

ESTIMATIVA DE EMBARQUE E DESEMBARQUE DE PASSAGEIROS POR DETECÇÃO DE SINAL DE WI-FI DE DISPOSITIVOS MÓVEIS

Diego Benites Paradedá

Werner Kraus Junior

Rodrigo Castelan Carlson

Universidade Federal de Santa Catarina

Programa de Pós Graduação em Engenharia de Automação e Sistemas

RESUMO

Apresenta-se método baseado em detecção de dispositivos móveis Wi-Fi, tipicamente *smartphones*, para estimação de embarques e desembarques de passageiros do transporte público. Inicia-se pela descrição de equipamento detector de *smartphones* que registra sinalizações de dispositivos em busca de conexão, associando-os aos respectivos endereços identificadores únicos. Experimentos controlados são realizados para verificação da acurácia do detector, medida pelo tempo até a primeira detecção e pelo intervalo entre sucessivas detecções do mesmo dispositivo. Em seguida, aplica-se o detector em ponto de parada de ônibus movimentado. Detecções de dispositivos são combinadas com dados de rastreamento de ônibus para estimação dos embarques e desembarques. Levantamento manual de sobe-e-desce no ponto de parada serve de referência para o método automático de estimação. Resultados de campo indicam a validade da proposta, sujeita a limites estatísticos do número de usuários do transporte público portando *smartphones* com interface de rede Wi-Fi ativada.

ABSTRACT

We present a method based on the detection of Wi-Fi mobile devices, typically smartphones, for estimation of public transport boarding and alighting. It starts with the description of the smartphone detector equipment that captures signaling for connection by nearby devices, associating them with their unique identifier address. Controlled experiments are performed to verify the accuracy of the detector, measured by the time to first detection and by the interval between successive detections of the same device. Next, the detector is applied in a busy bus stop. Device detections are combined with bus tracking data for estimation of boarding and alighting. A manual survey at the stop serves as a reference for the automatic estimation method. Results indicate the validity of the proposal, subject to statistical limits on the number of users with smartphones with enabled Wi-Fi interfaces.

1. INTRODUÇÃO

A confiabilidade do serviço de transporte público por ônibus é bastante afetada pela imprevisibilidade do tempo perdido quando o veículo realiza uma parada em ponto. O fator que mais impacta nesse tempo é a duração de embarques e desembarques. Assim, o conhecimento do volume de passageiros e das origens e destinos ao longo de diferentes seções de uma linha de transporte permite tornar mais preciso o planejamento e operação desta (Arhin *et al.*, 2016).

De acordo com Vuchic (2005), dados detalhados sobre o volume de passageiros em linhas de transporte público podem ser obtidos por contagens de embarque e desembarque (pesquisa “sobe-e-desce”) em cada ponto de parada ao longo do trajeto. Maneira usuais de realizar esse tipo de contagem consistem em colocar pesquisadores em cada ponto ou embarcados em cada ônibus para realizar a pesquisa (O’Flaherty, 1997).

A coleta de dados feita manualmente exige esforço considerável e tem custo alto. Por conta disso, não costuma ser realizada com a frequência necessária para atualização de dados, a despeito da importância que têm para o planejamento e a operação do transporte coletivo. Assim, é desejável o desenvolvimento de novas técnicas de contagem de passageiros, que reduzam o esforço necessário para a coleta de dados e o custo associado.

Com a disseminação de dispositivos móveis como *smartphones*, surge a oportunidade de estimação de informações de usuários por intermédio da tecnologia Wi-Fi com a qual estão equipados. Note-se que no Brasil 86% das pessoas utilizam Wi-Fi fora de suas residências (Barbosa, 2016), número que ultrapassa a utilização dos sistemas 3G ou 4G para acesso à internet. A predominância de uso de conexão Wi-Fi deve-se, principalmente, pela oferta gratuita de acesso em muitos locais e pelo consumo de energia dos dispositivos ser menor no caso do Wi-Fi.

O interessante da tecnologia Wi-Fi é que pode ser usada de forma passiva, isto é, sem a necessidade de intervenção do usuário. Basta que o Wi-Fi esteja ativado para que o monitoramento do tráfego de requisições e de pacotes em redes sem fio Wi-Fi (NTM, do inglês *Network Traffic Monitoring*) detecte requisições de conexão em pontos de acesso de internet. A detecção de requisições combinada com informação de localização de ônibus permite, em princípio, identificar a partida e a chegada de passageiros. Nessa operação, obtém-se o identificador único do aparelho do usuário (endereço MAC) sem que a identidade deste seja revelada. Entretanto, no caso de haver cadastros digitais que associem o MAC a uma pessoa, deve-se anonimizar a informação coletada para garantia da privacidade do usuário. No Brasil, não há legislação sobre a coleta de dados Wi-Fi de usuários; já nos Estados Unidos existe legislação que regulariza essa coleta para fins acadêmicos e em empresas para proteção da anonimidade dos usuários e para limitação da quantidade de informação coletada (Sicker *et al.*, 2007).

A localização e velocidade dos ônibus podem ser obtidas por meio de localização automática de veículos (AVL, do inglês *Automated Vehicle Location*), que fornece os dados com base em rastreamento por satélite. O processamento em tempo real dos dados de AVL permite, por exemplo, realizar previsões de chegadas em pontos de parada (Cathey e Dailey, 2003). A dificuldade é saber a demanda de passageiros que está aguardando em pontos para embarcar, a quantidade de passageiros a bordo, e também a demanda por desembarques em cada ponto.

Este artigo trata da estimação do sobe-e-desce em pontos de parada com método baseado em amostras de dados de Wi-Fi e informações de chegadas e partidas de ônibus dos pontos. O objetivo é determinar o nível de confiança que se pode obter com o uso das tecnologias Wi-Fi e AVL. Para tanto, dois tipos de experimentos foram realizados. O primeiro experimento determinou a taxa de detecção de dispositivos, sendo realizado em cenário controlado no qual conhecem-se todos os dispositivos na área de detecção. O segundo experimento foi realizado em cenário real para determinação do número de detecções em relação aos volumes de sobe-e-desce obtidos por pesquisadores. No primeiro experimento observou-se que todos os dispositivos são detectados pelo menos uma vez, e que o intervalo médio entre detecções sucessivas do mesmo dispositivo é de 68 s. No segundo experimento, foram detectados 80% dos embarques e desembarques de passageiros que ocorreram no ponto.

A apresentação do trabalho estrutura-se da forma seguinte. Na Seção 2 são apresentados trabalhos relacionados da literatura. A Seção 3 apresenta os dispositivos usados na construção do equipamento detector e os procedimentos aplicados nos experimentos. A Seção 4 apresenta o experimento realizado para determinação da acurácia do detector e a aplicação em campo. Por fim, a Seção 5 faz um sumário das principais conclusões do estudo.

2. COLETA DE DADOS A PARTIR DE DISPOSITIVOS MÓVEIS

A primeira forma de se obter dados de trajetos de usuários de smartphones é por meio do próprio sinal da operadora de telefonia. Çolak *et al.* (2015), Jiang *et al.* (2017), Bekhor *et al.* (2013), Zhang *et al.* (2010), Lind *et al.* (2017) e Alexander *et al.* (2015) determinaram trajetórias de usuários por meio de dados coletados das informações dos dados das operadoras de telefonia. Os dados usados por esses autores foram gravados por operadoras, tais como horário que um dispositivo foi registrado em uma antena, horário de ligação, horário que recebeu uma mensagem, e horário que acessou a internet por de 3G ou 4G. Uma vantagem dessa abordagem é que, tipicamente, este modo de operação do dispositivo móvel (*smartphone*) está sempre habilitado e não precisa contar com intervenção do usuário. Por outro lado, tais dados não são armazenados por padrão em todos os países e tampouco são públicos.

No Brasil, informações de acesso à telefonia celular são armazenadas apenas em casos específicos e por interesse da operadora. Por exemplo, dados coletados durante os jogos olímpicos do Rio de Janeiro em 2016 permitiram que Çolak *et al.* (2015) obtivessem estimativas de demanda de viagens durante o evento. Neste caso, foram utilizados apenas dados de ligações. O problema com ligações é que o seu uso está cada vez menos frequente, já que estão sendo preferidas as trocas de mensagens via internet por aplicativos (Gillet *et al.*, 2017).

Embora o uso das tecnologias Bluetooth e Wi-Fi dependam da ativação do usuário, o uso cada vez mais frequente delas faz com que permaneçam sempre ativadas pela maior parte dos usuários. Abedi *et al.* (2013) coletaram dados tanto por Bluetooth como por Wi-Fi. Eles realizaram testes para determinar a velocidade de detecção das duas tecnologias e determinaram que com Wi-Fi é possível detectar um usuário oito segundos mais rápido do que com Bluetooth. A diferença de tempo pode parecer pequena, mas quando se trata de uma grande massa de dados, essa diferença é significativa. El-Tawab *et al.* (2016) utilizaram um sistema de coleta de dados por Wi-Fi para determinar o tempo de espera de passageiros em dois pontos de ônibus usando dispositivos móveis previamente conhecidos. Os resultados mostraram que foi possível obter uma estimativa de tempo de espera dos usuários em cada um dos pontos de ônibus e ainda obter uma confiança de que o dispositivo de coleta estava funcionando corretamente.

A escolha pelo método de coleta de dados deve se dar baseada na facilidade, velocidade e confiança de obter a informação que se deseja coletar. Preferencialmente deve-se minimizar a necessidade de interferência do usuário e optar por métodos que não afetem a experiência do usuário em termos de consumo de bateria. A diferença entre a escolha destas duas tecnologias pode se dar pela velocidade com que se coleta os dados, maior com Wi-Fi, o que para aplicações em tempo real é extremamente importante.

3. MÉTODO DE CONTAGEM AUTOMÁTICA DE PASSAGEIROS

A contagem automática de passageiros é feita com base nos dados de ônibus coletados por meio de AVL e com base nas detecções de requisições de conexão Wi-Fi realizadas no ponto de ônibus. Esses dados são processados apropriadamente para determinar os embarques e desembarques. A infraestrutura de coleta de dados de localização dos ônibus (AVL) e o equipamento detector de dispositivos com Wi-Fi construído para esta pesquisa são descritos a seguir.

3.1. Automated Vehicle Location

Foram usados dados obtidos pelo sistema AVL Virloc 10 da marca Virtec instalado em ônibus do Consórcio Fênix que opera na cidade de Florianópolis. O dispositivo embarcado no ônibus conta com GPS, modem GSM, acelerômetro e *Data Logger* para até 200 mil registros. O sistema coleta dados cada 30 segundos e transmite para uma central que os registra numa base de dados e exporta relatórios. Os dados coletados são os seguintes:

- Veículo: número de identificação do veículo na operadora de transporte;
- Linha: identificação da linha contendo código e nome;
- Sentido: sentido da linha;
- Hora do Evento: horário que os dados do ônibus foi capturado;
- PC: nome do ponto de ônibus onde os dados do ônibus foram obtidos;
- Latitude/Longitude: posição geográfica do ônibus quando os dados foram obtidos; e
- Hora da inserção: horário que o ônibus foi detectado pelo sistema AVL.

3.2. Dispositivo detector Wi-Fi

Os dispositivos Wi-Fi têm como característica a troca de pacotes mesmo sem estar conectados a um ponto de acesso (Turner, 1988). Por meio do NTM, é possível obter os dados de cada dispositivo Wi-Fi a cada troca de pacotes. Os dados obtidos pelo dispositivo detector de todo dispositivo Wi-Fi conectado ou buscando por um ponto de acesso para realizar a conexão em seu raio de captura são os seguintes:

- Endereço MAC: endereço físico único de cada dispositivo Wi-Fi;
- Primeira vez visto: primeira vez que um endereço MAC foi registrado;
- Última vez visto: última vez que um endereço MAC foi registrado;
- Potência do sinal: valor de potência reportado pelo dispositivo Wi-Fi;
- Pacotes: número de pacotes enviados para um usuário;
- BSSID: endereço MAC de um ponto de acesso ao qual o dispositivo está conectado; e
- *Probed ESSIDS*: endereço MAC do ponto de acesso ao qual um usuário está tentando se conectar.

Para a detecção dos dispositivos Wi-Fi foi desenvolvido um equipamento detector baseado em um computador Raspberry Pi. A arquitetura física do equipamento é ilustrada na Figura 1. A seguir são descritos os componentes do equipamento.

Raspberry Pi Model B - MicroSd 32GB Classe 10: o dispositivo Raspberry Pi é um computador do tamanho de um cartão de crédito com as mesmas configurações de um computador de médio porte o qual pode ser utilizado com sistema operacional Linux ou Windows. Para o trabalho foi escolhido o sistema Linux pela facilidade de modificação da programação do hardware.

Adaptador Wireless Intelbras WBN-300: o modelo de Raspberry utilizado não contém uma placa Wi-Fi integrada. Assim, foi necessário acoplar via USB um adaptador Wi-Fi. A escolha por esse modelo de adaptador se deve a sua capacidade de entrar em modo monitor, permitindo capturar as trocas de informações entre dispositivos Wi-Fi sem que haja conexão a uma rede. Esse modelo de adaptador também permite ajustar a potência do alcance do sinal e assim limitar o raio de captura.

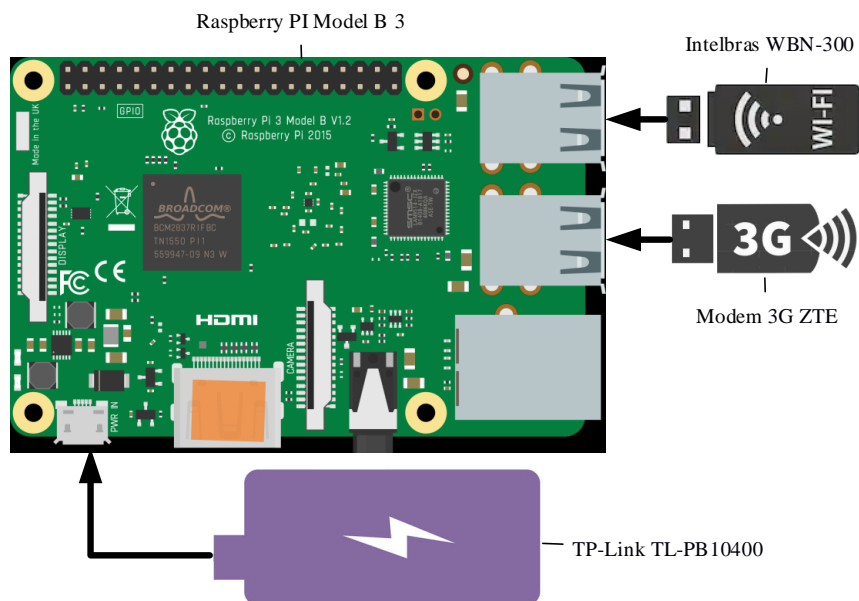


Figura 1: Arquitetura física do dispositivo de coleta de dados por detecção de Wi-Fi.

Modem 3G ZTE: para diminuir os riscos de perda de dados, seja por falhas ou furto do dispositivo, optou-se pelo envio imediato dos dados para um servidor. Para isso foi utilizado um modem 3G com chip de operadora de telefonia celular e um limite de 1GB por dia para envio dos dados. Estes são enviados para um servidor em nuvem.

Carregador portátil TP-Link TL-PB10400: a falta de rede elétrica ou de acesso a esta nos pontos de ônibus exige que o dispositivo seja auto suficiente em energia. Assim, o dispositivo conta um carregador portátil capaz de mantê-lo operando por treze horas sem o envio de dados para o servidor na nuvem e por seis horas com o envio dados para o servidor.

3.3. Procedimento de contagem automática de embarques e desembarques

O método de coleta e processamento consiste em detectar usuários que estejam com dispositivo Wi-Fi ligado e ativado dentro do raio de captura do dispositivo de coleta. Posteriormente os dados são processados em um servidor de aplicativos. O processamento relaciona o tempo de permanência na região de captura, número de detecções do usuário e potência do sinal com o horário aproximado de chegada/partida do ônibus para determinar se ocorreu um embarque ou desembarque e distinguir de detecções de transeuntes e pessoas que continuam no ponto. Embora o processamento possa ser feito em tempo real, neste trabalho o processamento foi todo realizado *offline*.

A Figura 2 apresenta a arquitetura geral do sistema de coleta. Os dados dos dispositivos Wi-Fi são obtidos pelo dispositivo de coleta apresentando na seção anterior por intermédio do NTM e, assim que disponíveis, são enviados pela rede 3G para o servidor em nuvem da IBM (Molloy, 2016). Ao final de um experimento, os dados armazenados no servidor e os dados obtidos do AVL por intermédio de relatórios são enviado para um servidor de aplicativos.

Em função da grande quantidade de dados coletados, torna-se inviável o processamento por meio

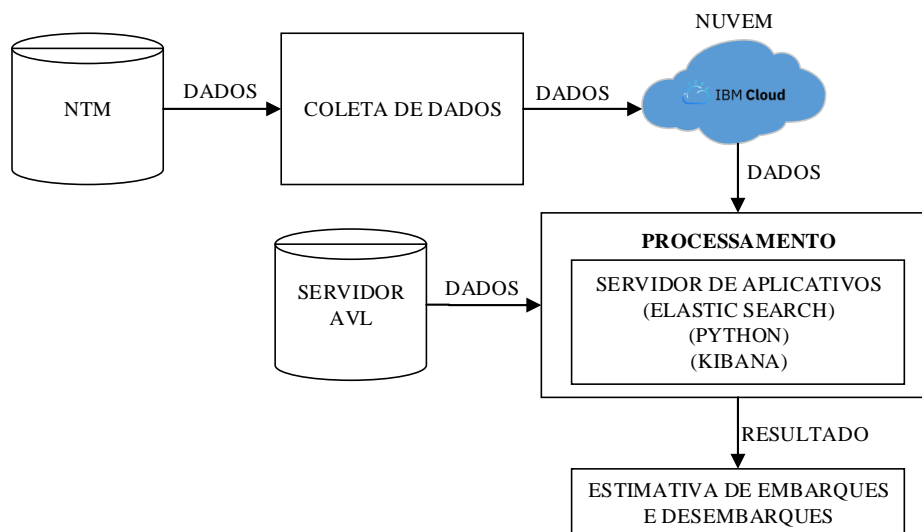


Figura 2: Método de coleta e processamento dos dados.

de planilhas eletrônicas. Por conta disso, foi usado um servidor de aplicativos ElasticSearch (Gormley e Tong, 2015; Bai, 2013) que permite a análise e processamento de grande quantidades de dados e um algoritmo de processamento em Python. Nessa etapa do processo é realizado o cruzamento dos dados de AVL com os dados de NTM. A aplicação Kibana (Bagnasco *et al.*, 2015) é utilizada para a visualização dos dados processados.

O processamento é dividido em quatro etapas: identificação e descarte de transeuntes, isto é, pessoas que apenas passaram pelo ponto; identificação de pessoas de pessoas paradas no ponto que ainda não embarcaram em ônibus; identificação de pessoas que embarcaram em um ônibus; e identificação de pessoas que desembarcaram de um ônibus.

A identificação e descarte de pessoas que estão somente passando pelo ponto é feita pelo número de vezes que um endereço MAC foi detectado e a potência do sinal dos dispositivos Wi-Fi registrada durante as detecções. Pessoas que apenas passam pelo ponto têm seus endereços MAC detectados poucas vezes, o tempo de detecção é curto, ou seja, o tempo entre a primeira e a última detecção é pequeno, e a potência do sinal é baixa nas detecções iniciais e finais.

Uma pessoa parada no ponto tem seu endereço MAC detectado várias vezes, o tempo de detecção é mais longo, ou seja, entre a sua primeira detecção e a última há um tempo que pode caracterizar uma espera por ônibus, e a potência do sinal é alta nas vezes que é detectado, indicando que está próximo ao detector. Se o fim da detecção de um endereço MAC coincide com a partida de um ônibus é possível que essa pessoa tenha embarcado no ônibus e um embarque é registrado. Podem ocorrer casos em que usuários estão correndo para chegar ao ônibus e são detectadas poucas vezes sendo considerado um usuário passando. Nesse caso seria preciso ter outro dispositivo de coleta em outro ponto e identificar se o mesmo endereço MAC aparece em outro dispositivo.

No caso de desembarque, um novo endereço MAC é detectado no momento em que o ônibus chega, e após alguns segundos esse mesmo endereço desaparece, mesmo com o ônibus ainda

parado no ponto. Como usuários embarcados no ônibus também são detectados, eles não são considerados desembarques quando a detecção termina na mesma hora que o ônibus sai. Podem ocorrer casos em que usuários que acabaram de desembarcar ainda estão sendo detectados no momento que o ônibus saiu. Nesse caso é preciso ter outro detector em outro ponto para comparar se um mesmo endereço MAC foi detectado em dois pontos. Dados que não se enquadram nestas quatro categorias são desconsiderados.

O uso de dispositivos de detecção em diversos pontos de uma mesma linha permite inclusive o cruzamento de informações entre pontos. Esta abordagem permite reduzir erros, por exemplo, entre desembarques e pessoas que permanecem no ônibus. Porém, esta aplicação não foi realizada neste trabalho, permanecendo como perspectiva de aplicação em novos experimentos.

4. EXPERIMENTOS REALIZADOS

Dois tipos de experimentos foram realizados. O primeiro experimento visou determinar a taxa de detecção de dispositivos e foi realizado em cenário controlado no qual conhecem-se todos os dispositivos na vizinhança de detecção. Contou-se com a colaboração de voluntários que portavam os smartphones em diferentes condições operacionais para emulação de condições típicas do ambiente de produção. O segundo experimento foi realizado em cenário típico de produção para determinação do número de detecções em relação aos volumes de sobe-e-desce medidos diretamente por pesquisadores. Um ponto de parada de ônibus com movimento adequado (cerca de dois mil embarques e desembarques por dia útil) foi escolhido.

Para ambos experimentos o detector de Wi-Fi descrito na Seção 3.2 foi configurado com potência de 1 mW, o que permite um raio de captura de aproximadamente 12 m. Esse raio é o mínimo que pode ser obtido pela placa Wi-Fi acoplada no detector, a qual não opera abaixo de 1 mW. Como o interesse é colocar em um ponto de ônibus e detectar somente dispositivos próximos ao ponto, o raio de 12 m é suficiente.

4.1. Experimento 1 - Cenário controlado

Este experimento tem como objetivo principal validar a possibilidade de uso de Wi-Fi para a detecção de dispositivos e seu uso na contagem automática de embarques e desembarques. Mais especificamente, deseja-se saber se as detecções ocorrem, quanto tempo é necessário para uma detecção e com que frequência ocorre, e se o raio de captura escolhido se verifica. Os parâmetros encontrados neste experimento serviram de padrão para análise dos dados coletados no cenário real.

4.1.1. Projeto do experimento

Inicialmente, optou-se por um cenário controlado em que fosse possível conhecer previamente as informações de todos os dispositivos Wi-Fi envolvidos. Para este fim, o campo de futebol da Universidade Federal de Santa Catarina foi escolhido e alguns dados dos dispositivos móveis dos participantes, como o endereço MAC, foram obtidos antecipadamente.

Ao todo doze pessoas participaram do experimento, cada um portando um dispositivo móvel com Wi-Fi. A Figura 3 mostra a configuração do experimento no campo. Um dispositivo detector de Wi-Fi foi posicionado entre dois marcadores afastados um do outro por 24 m, perfazendo o diâmetro da área de captura do dispositivo. O papel de cada participante era caminhar entre

os dois marcadores conforme instruções que serão apresentadas a seguir e registrar em seu dispositivo, por meio de um aplicativo, os instantes de entrada e saída da área de captura teórica do dispositivo de detecção. Um segundo dispositivo detector foi posicionado a 12 m de um dos marcadores com o objetivo de detectar os dispositivos que saíam do raio do detecção do dispositivo central por aquela extremidade.

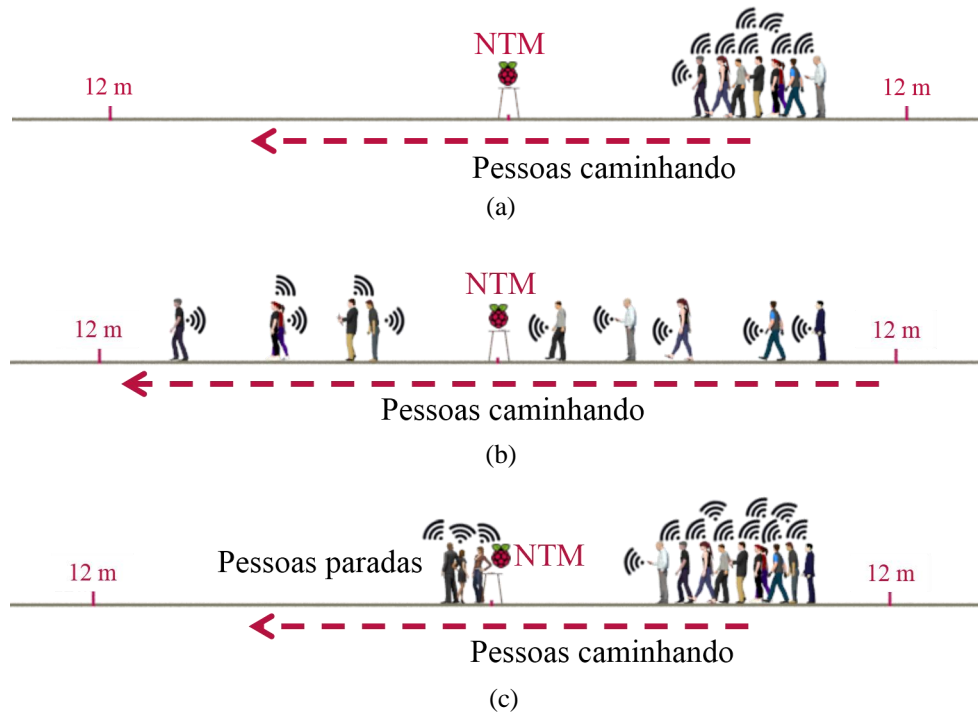


Figura 3: Configuração e etapas do Experimento 1: (a) Etapa 1- pessoas passaram em grupo em frente ao detector; (b) Etapa 2 - pessoas passaram em intervalos diferentes em frente ao detector; (c) Etapa 3 - pessoas paradas enquanto outras passaram em frente ao detector.

O experimento foi dividido em três etapas. A primeira etapa está representada na Figura 3(a). Consistiu no deslocamento simultâneo de todos os participantes de uma extremidade a outra da área do experimento passando em frente ao detector de Wi-Fi. O objetivo dessa etapa foi verificar o comportamento do detector quando da travessia de um grupo de pessoas simultaneamente pela área de captura.

A segunda etapa está representada na Figura 3(b). Consistiu no deslocamento dos participantes em momentos diferentes de uma extremidade à outra da área do experimento passando em frente ao detector Wi-Fi. Esta etapa foi realizado verificar a capacidade de detectar dispositivos em momentos diferentes.

A terceira etapa está representada na Figura 3(c). Nesta etapa tentou-se reproduzir o comportamento em um ponto de ônibus. Alguns participantes foram orientados a parar em frente ao detector enquanto outros passaram, simulando um desembarque. Ao mesmo tempo, outros participantes que estavam parados em frente ao detector foram orientados a continuar com o grupo, simulando um embarque.

4.1.2. Resultados

A primeira análise consiste em verificar o intervalo de detecções sucessivas de cada endereço MAC. Esse parâmetro é importante para determinar se uma pessoa possa estar parada em um ponto. Os dados da Tabela 1 indicam que há 94,66% de certeza que em até 120 s um usuário será detectado novamente.

Tabela 1: Intervalo entre detecções sucessivas.

RASP	Intervalo entre detecções (s)									Usuários
	40	50	60	70	80	90	100	110	120	
1	80,5	86,6	88,4	90,3	91,2	92,3	93,3	93,8	94,7	(%)
2	84,1	84,1	86,7	88,9	90,4	92,1	92,9	93,7	94,4	(%)

A segunda análise consiste da determinação do menor e maior intervalo e intervalo médio entre detecções. Isto é, se um usuário foi detectado uma vez, qual o menor e o maior intervalo em que este permaneceu sem ser detectado. A Tabela 2 indica que o pior caso foi um usuário que ficou 21 min sem ser detectado. O menor intervalo ficou em 1 s. Já a média entre detecções ficou em 70 s para um dispositivo de coleta e 67 s para outro dispositivo.

A terceira análise busca encontrar o tempo médio de travessia dos participantes. Os tempos de travessia foram obtidos por meio dos registros no aplicativo no qual os participantes marcavam o horário em que entraram no raio de coleta e o momento em que saíram do raio de coleta. Esse tempo ficou em 18 s, consistente com o raio de 12 m da região de detecção. A velocidade equivalente de caminhada resulta em 1,33 m/s ou 4,8 km/h, típicas de caminhadas em regiões planas sem obstáculos.

A quarta análise é retirada do dispositivo de coleta, e se baseia na potência do sinal quando os usuários estão passando em frente ao detector ou quando há usuários parados em frente ao detector. Quando os usuários passam em frente ao detector sem parar, a potência média do sinal fica acima de -78 dB, e quando os usuários estão em frente ao detector e param a potência fica abaixo de -78 dB. Isso significa que quanto mais próximo do detector maior a potência do sinal e quanto mais afastado, menor esta se torna.

A quinta análise relaciona o intervalo entre detecções e a potência do sinal, o que pode indicar um usuário parado. Foi obtido que em um intervalo entre detecções de mais de 50 s e a potência do sinal acima de -78 dB, o usuário estava parado.

4.2. Experimento 2 - Cenário real

Neste experimento, o método de coleta de dados é aplicado em um ponto de ônibus de movimento adequado em dia típico da semana. Neste caso, não se conhecem as características dos usuários e os seus comportamentos. Os parâmetros obtidos no experimento controlado são considerados válidos para este cenário e são utilizados para identificações de embarque e

Tabela 2: Menor, maior e intervalo médio entre detecções.

RASP	Intervalo entre detecções (s)		
	Menor	Maior	Média
1	1	1263	70
2	3	1118	67

desembarque. Somam-se aos dados do experimento controlado informações de contagem de sobre-desce realizada por pesquisadores no local com a finalidade de validar os resultados obtidos de forma automática.

4.2.1. Projeto do experimento

Para implantação do modelo de coleta de dados, o local escolhido é a Universidade Federal de Santa Catarina, na cidade de Florianópolis no estado de Santa Catarina. Inicialmente para aplicação do modelo foi selecionado um ponto de ônibus com demanda de passageiros considerada adequada para as finalidades da aplicação, da ordem de dois mil passageiros por dia útil. No ponto, foi instalado o equipamento detector para coletar informações dos passageiros que estão no ônibus ou nos pontos de parada.

O experimento foi realizado no dia 13/06/2018 no período das 14:00 às 19:00, sendo coletadas informações dos usuários que estavam com dispositivos Wi-Fi ligados. Dados dos ônibus foram obtidos por meio do sistema de AVL de uma linha específica, denotada "185 – UFSC SEMIDIRETO", que vem a ser a linha com maior demanda a operar no ponto escolhido. Durante esse intervalo foi realizado uma contagem manual a fim de aferir o modelo com os dados da realidade. Para isso dois pesquisadores se deslocaram até o ponto e realizaram a contagem manual de embarque e desembarque.

Cruzando os dados do AVL com os dados de detecção, espera-se observar aumento significativo na quantidade de detecções de usuários no detector no momento em que o ônibus passar ou parar no ponto. A Figura 4 ilustra o exemplo de um ônibus chegando ao ponto de ônibus, e ao mesmo tempo pessoas esperando no ponto de ônibus com o Wi-Fi dos seus dispositivos ligados, enquanto o dispositivo de coleta realiza a captura dos dados.

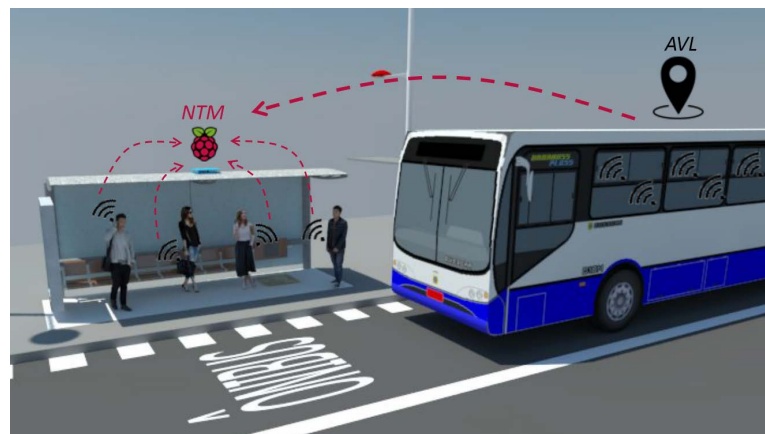


Figura 4: Ponto de ônibus com dispositivo de coleta captando informações de usuários que estão esperando para embarcar e de usuários dentro de ônibus.

4.2.2. Resultados

Com a coleta de dados realizado primeiro foi realizado uma análise exploratória dos dados para identificar possíveis campos com erros. Isto porque como a quantidade de informações enviadas ao mesmo tempo é alta, é possível que se tenha erros no registro dos dados captados. A Tabela 3 ilustra o resultado da análise do ponto estudado.

Tabela 3: Relação entre o número total de dados e a perda de informações.

Ponto	Total de dados	Dados válidos	Erro (% do total)
A	54.298	54.054	0,44

Após a análise exploratória, foi realizado uma análise para identificar a quantidade de endereços MAC's únicos. A Figura 5(a) ilustra o primeiro momento em que o usuário foi detectado e a Figura 5(b) o último momento em que foi detectado. Ao todo foram identificados 13.046 endereços MAC. Em seguida foi realizado o processamento para encontrar a média do intervalo entre detecções de um mesmo dispositivo; a média ficou em 80 s, 10 s a mais do que no caso do experimento controlado.

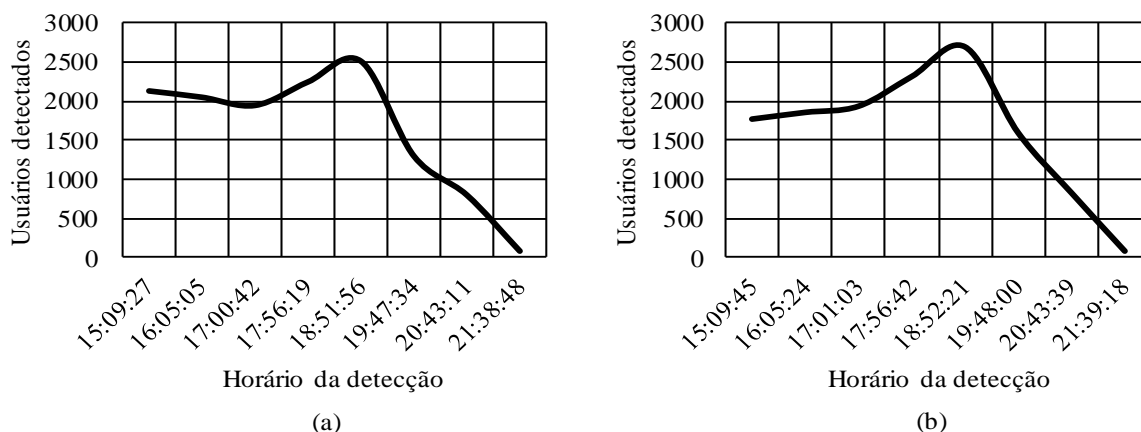


Figura 5: (a) Usuários detectados pela primeira vez e (b) Usuários detectados pela última vez.

Para os dados de AVL, no período do experimento foram obtidos 55 horários da linha 185 – UFSC SEMIDIRETO.

Como citado anteriormente durante os testes dois pesquisadores foram até o ponto realizar uma contagem manual com a finalidade de aferir o modelo. A pesquisa foi realizada durante 40 min, sendo contados quatro ônibus da linha pesquisada. O cruzamento dos dados de coleta com os dados pesquisado no ponto estão na Tabela 4.

Tabela 4: Cruzamento entre dados de pesquisa de contagem no ponto com dados obtidos através do modelo de coleta.

Linha	Chegada	Partida	Pesquisa		Modelo	
			Embarque	Desembarque	Embarque	Desembarque
UFSC3	17:19:04	17:21:12	23	22	29	18
UFSC4	17:23:56	17:24:52	6	26	12	28
UFSC8	17:31:47	17:32:48	20	15	23	14
UFSC11	17:39:55	17:41:58	28	25	25	23

É possível observar que os dados do detector estão próximos dos dados da pesquisa de campo. Sabe-se que o modelo pode apresentar falsos positivos, ou seja usuários podem estar contados como embarcados ou desembarcados mas na verdade não o fizeram. Como também podem ocorrer falsos negativos, como usuários podem ter embarcado e desembarcado e não foram contados. Sabe-se que nem todos os usuários mantém seus dispositivos Wi-Fi ligados, o que pode contribuir para esses falsos negativos. Os falsos positivos podem ocorrer por pessoas

estarem passando pelo ponto no momento em que o ônibus sai do ponto, e elas podem acabar sendo contadas.

5. CONCLUSÃO E TRABALHOS FUTUROS

Para uma estimativa dos dados de contagem os resultados são satisfatórios. No futuro pretende-se incluir outros três dispositivos de coleta, um no terminal de integração da cidade de Florianópolis e os demais nos pontos seguintes ao pesquisado seguindo o traçado da linha. Pretende-se realizar o cruzamento entre os dados dos pontos para saber se um usuário está embarcado com base na detecção em vários pontos. Para os dados de desembarque ainda será possível fazer o cruzamento com os dados de Wi-Fi da universidade.

Agradecimentos

Os autores agradecem o apoio da CAPES; do Consórcio Fênix, pela disponibilidade dos dados AVL; e dos voluntários, que ajudaram nos experimentos.

REFERÊNCIAS

- Abedi, N.; A. Bhaskar e E. Chung (2013) Bluetooth and Wi-Fi MAC address based crowd data collection and monitoring: benefits, challenges and enhancement. *36th Australasian Transport Research Forum*. Queensland University of Technology, Brisbane.
- Alexander, L.; S. Jiang; M. Murga e M. C. González (2015) Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, v. 58, p. 240–250.
- Arhin, S.; E. Noel; M. F. Anderson; L. Williams; A. Ribisso e R. Stinson (2016) Optimization of transit total bus stop time models. *Journal of Traffic and Transportation Engineering (English Edition)*, v. 3, n. 2, p. 146–153.
- Bagnasco, S.; D. Berzano; A. Guarise; S. Lusso; M. Masera e S. Vallero (2015) Monitoring of IaaS and scientific applications on the cloud using the Elasticsearch ecosystem. *Journal of Physics: Conference Series*, v. 608, n. 1.
- Bai, J. (2013) Feasibility analysis of big log data real time search based on hbase and elasticsearch. *2013 Ninth International Conference on Natural Computation*, p. 1166–1170.
- Barbosa, A. F. (2016) Pesquisa sobre o uso das tecnologias de informação e comunicação no Brasil: TIC Domicílios e Empresas 2016. *São Paulo: Comitê Gestor da Internet no Brasil*.
- Bekhor, S.; Y. Cohen e C. Solomon (2013) Evaluating long-distance travel patterns in Israel by tracking cellular phone positions. *Journal of Advanced Transportation*, v. 47, n. 4, p. 435–446.
- Cathey, F. e D. J. Dailey (2003) A prescription for transit arrival/departure prediction using automatic vehicle location data. *Transportation Research Part C: Emerging Technologies*, v. 11, n. 3-4, p. 241–264.
- Çolak, S.; L. P. Alexander; B. G. Alvim; S. R. Mehndiratta e M. C. González (2015) Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. *Transportation Research Record: Journal of the Transportation Research Board*, v. 2526, p. 126–135.
- Gillet, J.; M. Meyer e B. A. Lucini (2017) GMEI 2017: Global Mobile Engagement Index.
- Gormley, C. e Z. Tong (2015) *Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine*. O'Reilly Media, Inc.
- Jiang, S.; J. Ferreira e M. C. Gonzalez (2017) Activity-based human mobility patterns inferred from mobile phone data: a case study of Singapore. *IEEE Transactions on Big Data*, v. 3, n. 2, p. 208–219.
- Lind, A.; A. Hadachi e O. Batrashev (2017) A new approach for mobile positioning using the CDR data of cellular networks. *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems*, p. 315–320.
- Molloy, D. (2016) *Exploring Raspberry Pi: interfacing to the real world with embedded Linux*. John Wiley & Sons.
- O'Flaherty, C. A. (1997) *Transport planning and traffic engineering*. Wiley, London.
- Sicker, D. C.; P. Ohm e D. Grunwald (2007) Legal issues surrounding monitoring during network research. *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, p. 141–148.
- El-Tawab, S.; R. Oram; M. Garcia; C. Johns e B. B. Park (2016) Poster: Monitoring transit systems using low cost WiFi technology. *2016 IEEE Vehicular Networking Conference*.

- Turner, J. S. (1988) Design of a broadcast packet switching network. *IEEE Transactions on Communications*, v. 36, n. 6, p. 734–743.
- Vuchic, V. R. (2005) *Urban transit: operations, planning and economics*. J. Wiley & Sons, Hoboken, N.J.
- Zhang, Y.; X. Qin; S. Dong e B. Ran (2010) Daily O-D Matrix Estimation Using Cellular Probe Data. *Transportation Research Board 89th Annual Meeting*. Washington DC, United States.

Diego Benites Paradedá (diegoparadedá@gmail.com)
Werner Kraus Junior (werner.kraus@ufsc.br)
Rodrigo Castelan Carlson (rodrigo.carlson@ufsc.br)

Programa de Pós Graduação em Engenharia de Automação e Sistemas, Universidade Federal de Santa Catarina

Florianópolis, SC, 88040-900, Brazil