente, pode levar a uma melhora na

precisão de até 13% em média, sem qualquer aumento na complexidade ou esforço de implementação.

A sequência deste artigo é organizada da seguinte forma. Na Seção 2 são apresentados os trabalhos relacionados com técnicas para estimar a localização. Na Seção 3 e 4 são apresentadas duas arquiteturas para Sistemas de Localização. A técnica e modelo utilizados no sistema proposto são apresentados nas Seções 5 e 6. A Seção 7 mostra técnicas utilizadas para melhorar a precisão do sistema e a Seção 8 apresenta os resultados obtidos experimentalmente. Finalmente na Seção 9 são apresentadas as conclusões.

2. Trabalhos Relacionados

Sistemas de localização atuais, em sua grande maioria, funcionam em duas fases. Uma primeira fase de calibragem, chamada na literatura também de fase *off-line*, onde medições sobre a qualidade do sinal de RF são realizadas em diversos pontos específicos do ambiente no qual se deseja implementar o sistema de localização. A segunda fase consiste em coletar informações do sinal recebido por um dispositivo sem fio e utilizá-las, junto com os dados colhidos na primeira fase, para localizar tal dispositivo. O problema é que essa calibragem pode durar várias horas e muitas vezes ela precisa ser refeita, pois o ambiente monitorado é constantemente alterado, principalmente por pessoas [3]. Em [4] é possível ver um exemplo do alto custo de implantação e manutenção destes sistemas. Em um dos experimentos relatados em [4], os autores levaram mais de 10 horas para realizar as medições necessárias à calibragem do sistema de localização, em uma área de aproximadamente 68m x 26m.

Uma das vantagens inerentes ao mecanismo proposto neste trabalho é que o mesmo não necessita do mencionado esforço inicial de calibragem, o que reduz consideravelmente a relação custo/benefício para esse tipo de sistema de localização.

Trabalhos anteriores sobre a construção de modelos para estimativa de localização de dispositivos sem fio incluem, na grande maioria dos casos, esforços de calibragem onde cada ponto do local de cobertura da rede sem fio é mapeado em um vetor de RSSI (*Received Signal Strength*). Uma vez criado o mapa/modelo de localização, alguma técnica deve ser utilizada, na fase *on-line*, para comparar as medições de RSSI realizadas em tempo real, com os valores contidos no modelo. Diferentes técnicas de comparação entre o RSSI medido em tempo real e os valores registrados no mapa de calibragem vêm sendo utilizadas. Dentre estas técnicas é possível citar as que utilizam redes neurais [5], modelos probabilísticos [6, 7], distância euclidiana ou distância de Manhattan entre o valor medido e o valor registrado no mapa e o método dos k vizinhos mais próximos [8]. A técnica escolhida nessa comparação é crucial para a precisão do sistema.

A maior parte das técnicas atuais utilizam um esforço substancial de calibragem para gerar o modelo de localização [6, 7, 8, 9, 10, 11, 12] e sofrem algum tipo de degradação quando existem mudanças no local a ser monitorado. Até mesmo soluções comerciais, como Ekahau [13], baseiam-se neste tipo de procedimento. Em ambientes dinâmicos, com grande fluxo de pessoas ou onde existe mudança na disposição da mobília, o mapeamento pode sofrer modificações significativas e a calibragem, nestes casos, deveria ser refeita. Estudos como em [3, 14] mostram a necessidade de modelos adaptativos para o mapeamento do RSSI, mesmo em ambientes aparentemente estáticos, pois os valores oscilam no tempo.

Em [1], os autores apresentaram uma técnica para a estimativa da localização sem o uso da fase de calibragem. Naquele trabalho, foi utilizado o padrão IEEE 802.11b e a

atenuação do sinal devido a obstáculos, no local onde foram aferidos os valores de sinal, não foi levada em consideração. Conforme será mostrado no presente artigo, levar em consideração esta atenuação na construção do modelo de localização, pode gerar considerável melhora na precisão.

3. Arquitetura Baseada em Cliente-Servidor

Quando se usa arquitetura cliente-servidor, o processo de estimativa de localização ocorre sempre em duas etapas, *off-line* e *on-line*. Na Figura 1(a) é possível ver um exemplo de uma rede sem fio onde um mecanismo de localização baseado no paradigma cliente-servidor foi implantado. Já a Figura 1(b) mostra um cenário baseado em *sniffers*. Na Figura 1(a), os clientes sem fio estão associados a um determinado ponto de acesso (AP - *Access Point*), mas recebem também sinal de outros APs dentro da área de cobertura. Neste ambiente, cada cliente mede o RSSI de cada um dos APs em sua área de cobertura e os envia na forma de um vetor $[RSSI_1, RSSI_2, \dots, RSSI_n]$ (onde n indica o índice do AP transmissor) para o servidor de localização. Este último compara o vetor recebido de cada cliente com as informações dos vetores de RSSI registrados no banco de dados, através de medidas de nível de sinal realizadas na fase de calibragem. Esta comparação é feita utilizando algum algoritmo específico; desta forma o servidor de localização estima a posição do cliente.

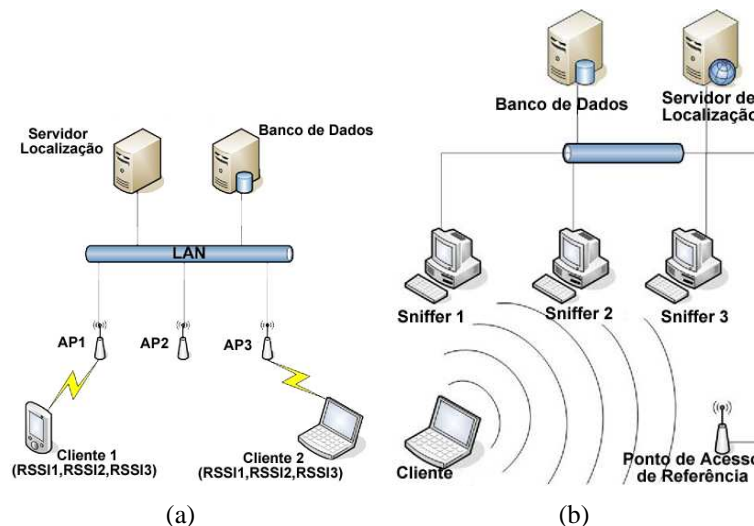


Figura 1. (a) Arquitetura baseada em Cliente-Servidor. (b) Arquitetura baseada em Sniffers.

4. Arquitetura Baseada em Sniffers

Sniffers são softwares que monitoram uma interface de rede e capturam todas as informações que trafegam por ela. Neste trabalho serão chamados *sniffers* entidades formadas por um computador (PC - *Personal Computer*), uma interface de rede sem fio instalada e configurada neste PC e um *software* que captura o tráfego nesta interface de rádio. Estas entidades podem estar co-aloçadas aos APs ou em máquinas distintas.

A principal motivação para o uso de *sniffers* na detecção e localização de dispositivos wi-fi, é a possibilidade de detectar qualquer dispositivo, seja ele um usuário ou um AP, legítimos ou não. Quando se utiliza uma arquitetura cliente-servidor, se o cliente, por algum motivo, não possuir o software instalado e executando, o sistema não será capaz de localizá-lo. Outro ponto negativo é a preocupação com consumo de energia. Existe a necessidade de que o usuário faça o *download* do software que monitora o RSSI e envie as informações para o servidor de localização. Este processamento extra pode ter influência

nos requisitos de energia dos dispositivos sem fio, o que é uma preocupação neste tipo de ambiente. Com o uso desta arquitetura, o cliente gera tráfego periodicamente, quando envia os dados recolhidos para o servidor de localização, aumentando a atividade na rede e a disputa pelo meio. Quando se utiliza *sniffers*, estas preocupações deixam de existir, pois tanto os *sniffers* quanto o servidor de localização podem estar ligados na rede cabeada e na rede elétrica, de forma que não haja limitações de energia.

4.1. O Sniffer

Os *sniffers* foram implementados para desempenhar duas tarefas na arquitetura proposta:

1. Detectar estações sem fio e aferir os valores de RSSI para cada uma;
2. Medir o nível de sinal recebido de um ou mais pontos de referência para construir o modelo de localização.

Estas ações são executadas simultaneamente e de forma ininterrupta ao longo do tempo. A primeira consiste em capturar pacotes de todas as interfaces sem fio detectadas pelo *sniffer* e extrair informações sobre RSSI e MAC do transmissor x . Para cada interface transmissora detectada, são capturados pacotes durante um determinado período. Esse período é definido aqui como *IC (Intervalo de Captura)* e foi utilizado nos experimentos apresentados com o valor de 1 segundo. Escolheu-se utilizar o RSSI médio medido durante um pequeno intervalo de tempo de forma a tirar vantagem da alta autocorrelação existente entre medidas de RSSIs dentro deste intervalo de tempo [4]. Cada pacote possui seu RSSI respectivo e cada *sniffer* calcula o RSSI médio a partir dos pacotes capturados durante o IC, para cada transmissor detectado. Em seguida, cada *sniffer* envia à base de dados a média calculada para cada transmissor detectado, repetindo sempre este procedimento a cada IC. Desta forma, no banco de dados serão armazenadas as tuplas $(RSSI_{1;x}, RSSI_{2;x}, \dots, RSSI_{i;x}, \dots, RSSI_{k;x}, MAC_x)$, onde $RSSI_{i;x}$ é a média dos RSSIs medidos a partir de todos os pacotes capturados pelo *sniffer* de índice i para o dispositivo transmissor detectado de índice x . Isso é feito em todos os k *sniffers* que monitoram o local.

Na segunda, são capturados pacotes de *beacon*¹ oriundos do(s) ponto(s) de acesso que compõe a rede sem fio, chamados aqui de *pontos de acesso de referência*. Os *sniffers* então extraem destes pacotes a informação de RSSI. Cada *sniffer* faz isso para cada ponto de acesso de referência M vezes. Assim, de posse de M medições de RSSI seguidas, o *sniffer* calcula média μ e desvio padrão σ destes M dados coletados, enviando em seguida para o banco de dados o par $(\mu_{i;n}, \sigma_{i;n})$, onde $\mu_{i;n}$ e $\sigma_{i;n}$ são respectivamente média e desvio padrão calculados pelo *sniffer* i a partir o ponto de acesso de índice n . Neste trabalho e na geração dos resultados foram utilizados $M = 1000$ medições de RSSI, 3 *sniffers* e apenas 1 ponto de acesso de referência.

4.2. O Servidor de Localização

O componente mais importante que integra o sistema proposto é o *Servidor de Localização*. Sua função é estimar a posição dos dispositivos sem fio detectados pelos *sniffers*. Este componente é um *software*, implementado em C/C++, que lê informações do banco de dados, as processa e fornece como saída a posição dos dispositivos sem fio detectados. Nele também está implementado o modelo de localização proposto.

O servidor de localização funciona da seguinte forma. Uma vez iniciado o programa servidor, ele acessa o banco de dados e lê as informações necessárias para montar o modelo de localização e aplicá-lo. Estas informações são:

¹Pacotes enviados pelo AP para sincronização da rede sem fio

1. Tamanho do local monitorado (X_{max}, Y_{max}).
2. Resolução do *grid*. Define-se aqui, *grid*, como sendo um conjunto de pontos que cobre o local monitorado, espaçados de uma distância fixa chamada resolução do *grid*. Neste trabalho a resolução utilizada foi de 1 metro.
3. Posição (x,y) do(s) AP(s) de referência (apenas 1 neste trabalho).
4. Posição (x,y) dos *sniffers*.
5. Endereços MAC de todos os dispositivos sem fio detectados pelos *sniffers*.
6. Níveis de sinal recebido a partir dos dispositivos detectados $RSSI_{i;x}$. Isso é feito para todos os *sniffers* e todos os dispositivos detectados por cada *sniffer*.
7. Par $(\mu_{i;n}, \sigma_{i;n})$ calculado para cada *sniffer* i com os RSSIs medidos a partir dos *beacons* do AP de referência de índice n . Isso é feito para todos os *sniffers* e todos os APs de referência.

De posse destes dados o servidor pode construir o modelo de localização e estimar a posição dos dispositivos detectados.

4.3. O Ponto de Acesso

O ponto de acesso, além de desempenhar as funções corriqueiras de um equipamento desta natureza, tais como permitir conectividade entre estações sem fio associadas a ele e funcionar como uma “ponte” para a rede cabeada, é também utilizado como ponto de referência para construção do mapa de propagação.

Este equipamento transmite pacotes de gerenciamento, controle e pacotes de dados [2]. No sistema proposto, os *sniffers* capturam todos esses pacotes, identificam quais foram transmitidos pelo ponto de acesso de referência e destes, extraem a informação de RSSI. Apenas *beacons* são utilizados para este fim. Desta forma, os *sniffers* calculam o RSSI médio e o desvio padrão, par $(\mu_{i;n}, \sigma_{i;n})$, como indicado anteriormente, para a posição onde o AP de referência está localizado. O ponto onde o AP de referência se encontra, será chamado no texto de l_0 e a exemplo do que acontece também com os *sniffers*, é fixo e conhecido.

5. Construção do Mapa de Propagação (MP)

Como dito anteriormente, o servidor de localização lê do banco de dados informações sobre o local monitorado, *sniffers* e ponto(s) de acesso(s) de referência. De posse destes dados, o servidor usa (X_{max}, Y_{max}) , a resolução do *grid*, as posições dos *sniffers* e do AP de referência e os valores $(\mu_{i;n}, \sigma_{i;n})$ de cada par *sniffer*+AP de referência, para construir um Mapa de Propagação (MP) do local monitorado. Para cada *sniffer*+AP de referência, o servidor de localização gera um mapa próprio, de forma que um total de k *sniffers* e N APs de referência, geram $(k \cdot N)$ MPs no servidor.

Esses mapas são compostos por um *grid* de duas dimensões que cobre toda a área do local monitorado e cujos pontos estão espaçados a uma distância igual a resolução do *grid*. É atribuída a cada local l , definido como sendo um ponto $[x, y]$ no *grid*, deste mapa uma distribuição de probabilidade $P(s|l)$, que define a probabilidade do *sniffer* medir, a partir do ponto fixo em que ele se encontra (lembrando que existe um mapa para cada par *sniffer*+AP de referência), um nível de sinal s dado que o transmissor está localizado em l . Em outras palavras, cada ponto l do mapa indica a probabilidade do *sniffer* em questão receber um nível de sinal s dado que um dispositivo hipotético estaria transmitindo a partir de l . Apesar do RSSI ser uma grandeza discreta, utilizou-se a distribuição Gaussiana para representar $P(s|l)$, em cada posição l :

$$P(s|l) = \frac{1}{\sigma_{(l)}\sqrt{2\pi}} \exp\left(-\frac{(s - \mu_{(l)})^2}{2\sigma_{(l)}^2}\right), \quad (1)$$

onde o parâmetro $\mu_{(l)}$ representa o valor esperado para o nível de sinal medido pelo *sniffer* em questão, dado que o transmissor está posicionado em l e $\sigma_{(l)}$ é o desvio padrão da distribuição. Para estimar o valor $\mu_{(l)}$ foi utilizado um modelo de propagação, tal que $\mu_{(l)}$ fosse dado em função da distância de l (local onde hipoteticamente se encontra o transmissor) até o *sniffer* (receptor).

A distribuição de probabilidades utilizada para representar os RSSI foi definida através de uma análise estatística obtida em [15], bem como em [6, 7, 10]. Esses trabalhos mostraram que podemos aproximar o valor de $P(s|l)$ por uma distribuição Gaussiana.

Desta forma, baseado no modelo de propagação dado pela Equação 2, podemos calcular $\mu_{(l)}$ para cada par *sniffer*+AP de referência, através de equação análoga:

$$\mu_{(l)}(d) = \mu_0(d_0) - 10n_0 \log\left(\frac{d}{d_0}\right) - \alpha, \quad (2)$$

onde α (chamado de WAF (*Wall Attenuation Factor*) em [8]) representa o valor da atenuação provocada por obstáculos entre o transmissor localizado no ponto l e o *sniffer*. Este parâmetro pode ser visto como um ajuste fino do modelo, para refletir de forma mais realista as características físicas do ambiente monitorado podendo assim incrementar consideravelmente a performance do sistema, independente do algoritmo de localização utilizado. O valor d representa a distância entre o transmissor localizado no ponto $l = (x, y)$ e o *sniffer*, d_0 é a distância entre o AP de referência localizado no ponto $l_0 = (x_0, y_0)$ e o *sniffer*, n_0 indica a taxa de decaimento do sinal transmitido proporcional à distância e $\mu_0(d_0)$ é a média dos valores de RSSI aferidos pelo *sniffer* i a partir de seu respectivo AP de referência de índice n . $\mu_0(d_0)$ é dado por $\mu_0(d_0) = \mu_{i;n}$. As distâncias d e d_0 são respectivamente por $d = \sqrt{(|x_{sniffer} - x|)^2 + (|y_{sniffer} - y|)^2}$ e $d_0 = \sqrt{(|x_{sniffer} - x_0|)^2 + (|y_{sniffer} - y_0|)^2}$.

Sendo o valor de $\mu_{(l)}$ dado pela Equação 2, resta definir o valor do desvio padrão $\sigma_{(l)}$ da distribuição de RSSI para cada posição l do transmissor. Assim como $\mu_{(l)}$, $\sigma_{(l)}$ também muda com a posição l do transmissor, no entanto, utilizou-se aqui $\sigma_{(l)} = \sigma_{i;n}$. O valor de $\mu_{(l)}$ pode ser estimado através de um modelo de propagação, mas o valor de $\sigma_{(l)}$ poderia ser determinado apenas através de medições empíricas realizadas em cada ponto l do local monitorado. Desta forma, para evitar essas medições (como na fase de calibragem), utilizou-se $\sigma_{(l)} = \sigma_{i;n}$. Essa igualdade fixa o valor de $\sigma_{(l)}$ para todo l como sendo o valor do desvio padrão calculado através dos níveis de sinal recebidos pelo *sniffer* i a partir do AP de referência n . Os resultados experimentais apresentados na Seção 8 mostram que essa é uma boa aproximação para $\sigma_{(l)}$.

Supondo que em cada ponto l do *grid* existe um transmissor hipotético, deseja-se encontrar $\mu_{(l)}$, que determina o valor esperado de nível de sinal medido pelo *sniffer* i , para cada um destes transmissores.

5.1. Reconstrução do MP

Em trabalhos anteriores e nas medidas de RSSI claramente verifica-se que o RSSI segue uma distribuição normal cuja média μ e desvio padrão σ podem ser estimados para cada par *sniffer* + AP de referência. Assim, para o sistema proposto, é possível reconstruir o MP para cada par *sniffer* + AP de referência sempre que o *sniffer* k observar que RSSIs medidos a partir do seu ponto de acesso de referência i apresentam alterações estatísticas significativas. Em outras palavras, caso n em N medidas de RSSI seguidas estiverem fora do intervalo $(\mu_{k,i} + \sigma_{k,i}, \mu_{k,i} - \sigma_{k,i})$, um novo MP deve ser construído.

Uma alternativa ao procedimento de atualização do MP descrito acima é fazê-lo a cada T segundos. Ou seja, periodicamente reconstruir o MP independente da existência de alterações estatísticas significativas. O motivo de se utilizar a reconstrução periódica está no fato de que caso não exista necessidade de reconstruir o MP (média e desvio padrão não têm alterações significativas), a reconstrução ocorrerá assim mesmo e os novos valores de $\mu_{k,i,t}$ e $\sigma_{k,i,t}$ serão muito próximos dos valores de $\mu_{k,i,t-T}$ e $\sigma_{k,i,t-T}$. Caso fosse necessário atualizar o MP devido a alterações nos valores de $\mu_{k,i,t}$ e $\sigma_{k,i,t}$ essa necessidade duraria, no pior caso, T segundos.

6. O Modelo de Localização Proposto

Antes de descrever o problema de localização de um dispositivo sem fio são necessárias algumas definições. Seja L um espaço físico bidimensional, a partir de cada posição $l \in L$ é possível obter medições de nível de sinal de k *sniffers* (receptores), dado um transmissor posicionado em l . Neste trabalho assume-se que L é discreto. Define-se assim, um espaço amostral S de k dimensões, onde cada elemento deste espaço é um vetor de dimensão k , cujas posições representam leituras de níveis de sinal feitas por k diferentes *sniffers*. Desta forma, amostras do espaço S são referenciadas por s .

O problema então, pode ser descrito como de máximo a *posteriori*, ou seja, dado um vetor de medições de nível de sinal $s = (s_1, s_2, \dots, s_k)$, deseja-se determinar uma posição $l \in L$, tal que a probabilidade $P(l|s)$ seja maximizada. Pode-se dizer de forma mais explícita que $P(l|s)$ é a probabilidade de um transmissor estar localizado fisicamente em uma posição l , dado que medidas de nível de sinal s foram feitas por k *sniffers* diferentes.

6.1. Estimador Proposto

A idéia por trás do método proposto aqui é a de que não se pode esperar nenhuma colaboração por parte dos dispositivos móveis cuja posição deseja-se estimar. O que se faz é medir o nível de sinal recebido por k *sniffers* a partir do dispositivo que se deseja localizar. De posse deste conjunto de dados (medições realizadas pelos *sniffers*) o servidor de localização pode estimar a posição do dispositivo através do modelo descrito a seguir.

Dado um vetor de medições $s = (s_1, s_2, \dots, s_k)$, deseja-se encontrar uma posição $l \in L$, tal que a probabilidade $P(l|s)$ seja maximizada. Aplicando-se a regra de *Bayes*, minimiza-se a probabilidade de erro e é possível encontrar a distribuição a *posteriori* da localização:

$$P(l|s) = \frac{P(s|l) \cdot P(l)}{P(s)} = \frac{P(s|l) \cdot P(l)}{\sum_{l' \in L} P(s|l') \cdot P(l')}; \quad (3)$$

onde o somatório segue através de todos os valores possíveis para posições $l \in L$. $P(s|l)$ é a probabilidade de um *sniffer* receber o sinal s , dado que o mesmo foi supostamente transmitido a partir de um local l . Em caso de um espaço L contínuo, o somatório deverá ser substituído pela integral correspondente. $P(l)$ por sua vez, é a probabilidade a *priori* de se encontrar o transmissor em uma posição l , antes de conhecidas as leituras de níveis de sinal s , por parte dos *sniffers*. Este parâmetro pode ser usado em um mecanismo de localização para dar “peso” a determinadas posições mais prováveis de se encontrar um transmissor. Esta probabilidade a *priori* possibilita uma forma simples de incorporar ao sistema, informações sobre padrão de mobilidade e/ou rastreamento. A probabilidade $P(l)$ poderia ser determinada também por perfis de mobilidade, atendo-se ao fato de que se um usuário estivesse posicionado em um dado local, seria mais provável localizá-lo em alguma posição adjacente em um futuro próximo. Caso o perfil do usuário seja

desconhecido ou simplesmente não seja considerado, pode-se assumir que é igualmente provável encontrá-lo em qualquer das localidades $l \in L$. Sendo assim, $P(l)$ seria dada uniformemente sem apresentar tendências de encontro em qualquer posição l particular.

$P(s|l)$ foi estimada para cada local $l \in L$, construindo-se assim um mapa, chamado aqui de *mapa de propagação (MP)*. Cada posição deste mapa possui a distribuição de probabilidade dos k *sniffers* medidos os níveis de sinal $s = (s_1, s_2, \dots, s_k)$, dado que um transmissor esteja posicionado neste local l do mapa. Na realidade, cada *sniffer* i possui seu próprio mapa de propagação. Cada posição deste mapa possui uma distribuição de probabilidade $P(s_i|l)$ que denota a probabilidade deste *sniffer* medir o nível de sinal s_i , dado que o transmissor estivesse, supostamente localizado na posição l . Neste trabalho, para estimar $P(s_i|l)$, em cada posição, foi utilizada uma distribuição Gaussiana, como já mencionado anteriormente.

Os valores $\mu_{(l)}$ e $\sigma_{(l)}$ são dados em função da posição l no mapa e denotam o nível de sinal médio e o desvio padrão percebido para esta posição, respectivamente. Estes valores, para cada posição do MP, são calculados a partir de um modelo de propagação escolhido. Os parâmetros deste modelo são estimados por medições realizadas pelo *sniffer* i a partir do sinal transmitido pelo ponto de acesso de referência. Desta forma, assumindo que as medições realizadas pelos *sniffers* são independentes, pode-se escrever que:

$$P(l|s) = P(l|s_1, s_2, \dots, s_k) = P(l|s_1) \cdot P(l|s_2) \cdot \dots \cdot P(l|s_k) = \prod_{i=1}^k P(l|s_i); \quad (4)$$

onde $P(l|s_i)$ é dado pela equação 3. A equação 4 denota a probabilidade de um transmissor estar localizado na posição l , dado que k *sniffers* mediram os níveis de sinal a partir deste transmissor, criando o vetor $s = (s_1, s_2, \dots, s_k)$. O resultado do estimador proposto poderá ser a posição $l \in L$ que proporcionar à Equação 4 o maior valor.

7. Precisão do Sistema

Nas seções anteriores, foi descrita a técnica proposta para localizar dispositivos sem fio, baseada no nível de sinal recebido. Este método diferencia as possíveis posições no *grid* do local monitorado, calculando a probabilidade do dispositivo transmissor em questão estar localizado em cada um destes pontos. A Figura 2(a) mostra um exemplo da distribuição de probabilidade da localização de um dispositivo, calculada pela técnica proposta durante uma tentativa real de localização. Ainda assim, uma última questão permanece em aberto: *baseado nestes cálculos, qual será a saída do estimador?* A técnica utilizada na estimativa de posição baseada na probabilidade $P(l|s)$ é de fundamental importância para a precisão do sistema. Desta forma, foram utilizadas duas técnicas para estimar de forma mais acurada a localização dos dispositivos: *Janela de Estimativas* e *Centro de Massa*.

7.1. Janela de Estimativas

A primeira proposta para melhoria na precisão baseia-se em escolher o ponto l do *grid* mais provável de se encontrar o transmissor, dado um vetor de medições $s = (s_1, s_2, \dots, s_k)$, para k *sniffers*. O exemplo observado na Figura 2 mostra o *grid* formado por pontos espaçados de 1 metro, sobre o local monitorado de dimensões $16m \times 10m$. É possível observar que o ponto de maior probabilidade é o $(0, 9)$, resultado da estimativa. Esse resultado é então computado como sendo a saída do sistema.

Na medida em que os *sniffers* medem novos valores de RSSI, as probabilidades em cada ponto do *grid* são recalculadas e modificadas ao longo do tempo. Em algumas

ocasiões, diversos pontos l diferentes podem possuir valores de $P(l|s)$ próximos, sendo que o maior valor pode, na verdade, ser encontrado em um ponto distante da posição real do dispositivo sem fio que se deseja localizar, o que caracteriza um erro grande. Um exemplo deste fenômeno pode ser visto na Figura 2(b).

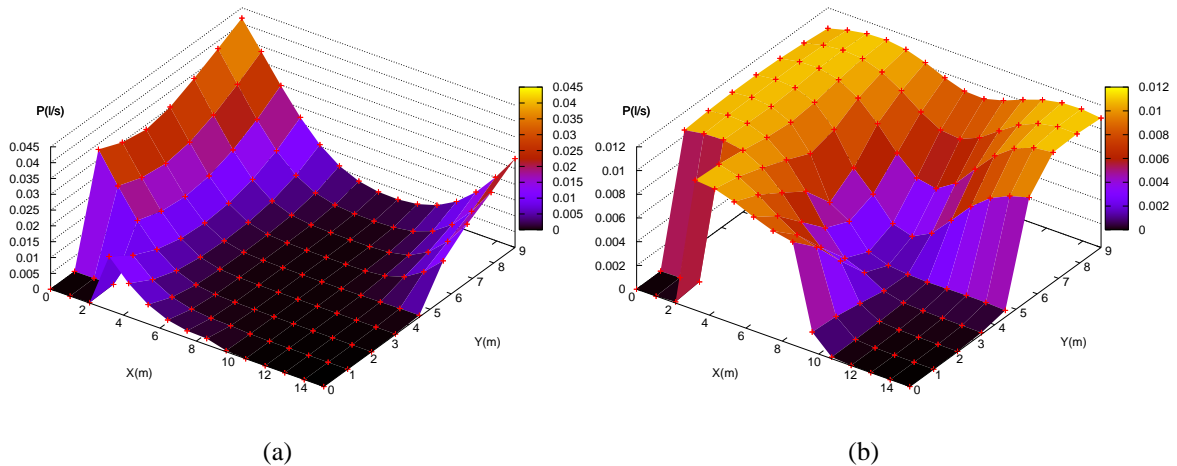


Figura 2. (a) Distribuição de probabilidade de se encontrar o dispositivo transmissor em algum ponto l do *grid* sobre o local monitorado, calculada pela técnica proposta. Cada ponto possui uma probabilidade do transmissor estar em l , dado que os k *sniffers* mediram os níveis de sinal $s = (s_1, s_2, \dots, s_k)$. (b) Distribuição de probabilidade de se encontrar o dispositivo transmissor em algum ponto l do *grid* sobre o local monitorado, quando ocorre um erro na estimativa de posição.

Durante experiências realizadas, um dispositivo sem fio foi colocado em 10 posições diferentes escolhidas aleatoriamente no laboratório. Para cada uma destas posições foram feitas 500 estimativas de posição pelo sistema, gerando um total de 5000 estimativas de localização. As figuras 2 (a) e (b) foram geradas a partir de duas estimativas espaçadas por apenas 1 segundo (nesta experiência o sistema gerava uma estimativa a cada 0.5 segundos), o que mostra a grande volatilidade do sinal medido e a capacidade de adaptação do mecanismo proposto. A situação de erro, como observada na Figura 2(b), não permanece por muito tempo e estatisticamente ocorre poucas vezes comparando-se com situações de erro pequeno, como pode ser visto a seguir.

Notou-se também que em situações onde o erro é grande, como na Figura 2(b), a probabilidade $P(l|s)$ é pequena e que quando esta probabilidade é grande, existe um destaque maior para um determinado ponto do *grid* (como na Figura 2(a)) sendo o erro é pequeno. Isso pode ser visto claramente na Figura 3. Quando a estimativa de posição tem probabilidade maior, o erro é menor e quando o erro é maior, a probabilidade é menor. Não se pode afirmar nada quando a probabilidade é baixa. Porém, esses resultados motivaram uma variação da proposta original.

Cada vez que o sistema proposto gera uma saída (grava no banco de dados coordenadas $[x, y]$ da posição l estimada), ele deve calcular as probabilidades $P(l|s)$ para todo $l \in L$ e gerar dados como observados nas Figuras 2(a) e 2(b). Assim, cada estimativa gera uma saída. Foi proposta então, uma variação que consiste em gerar uma saída a cada W estimativas, de forma que W seria o tamanho da Janela de Estimativas. Ou seja, para $W = 10$, seriam geradas 10 estimativas de posição l . O sistema então, verificaria qual destas 10 estimativas possui o maior valor de probabilidade $P(l|s)$ e apenas essa posição seria escolhida como saída do sistema, de forma que $W = 10$ corresponde a uma

saída para cada 10 estimativas e $W = 1$ é o cenário original onde cada estimativa gera uma saída. O valor de W seria então um parâmetro do sistema utilizado para “ajustar” o compromisso entre precisão e tempo de resposta.

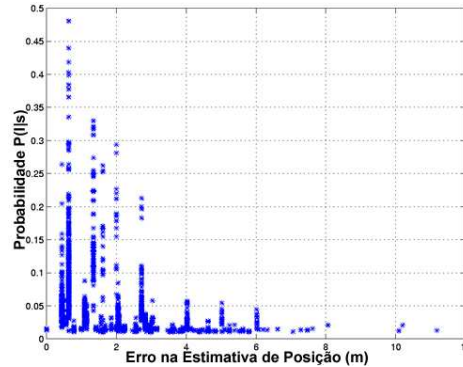


Figura 3. Dispersão das 5000 estimativas. Cada estimativa está associada a uma probabilidade $P(l|s)$ e a um erro.

7.2. Centro de Massa

A técnica descrita na seção anterior, pode ser considerada um estimador discreto, pois aponta como saída do sistema apenas pontos sobre o *grid*. Já a técnica apresentada nesta seção, pode ser considerada um estimador contínuo, por fornecer como saída qualquer coordenada $[x, y]$, para x e y contínuos. O *Centro de Massa* CoM (*Center of Mass*) foi utilizado previamente em [10] e também aqui, como o estimador contínuo do sistema.

A idéia básica por trás desta técnica é tratar cada local l do *grid* como um objeto no espaço físico, cuja massa é igual a probabilidade normalizada $P(l|s)$ calculada para todo $l \in L$, através do modelo proposto na Seção 6.1. Por simplicidade, a probabilidade $P(l|s)$ será tratada aqui como m ($P(l|s) = m$). Assim, sendo m_i a massa do local l_i , podemos definir a saída do sistema como sendo um local Z (o centro de massa) dado pela Equação 5, tal que \bar{L} é uma lista de todos os locais do *grid*, ordenados de forma decrescente de acordo com a probabilidade normalizada de cada um. Através desta equação, pode-se obter o centro de massa para os N objetos (locais) de maior massa, onde N é um parâmetro do sistema, tal que $1 \leq N \leq \|\bar{L}\|$, e onde $m_1 > m_2 > \dots > m_N > \dots > m_{\|\bar{L}\|}$. O leitor deve notar que para o caso particular onde $N = 1$, a técnica de CoM é equivalente a Janela de estimativa para $W = 1$.

$$Z = \left(\sum_{i=1}^N m_i \cdot \bar{L}(i) \right) / \left(\sum_{i=1}^N m_i \right) \quad (5)$$

8. Resultados Experimentais

Para avaliar as técnicas propostas, um ambiente de testes foi montado em laboratório. Interfaces sem fio foram instaladas em 3 estações de trabalho sob o sistema operacional Linux. O *software* do *sniffer* foi instalado e configurado em cada uma dessas estações. Uma quarta estação de trabalho Linux foi utilizada para desempenhar as funções do servidor de localização e do banco de dados. Um ponto de acesso 3COM modelo 3CRWE454A72 foi utilizado como AP de referência. Todas estas estações já estavam interconectadas por uma infraestrutura de rede cabeada pré-existente, de forma que apenas as interfaces sem fio precisaram ser instaladas. Os dispositivos sem fio clientes, cujas posições seriam estimadas, foram representados por um *notebook* com interface PCMCIA e por uma estação de trabalho com interface PCI.

Este ambiente foi montado dentro de um local de dimensões $16m \times 10m$, cujo mapa e disposição dos equipamentos podem ser vistos na Figura 4. Nesta figura, os *sniffers* são representados por quadrados, o ponto de acesso por um triângulo e os dispositivos sem fio detectados (dispositivos cliente), por um círculo. No exemplo desta figura, apenas um dispositivo cliente foi encontrado e localizado. As posições dos *sniffers* e do ponto de acesso são conhecidas e não se alteram durante os experimentos.

Para coleta dos resultados apresentados neste capítulo, os dispositivos clientes sem fio (*notebooks* e estações de trabalho) foram dispostos em 10 posições $[x, y]$ diferentes dentro do perímetro do laboratório e suas posições foram estimadas pelo sistema proposto 500 vezes para cada posição, através de cada uma das técnicas propostas na Seção 7 e suas variações. Ou seja, 500 saídas do sistema foram dadas para cada uma das 10 posições, totalizando 5000 estimativas de posição para cada uma das técnicas utilizadas ($W=10$ e CoM).

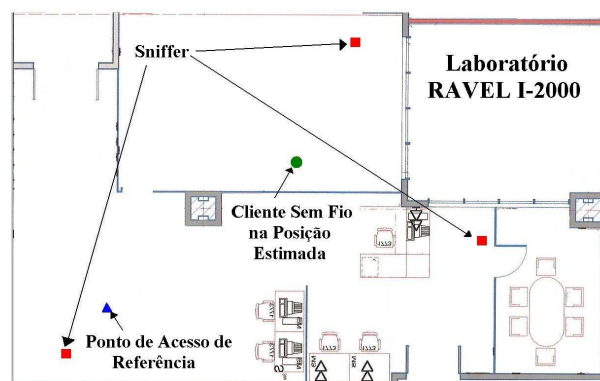


Figura 4. Mapa do local monitorado e disposição dos equipamentos sem fio.

8.1. Cálculo do Valor de WAF para o Ambiente Monitorado

O local monitorado possui obstáculos que atenuam o sinal, causando redução da precisão na medição da localização do dispositivo móvel. Como descrito anteriormente, o modelo possui um parâmetro chamado *Wall Attenuation Factor (WAF)*, responsável pela correção da atenuação do sinal causado pelos obstáculos no ambiente.

Os obstáculos presentes no local monitorado são placas de madeira compensada de 4 cm de espessura. O valor do parâmetro WAF, para o ambiente, foi determinado experimentalmente de acordo com o seguinte processo trivial: um dispositivo, equipado com uma interface de rede sem fio (Cliente) foi posicionado a uma distância de 4.5 metros de uma estação de trabalho, equipado, também, com uma interface sem fio (*Sniffer*). O cliente gerou constantemente tráfego e o *sniffer* capturou este tráfego extraindo os valores de RSSI de cada pacote. Um total de 500 valores de RSSI foram aferidos e a média dos valores RSSI foi determinada, apontando o valor -39.5dBm . Em seguida, uma parede de madeira compensada de 4 cm, igual a todas do ambiente utilizado no experimento, foi introduzida na distância média entre o cliente e o *sniffer*. Outra vez, 500 valores de RSSI foram capturados, apresentando uma média de -42.0 dBm . Com isso, obtivemos o decaimento do sinal, devido a parede de madeira compensada de -2.5 dBm .

8.2. Avaliação da Precisão do Sistema

Uma vez montado o ambiente descrito na seção anterior e de posse dos dados de localização gerados a partir dos experimentos descritos acima, foi necessário definir uma métrica para se avaliar a precisão do estimador proposto e das técnicas de localização

utilizadas. Assim, a métrica utilizada neste trabalho para avaliação do sistema foi, naturalmente, o erro na estimativa de localização. Este erro foi definido como a distância (em metros) entre o ponto $[x, y]$ indicado na saída do estimador e a real posição do dispositivo sem fio. Deseja-se assim que o erro seja o menor possível.

Em [1], foi mostrado que os resultados experimentais da técnica chamada Janela de Estimativas possui um resultado consideravelmente melhor quando utilizado o tamanho da janela com valor 10 ($W=10$) em detrimento ao valor 1 ($W=1$). Por este motivo, nos experimentos apresentados aqui, foi utilizado apenas o tamanho da janela de estimativas igual a 10. A Tabela 1 mostra os parâmetros utilizados no experimento.

Tabela 1. Parâmetros utilizados para gerar os resultados presentes.

Parâmetro	Valor
α	$= 0$ e > 0
n_0	2
T	10 segundos
Resolução do Grid	1 metro
(X_{max}, Y_{max})	(16m, 10m)
M	60 beacons
W	10
IC	1 segundo

Foram traçadas então, de forma empírica, as funções de distribuição cumulativas (ECDF - *Empirical Cumulative Distribution Function*) do erro na estimativa de localização para cada uma das técnicas descritas na Seção 7. A Figura 5(a) mostra estas distribuições.

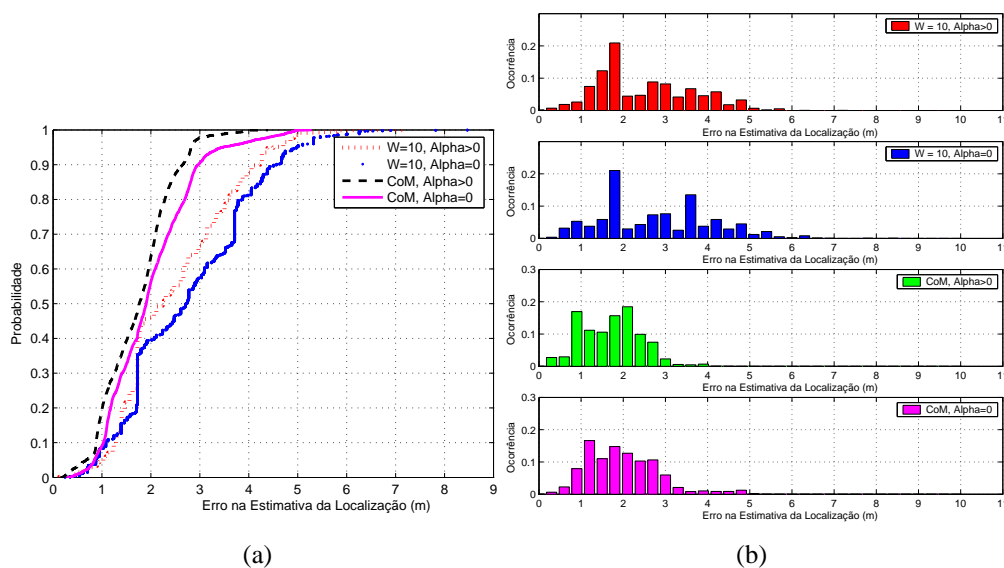


Figura 5. (a) Distribuição cumulativa empírica de probabilidade do erro na estimativa de posição. (b) Histograma normalizado do erro na estimativa de posição para variações nas técnicas de estimativa.

O comportamento de "escada" da curva $W=10$, difere do comportamento mais suave da curva CoM, pois a primeira é variação de um estimador discreto e a última mostra o resultado de um estimador contínuo, onde o erro pode assumir infinitos valores. No caso do estimador discreto, para o *grid* utilizado nestes experimentos com 160 pontos ($16m \times 10m = 160$, para uma distância entre pontos de 1m), dada a posição real do transmissor, o valor de erro na estimativa poderá assumir 160 valores diferentes, uma vez que o erro é dado pela distância entre a posição real e o ponto indicado pelo estimador e que o estimador discreto só indica como saída pontos do *grid*.

É importante citar que os experimentos foram realizados durante horário comercial, onde existia constante movimentação de pessoas dentro do local monitorado. Os resultados apresentados demonstram a grande capacidade que o método de localização proposto tem para se adaptar às variações do RSSI. Mesmo com o trânsito de pessoas, o sistema apresentou erro abaixo de 2.52 metros durante 75% do tempo e abaixo de 3.0 metros 90% do tempo, quando se utiliza CoM.

A Figura 5(b) ilustra a frequência com que o erro ocorre para as diferentes técnicas utilizadas. Percebe-se que para erros menores ou iguais a 1 metro, quando CoM é considerado, temos aproximadamente 20% de estimativas nessa faixa, quando o parâmetro $\alpha > 0$, contra aproximadamente 10% quando $\alpha = 0$. Para erros menores ou iguais a 2 metros, o uso do parâmetro $\alpha > 0$ aumenta em 65% o número de estimativas nesta faixa quando considerado o CoM, contra 56% quando $\alpha = 0$. É possível observar que o uso do parâmetro $\alpha > 0$ diminui drasticamente o erro na localização, especialmente quando usada a técnica CoM. A tabela 2 apresenta os valores de 50, 75 e 90 percentil da distribuição cumulativa empírica do erro de localização, dados em metros.

Tabela 2. Valores de 50, 75 e 90 percentil da distribuição cumulativa empírica do erro de localização (m).

Técnica	50	75	90	Média
W=10 ($\alpha > 0$)	2.24	3.33	4.22	2.48
W=10 ($\alpha = 0$)	2.72	3.71	4.57	2.77
CoM ($\alpha > 0$)	1.76	2.17	2.65	1.72
CoM ($\alpha = 0$)	1.91	2.51	2.97	1.98

9. Conclusões e Trabalhos Futuros

O objetivo principal deste trabalho foi melhorar a precisão de um sistema de localização livre de fase de calibragem, estimando-se o decaimento de sinal provocado por obstáculos entre o ponto de acesso e o *sniffer* sem fio. O modelo analítico de propagação de sinal utilizado para construção do modelo de localização, agora é capaz de refletir de forma mais realista o ambiente monitorado. Esta simples modificação melhora a precisão em até 13% em relação ao erro médio (1,98m), quando não se utiliza qualquer informação sobre o ambiente monitorado.

Em comparação com [1], observamos que os resultados são inferiores no presente trabalho. Atribui-se este fato, ao uso de hardware diferente, pois utilizou-se rádios sob padrão IEEE 802.11g ao invés do padrão IEEE 802.11b como os usados em [1]. A forma como são representados os valores de RSSI nas placas sem fio diferem dependendo do fabricante. Além disso, a disposição diferente dos *sniffers* no ambiente monitorado contribui para resultados diferentes. Concluimos que não se pode comparar os valores obtidos neste artigo com os valores de [1].

O sistema foi implementado em um ambiente real de produção, de forma que nenhum aspecto da rotina dos usuários tivesse sido afetado de qualquer maneira. A partir deste trabalho, torna-se interessante investigar o impacto do posicionamento de *sniffers* e pontos de acesso de referência na precisão, bem como, a influência da quantidade destes dispositivos no desempenho do sistema.

Por fim, a implementação de um mecanismo que utiliza fase de calibragem, como o RADAR, por exemplo, no mesmo ambiente onde foi testado o método proposto seria interessante para que resultados de precisão pudessem ser comparados efetivamente de forma quantitativa. O mapa de propagação gerado na fase de calibragem poderia também

ser comparado aquele gerado dinamicamente, através do uso de modelos de propagação em RF, e ainda utilizado para ajustar os parâmetros α e n_0 , otimizando-os para um ambiente específico e contribuindo também para melhoria na precisão.

Referências

- [1] B. A. A. Nunes and L. F. M. de Moraes, "Calibration-Free WLAN Location System Based on Dynamic Mapping of Signal Strength," *The 4-th ACM International Workshop on Mobility Management and Wireless Access - MobiWac2006*, Outubro 2006.
- [2] IEEE 802.11, *IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1999.
- [3] P. Krishnan, A. Krishnakumar, W.-H. Ju, C. Mallows, and S. Ganu, "A system for LEASE: location estimation assisted by stationary emitters for indoor RF wireless networks," *IEEE Infocom*, Março 2004.
- [4] M. Youssef and A. K. Agrawala, "Handling samples correlation in the horus system," *IEEE Infocom*, Março 2004.
- [5] R. Battiti, T. L. Nhat, and A. Villani, "Location-aware computing: A neural network model for determinating location in wireless LANs," University of Trento, Tech. Rep. DIT-02-0083, Fevereiro 2002.
- [6] M. Youssef and A. Agrawala, "WLAN location determination via clustering and probability distributions," *IEEE PerCom*, Fevereiro 2003.
- [7] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen, "A probabilistic approach to wlan user location estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, Julho 2002.
- [8] P. Bahl and V. N. Padmanabhan, "RADAR: An In-building RF-based User Location and Tracking System," *IEEE INFOCOM*, vol. 2, pp. 775–784, Março 2000.
- [9] P. Bahl, A. Balachandran, and V. N. Padmanabhan, "Enhancements to the RADAR user location and tracking system," *Microsoft Technical Report*, 2000.
- [10] M. Youssef and A. K. Agrawala, "The horus wlan location determination system," *Proceedings of the Third International Conference on Mobile Systems, Applications, and Services - MobiSys*, Junho 2005.
- [11] A. Taheri, A. Singh, and E. Agu, "Location fingerprinting on infrastructure 802.11 wireless local area networks (WLANs) using locus," *Fourth International IEEE Workshop on Wireless Local Networks*, Novembro 2004.
- [12] S. Saha, K. Chaudhuri, D. Sanghi, and P. Bhagwat, "Location determination of a mobile device using IEEE 802.11b access point signals," *IEEE (WCNC)*, Março 2003.
- [13] *Ekahau Positioning Engine*, 2006. [Online]. Available: <http://www.ekahau.com/>
- [14] S. Ganu, A. S. Krishnakumar, and P. Krishnan, "Infrastructure-based location estimation in wlan networks," *IEEE (WCNC)*, Março 2004.
- [15] B. A. A. Nunes, "Um sistema de localização para redes wi-fi baseado em níveis de sinal e modelo referenciado de propagação," *Dissertação Mestrado - COPPE/UFRJ*, Maio 2006.