

ESTIMATIVA DA VIABILIDADE DA RECONSTRUÇÃO DA MATRIZ ORIGEM DESTINO ATRAVÉS DE SINAIS WI-FI DE PASSAGEIROS

Paulo de Tarso Gonçalves Sousa Júnior

Marcos Paulino Roriz Junior

Kamilla Cruvinel Rodrigues

Ronny Aliagra Medrano

Universidade Federal de Goiás

Faculdade de Ciências e Tecnologia – Engenharia de Transportes

RESUMO

A obtenção da matriz de origem-destino (O/D) dos passageiros no transporte público permite compreender e propor melhorias no sistema. A estimativa destes deslocamentos é tipicamente realizada utilizando pesquisa de campo, que por motivos de custos não são realizadas frequentemente. Nesse sentido, têm-se investigado mecanismos alternativos. Dentre esses, destaca-se a utilização da tecnologia Wi-Fi, que utiliza os pacotes enviados pelos *smartphones* para rastrear os passageiros. Entretanto, a taxa de emissão desses pacotes não é padronizada, variando entre diferentes fabricantes. Além disso, a transmissão dos pacotes pode ocorrer em diferentes canais de frequência, levando a possíveis perdas de pacotes. Nesse sentido, este trabalho investiga a viabilidade de estimar a matriz O/D frente as limitações da tecnologia Wi-Fi. Para isso, é proposto um algoritmo para reconstrução desta matriz. Os resultados preliminares sugerem que os pares de origem-destino obtidos divergiram tipicamente em até 4 pontos de embarque/desembarque frente aos pontos reais utilizados pelos passageiros.

ABSTRACT

By obtaining the origin-destination matrix (O/D) of public transportation passengers' is possible to understand and propose improvements to the system. The estimation of such movements is typically performed using field research, which for cost reasons are often not performed regularly. Hence, alternative techniques have been proposed. Among such, we highlight the use of the Wi-Fi technology, which uses packets sent by smartphones to track passengers. However, the emission rate of such packets are not standardized, varying among different manufacturers. In addition, packets can be transmitted to different channels, which can lead to loss of packets. In this sense, this work investigates the feasibility of estimating the O/D matrix considering the intrinsic limitations of the Wi-Fi technology. To do so, it presents an algorithm for reconstructing such matrix. Preliminary results suggests that the O/D pairs obtained diverge typically up to 4 stations considering the real ones used by passengers.

1. INTRODUÇÃO

O transporte público urbano de ônibus tem sofrido uma queda expressiva na demanda de passageiros nos últimos anos (Soares, 2016). Para compreender essa transformação tem sido realizado pesquisas acerca do deslocamento dos passageiros, também conhecido como pesquisa de origem-destino (O/D). Entender o deslocamento das pessoas, a matriz origem-destino (O/D), do sistema de transporte possibilita compreender os locais e fatores envolvidos nesta queda e, conseqüentemente, a capacidade de propor melhorias e otimizações no sistema (Arbex e da Cunha, 2017; Chen *et al.*, 2016).

A estimativa dessas matrizes pode ser realizada utilizando métodos indiretos, através de modelos de previsão de demanda, ou métodos diretos, que se baseiam na realização de pesquisa de campo, sendo a última a mais utilizada. Embora seja uma prática relativamente comum, verifica-se que os empecilhos de realização da pesquisa de campo são cada vez maiores devido às dificuldades de acesso aos entrevistados. Além disso, os custos e o tempo de execução deste tipo de pesquisa são muito expressivos, restringindo a frequência de sua aplicação para prazos mais alongados (Guerra *et al.*, 2014). Nesse sentido, faz-se necessário investigar novas formas de obtenção de dados que permitam entender os deslocamentos dos passageiros.

Como alternativa, diversas pesquisas têm explorado dados gerados por *smartphones* de passageiros como meio de obtenção do deslocamento dos mesmos (Mikkelsen *et al.*, 2016), A

obtenção desses dados têm ocorrido através da captura de pacotes de rede da especificação (protocolo) IEEE 802.11, também conhecido como Wi-Fi (Crow *et al.*, 1997). O padrão Wi-Fi inclui um pacote de sinal de varredura (*probe request*), semelhante a uma sonda, que os dispositivos portáteis emitem periodicamente para descobrir pontos de acesso próximos, ou seja, roteadores que oferecem uma rede Wi-Fi.

Estudos preliminares indicam que estes dados possibilitam detectar a presença e contabilizar os passageiros dentro dos ônibus (Mikkelsen *et al.*, 2016; Myrvoll *et al.*, 2017; Oransirikul *et al.*, 2014; Paradedda Benites *et al.*, 2018; Roriz Junior e Medrano, 2018). Entretanto, um problema recorrente nesses trabalhos é a demora em capturar os dados de Wi-Fi (Oransirikul *et al.*, 2016; Roriz Junior e Medrano, 2018). Isso decorre da especificação IEEE 802.11 (Wi-Fi) não definir a taxa ou período padrão de transmissão dos sinais (pacotes) de sondagem (*probe requests*). Desta maneira, a fabricante do *smartphone* é livre para definir tais valores.

A irregularidade no envio destes pacotes pode dificultar a construção da matriz O/D, visto que o usuário pode ter embarcado no veículo em um ponto e só ter seu pacote detectado em paradas subsequentes. A mesma falha de sincronização pode ocorrer no desembarque. Neste caso, o último pacote do usuário pode não ser a parada de fato que o mesmo utilizou.

Um outro problema ainda não abordado pelos trabalhos apontados decorre de outra limitação da especificação IEEE 802.11: o canal de frequência em que a comunicação ocorre, isto é, onde o pacote é enviado. Precisamente, os trabalhos apontados utilizam um computador, denominado *sniffer*, para monitorar e processar os pacotes enviados em um determinado canal. A fim de monitorar o espectro público, o *sniffer* alterna entre os diferentes canais para ouvir os pacotes da respectiva banda. Porém, pode ocorrer de o *smartphone* do passageiro enviar o pacote de sonda em um canal diferente daquele em que o *sniffer* está ouvindo naquele momento. Quando isso ocorre, o pacote será perdido. O estudo de Freudiger (2015) mostra que em alguns cenários um único *sniffer* pode capturar apenas 57% dos pacotes de sondagem enviados.

Nesse sentido, com base nesses desafios, este artigo visa pesquisar a capacidade de se utilizar os dados Wi-Fi como meio de reconstruir a matriz O/D dos passageiros de transporte público. Precisamente, será investigado a viabilidade de se reconstruir a matriz O/D frente às limitações da periodicidade de transmissão de pacotes de sondagem e do canal (frequência) utilizado, que são inerentes ao protocolo IEEE 802.11, respondendo assim o quão similar tal matriz pode chegar da de fato realizada pelos passageiros. Para responder essas questões, é proposto um algoritmo para reconstrução da matriz O/D a partir dos dados de sondagem Wi-Fi capturados.

Como uma parte crucial do algoritmo envolve a classificação de um determinado pacote de sondagem como dentro ou fora do veículo, utiliza-se métodos existentes para tal tarefa, isto é, esta informação é dada como entrada para o algoritmo proposto. Especificamente, será feito o uso do método UAI-FI (Utilização de Aprendizado de máquina para contagem automática de passageiros utilizando wi-Fi) (Roriz Junior e Medrano, 2018). Este método utiliza *machine learning* para classificar os pacotes oriundo dos *smartphones* dos passageiros como dentro ou fora. Entretanto, ressalta-se que o algoritmo proposto é independente de tal método, sendo o filtro de classificação uma informação de entrada do método. Destaca-se que o algoritmo proposto neste trabalho foi implementado na linguagem Python e implantado na plataforma de *hardware* Raspberry Pi3B (Richardson e Wallace, 2014). Para avaliá-lo foi realizado testes de campo preliminares controlados em uma linha circular da cidade de Aparecida de Goiânia.

A fim de detalhar o algoritmo proposto, este trabalho divide-se da seguinte forma. A Seção 2 apresenta a fundamentação teórica do trabalho, abordando os conceitos de matriz O/D e características do protocolo Wi-Fi, além dos trabalhos relacionados. A Seção 3 descreve as etapas do algoritmo para reconstrução dos pares O/D. Os resultados preliminares da pesquisa são apresentados e discutidos na Seção 4. Por fim, a Seção 5 apresenta as conclusões e aponta questões para trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

Esta pesquisa, juntamente com o algoritmo proposto neste trabalho, se baseia nos conceitos relativos a matriz O/D e na utilização na utilização dos sinais (*probe requests*) emitidos pelos *smartphones* dos passageiros, além das próprias limitações da especificação IEEE 802.11 (Wi-Fi). Com a finalidade de embasar o fundamento do algoritmo, esta seção discorre sucintamente sobre o funcionamento dos mesmos e dos principais trabalhos relacionados a temática.

2.1. Matriz Origem-Destino (O/D)

O levantamento de viagens origem-destino é um instrumento fundamental no planejamento de sistemas de transporte. Os estudos origem-destino são usados para obtenção de informação de viagens existentes e para estimar o movimento de pessoas em uma região. Os resultados desses levantamentos são expressos em forma de matriz, daí o nome matriz O/D, as quais contém em suas células os números estimados das viagens realizadas de uma zona de origem (linha) para uma zona de destino (coluna), contemplando assim a distribuição das viagens geradas em uma zona para as demais (Ferraz e Torres, 2004).

Para a construção da matriz O/D é necessário obter o trajeto de deslocamento do passageiro do transporte público, desde a entrada do mesmo no sistema, até sua saída no destino final. Existe a dificuldade da obtenção desse deslocamento pelo sistema de bilhetagem eletrônica, tanto pelo fato de comumente só ter a validação do bilhete na entrada do veículo, quanto pela existência de terminais de transbordos, em que os passageiros tipicamente entram pela parte de trás do veículo e não precisam revalidar o cartão de viagem (Pelletier *et al.*, 2011). Com base nessas limitações, decidiu-se explorar outras tecnologias que possam contribuir ou complementar as viagens da matriz estimada. Dentre essas tecnologias, pode-se destacar o Wi-Fi.

2.2. Protocolo IEEE 802.11 (Wi-Fi)

O protocolo IEEE 802.11, comumente conhecido como Wi-Fi, é uma tecnologia de rede sem fio concebida e padronizada pelo Instituto dos Engenheiros Eletricistas e Eletrônico (IEEE) para possibilitar a comunicação sem fio entre dispositivos eletrônicos (Crow *et al.*, 1997). Esse protocolo especifica processos para realizar cada uma das atividades envolvidas na conexão dos dispositivos, dentre essas, pode-se destacar o processo de descoberta de rede (Gast, 2005). Para realizar tal atividade, o protocolo especifica um processo que sonda dispositivos próximos através de um sinal (pacote de rede) denominado *probe request*. Periodicamente, *Smartphones* emitem estes pacotes com intuito de encontrar roteadores Wi-Fi próximos (Freudiger, 2015).

Os sinais dos pacotes de sondagem são difundidos publicamente em um canal de onda próximo ao emissor, permitindo que todos dispositivos interessados possam recebê-lo. Com intuito de maximizar a transmissão de dados, a especificação IEEE 802.11 define 14 canais de frequência para envio de dados, como ilustrado na Figura 1. Tipicamente é utilizado os canais 1, 6 ou 11 devido os mesmos não se sobreporem, o que aumenta a vazão e diminui a taxa de colisão e perda de pacotes. Toda comunicação entre os dispositivos deve ocorrer em um único canal.

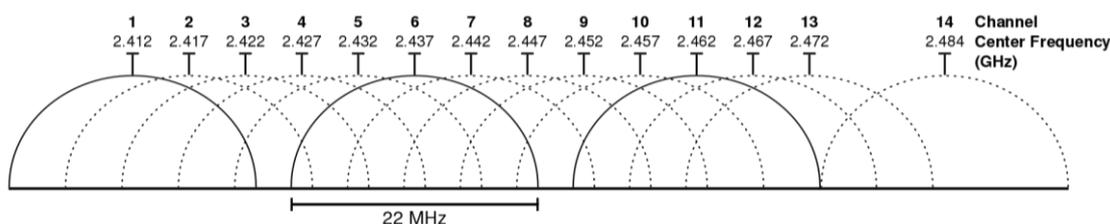


Figura 1: Canais da especificação Wi-Fi (2.4 GHz). Extraído de LifeHacker (2019).

Como os pacotes de sondagem são enviados periodicamente e em faixas de onda públicas, pode-se utilizar um computador para monitor e armazenar tais dados. Estes computadores, chamado de *sniffers*, ativam a placa Wi-Fi em modo monitor e ouvem os pacotes transmitidos. Por exemplo, ao situar o *sniffer* no canal 1 o mesmo poderá capturar pacotes do primeiro canal e possivelmente alguns do segundo, terceiro e quarto. Comumente, deseja-se que o *sniffer* ouça os pacotes transmitidos em todos os canais. Porém, como só é possível ouvir um canal por vez, o *sniffer* deve alternar entre eles. Tipicamente é feita a alternância entre os canais 1, 6 e 11 para cobrir o espectro. Esta atividade é feita milhares de vezes por segundo (Hertzog e O’Gorman, 2017). Entretanto, pode ocorrer de o pacote de sonda do usuário ser transmitido em um canal em que o *sniffer* não esteja ouvindo. Neste caso, o pacote não será detectado. Uma estratégia para resolver isso é utilizar 3 *sniffers* fixos, nos canais 1, 6 e 11, visto que os mesmos não se sobrepõem e podem ouvir a comunicação dos demais canais Freudiger (2015).

Cada pacote emitido inclui um número que identifica globalmente aquele aparelho, uma espécie de impressão digital do smartphone, chamado de MAC (*Media Access Control*), esse MAC é único e com ele é capaz de diferenciar cada aparelho. O alcance dos pacotes depende do perfil do padrão Wi-Fi sendo utilizado. Como a maioria dos chips são do perfil *b*, *g*, ou *n* do padrão, isto é, IEEE 802.11.b/g/n, sabe-se que a distância alcançada pelos pacotes é de até 100 metros em campo aberto e de até 30 metros dentro de residências (Gast, 2005).

Além do identificador MAC, os receptores recebem um indicador da potência do sinal da onda recebida, denominado *Received Signal Strength Indication* (RSSI). Ao obter um pacote, o receptor computa o indicador RSSI, em decibéis, que representa a proximidade dos aparelhos. Este valor situa entre 0 e $-\infty$, sendo que quanto mais próximo de zero, mais próximo os dispositivos se encontram. Note que devido a escala deste indicador estar em decibéis, pequenas variações causam grandes variações na distância. Por exemplo, um valor de RSSI de -65 dBm situa o passageiro a uma distância de aproximadamente 50 metros do *sniffer*, enquanto que -65 dBm o posiciona a 90 metros de distância (Mohammed, 2016).

Vários trabalhos utilizam esse indicador como meio de filtrar os passageiros que estão dentro do veículo. A escolha do valor limite a ser utilizado para classificar os pacotes e, conseqüentemente, os passageiros como dentro ou fora do veículo é tipicamente realizada manualmente (Roriz Junior e Medrano, 2018). Uma alternativa para lidar com a sensibilidade deste indicador é utilizar uma abordagem de *aprendizado de máquina* para escolher automaticamente um valor limite que melhor separe os pacotes.

Existem limitações consideráveis ao se trabalhar com esses dados, o principal é que o protocolo não define um intervalo de frequência para o envio de pacotes de sondagem (Gast, 2005), *i.e.*, não se sabe o valor de retransmissão desses pacotes, sendo estes dependente da marca e da fabricante. Assim, a fabricante do smartphone é livre para definir tais valores.

Além dos problemas apontados, Freudiger (2015) relata que a quantidade de pacotes emitidos está relacionada ao modo de uso do *smartphone*, *e.g.*, um aparelho com a tela travada transmite menos pacotes do que um destravado. Dentre os modos de uso, relata-se que o pior caso, isto é, aquele que apresenta o menor número de transmissões é quando o aparelho está com a tela desligada. Note que esta configuração tipicamente será a observada no transporte público, uma vez que os aparelhos se encontram guardados no bolso (tela desligada) por razões de segurança.

Um outro problema surge com os aparelhos da marca Apple. Precisamente, a partir da versão 8 do sistema operacional iOS os aparelhos enviam identificadores MAC aleatórios no processo de sondagem (Arnott, 2014). Entretanto, estudos recentes mostram que é possível descobrir e rastrear esses aparelhos através de correlações de pacotes (Martin *et al.*, 2017). Descobriu-se que os dados internos dos pacotes de sondagem não são aleatórios, como o número de sequência e total de pacotes de enviados. Através desses números, que aparecem em diferentes pacotes, pode-se associar pacotes com MAC diferentes ao mesmo aparelho. Porém, com intuito de simplificar a validação da abordagem proposta, a mesma utilizará somente aparelhos Android.

2.3. Trabalhos Relacionados

O potencial de obtenção de dados da matriz O/D através da tecnologia Wi-Fi tem atraído a atenção da comunidade científica. Dentre essas, pode-se destacar o trabalho de Ji *et al.* (2017) que realiza uma análise dos dados históricos para reconstrução da matriz. O trabalho considera que o tempo de transmissão dos pacotes é uma constante λ conhecida. Entretanto, é difícil estimar tal parâmetro devido a irregularidade dos aparelhos e de taxa de transmissão conforme o uso do mesmo (*e.g.*, tela bloqueada, navegando). Além disso, os autores não lidam com os problemas referentes a transmissão de pacotes em diferentes canais.

Outro trabalho em destaque é o de Danalet *et al.* (2014), que aborda a reconstrução da matriz O/D utilizando dados Wi-Fi de pedestres. O objetivo dos autores era detectar as atividades de pedestres em um campus. Entretanto, o método proposto não considera a periodicidade do envio de pacotes, nem a comunicação em diferentes canais.

3. METODOLOGIA

Para abordar essa problemática, apresenta-se um algoritmo voltado a reconstrução da matriz O/D dos passageiros que considera as limitações inerentes ao Wi-Fi discutidas. Primeiramente, o método capturar os pacotes para, em seguida, reconstruir a origem e destino dos mesmos. Esta seção descreve os passos envolvidos em cada etapa.

3.1. Plataforma de Captura

A captura dos pacotes de sondagem emitidos pelos *smartphones* é feita utilizando um ou mais computadores intermediários (*sniffers*). Como eles devem situar dentro do veículo, recomenda-se utilizar computadores embarcados, tais como Raspberry Pi e Arduino, devido ao tamanho reduzido e facilidade de estendê-los com sensores (Richardson e Wallace, 2014).

Neste trabalho, utilizou-se a plataforma de hardware Raspberry Pi 3B como *sniffer* devido o mesmo vir por padrão com um chip Wi-Fi (IEEE 802.11.b/g/n) que pode ser utilizado em modo monitor para captura de pacotes de rede. Também foi acoplado um sensor de GPS (BU-353) para receber os dados de posicionamento do *sniffer* e, conseqüentemente, dos passageiros. Utilizando os componentes de Wi-Fi e GPS do sistema, é possível obter a posição (latitude e longitude), intensidade do sinal (RSSI), além do emissor (MAC) e hora de envio do pacote.

3.2. Reconstrução da Matriz O/D

O algoritmo de reconstrução da matriz O/D proposto utiliza como entrada os pacotes capturados \mathcal{P} e uma função \mathcal{F} de classificação que é capaz de filtrar pacotes dos passageiros como dentro ou fora do veículo. Especificamente, como dito, neste trabalho será utilizado um classificador baseado em aprendizado de máquina, conforme o método UAI-FI, apresentado em Roriz Junior e Medrano (2018). Entretanto, ressalta-se que o algoritmo proposto espera tal classificador como entrada, sendo possível utilizar outros métodos (Mikkelsen *et al.*, 2016; Myrvoll *et al.*, 2017; Oransirikul *et al.*, 2014; Paradedda Benites *et al.*, 2018) para tal.

O algoritmo é dividido em duas partes, sendo a primeira responsável por reconstruir a origem do passageiro, enquanto que a segunda se propõe a encontrar o par destino. A primeira etapa baseia-se na utilização do filtro de classificação \mathcal{F} para detectar quando o pacote de um usuário está dentro do veículo. Ressalta-se que os filtros de classificação podem levar um tempo até serem capazes de associar que pacotes oriundos de um passageiro estejam dentro do ônibus. Por exemplo, o filtro de classificação de Mikkelsen *et al.* (2016) requer que o aparelho esteja transmitindo pacotes por no mínimo 6 minutos abaixo de um valor limite para ser considerado como dentro. Nesse sentido, o pacote classificado como dentro do passageiro não será o primeiro pacote transmitido pelo usuário, conseqüentemente, não estará próximo da origem de embarque do mesmo. Assim, destaca-se a importância de armazenar todos os pacotes de sondas recebidos no experimento para evitar perdas durante a análise.

A Figura 2 ilustra as etapas do Algoritmo 1 para reconstruir as origens a partir dos pacotes de sonda \mathcal{P} . Ao ler um pacote de sonda p de um usuário, inicialmente se obtém o endereço *mac* do mesmo. Caso ainda não tenha detectado a origem do pacote o algoritmo prossegue, no caso oposto o algoritmo passa para o próximo pacote, visto que já se teve a origem deste usuário.

Algoritmo 1: RECONSTRUÇÃO DAS ORIGENS

Entrada: O conjunto de pacotes \mathcal{P} , um filtro de pacotes \mathcal{F} , o raio r de influência de um ponto de parada e a lista de pontos de parada \mathcal{L}

Saída: A tabela \mathcal{O} com os pontos de origem de cada endereço MAC

```

1  $\mathcal{O} \leftarrow \emptyset$            ▷ Tabela com as origens de cada MAC, inicialmente vazia
2  $\mathcal{M} \leftarrow \emptyset$       ▷ Tabela com os pacotes de cada MAC, inicialmente vazia
3 para cada pacote  $p$  de  $\mathcal{P}$  faça
4    $mac \leftarrow$  obtenha o endereço MAC de  $p$ 
5   se ainda não calculou a origem de  $mac$  em  $\mathcal{O}$  então
6     armazene o pacote  $p$  do  $mac$  na tabela  $\mathcal{M}$ 
7     se filtro  $\mathcal{F}(p)$  diz que  $p$  está dentro do ônibus então
8        $q \leftarrow$  obtenha o primeiro pacote do  $mac$  em  $\mathcal{M}$ 
9       se  $q$  está dentro do raio  $r$  de algum ponto  $l$  da lista  $\mathcal{L}$  então
10         $\mathcal{O}[mac] \leftarrow l$ 
11      senão
12         $\mathcal{O}[mac] \leftarrow$  ponto anterior a posição de  $p$ 
13      fim
14    fim
15  fim
16 fim
17 retorna  $\mathcal{O}$ 

```

Figura 2: Algoritmo para detecção dos pontos de embarque a partir dos pacotes Wi-Fi.

Nos próximos passos, armazena-se p caso o classificador venha a demorar detectar os pacotes deste usuário como dentro do veículo. Em seguida, é feita uma verificação se o mesmo se encontra dentro do veículo. Para isso, utiliza-se o filtro \mathcal{F} recebido como entrada. Caso negativo, passa-se para o próximo pacote. Caso positivo, recupera-se o primeiro pacote q emitido daquele MAC para analisar sua localização. Caso a emissão tenha ocorrido dentro de um raio r de algum de ponto de parada l , e.g. a 15 metros de l , associa-se a origem do embarque ao respectivo ponto; caso esteja fora do raio de todos os pontos, associa-se o embarque ao ponto anterior à captura do primeiro pacote q como origem.

A fim de ilustrar a idéia do algoritmo considere o cenário ilustrado na Figura 3. Considere que o *smartphone* embarcou em P1 e transmitiu 3 pacotes de sondas. Somente no terceiro pacote p a função classificadora \mathcal{F} associou o pacote como dentro do ônibus. Nesse estágio, o algoritmo retorna ao primeiro pacote de sonda recebido q , mesmo que este não seja classificado como dentro pelo classificador. O próximo passo é analisar q . Se o mesmo ocorreu dentro do raio de algum ponto de parada, associa-se o respectivo ponto como origem. Caso contrário, volta-se um passo e associa-se o ponto de parada anterior. Por exemplo, na ilustração o pacote ocorre entre P1 e P2. Como o mesmo não se encontra dentro do raio dos pontos, associa-se a origem o ponto P1. Voltando um passo espera-se poder compensar o tempo de atraso na transmissão dos pacotes de sonda. Assim, assume-se que a origem veio anterior a transmissão, visto que o pacote não se encontra dentro da vizinhança de um ponto de parada.

Semelhante a reconstrução da origem, para determinar o destino dos passageiros o algoritmo extrapola o desembarque conforme a localização da última posição obtida no pacote de sonda. A Figura 4 sintetiza as etapas envolvidas no algoritmo de detecção do local de desembarque. Aqui, assume-se que um passageiro desembarcou se o *smartphone* do mesmo não transmitiu nenhum pacote de sonda em um período máximo de Δ segundos. Caso o mesmo não possua a origem previamente cadastrada em \mathcal{O} , isto é, o mesmo não possui nenhum pacote classificado anteriormente como dentro do veículo, considera-se que este endereço MAC não esteve no veículo e prossegue-se para o próximo passageiro.

Nos casos positivos, quando o endereço MAC esteve dentro do ônibus, isto é, possui pelo menos um pacote de sonda classificado como dentro do ônibus, recupera-se o último pacote emitido q pelo passageiro. Se a localização em que tal pacote foi emitido estiver próxima a um ponto de parada (utilizando o raio r) associa-se o mesmo ao local de desembarque do passageiro. Caso contrário, associa-se o ponto de parada posterior a localização do último pacote lido. De maneira análoga ao processo de origem, a idéia é considerar o próximo ponto de parada, frente ao último pacote capturado, para compensar o atraso na transmissão dos pacotes.

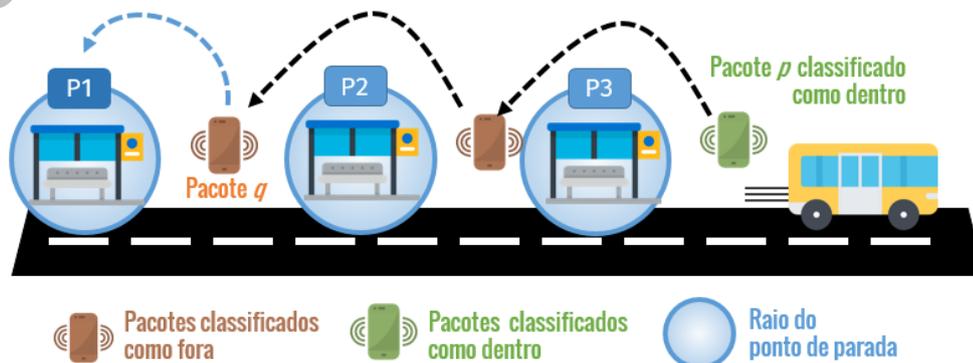


Figura 3: Exemplo de reconstrução da origem.

Algoritmo 2: RECONSTRUÇÃO DOS DESTINOS

Entrada: O conjunto de pacotes \mathcal{P} , a tabela de origens \mathcal{O} , o tempo máximo de transmissão Δ , o raio r de influência de um ponto de parada e a lista de pontos de parada \mathcal{L}

Saída: A tabela \mathcal{D} com os pontos de destino de cada endereço MAC

```
1  $\mathcal{D} \leftarrow \emptyset$  ▷ Tabela com os destinos de cada MAC, inicialmente vazia
2 para cada  $mac$  que não emitiu um pacote nos últimos  $\Delta$  s faça
3   se existe a origem de  $mac$  em  $\mathcal{O}$  então
4      $q \leftarrow$  obtenha o último pacote do  $mac$  em  $\mathcal{P}$ 
5     se  $q$  está dentro do raio  $r$  de algum ponto  $l$  da lista  $\mathcal{L}$  então
6        $\mathcal{D}[mac] \leftarrow l$ 
7     senão
8        $\mathcal{D}[mac] \leftarrow$  ponto posterior a posição de  $p$ 
9     fim
10  fim
11 fim
12 retorna  $\mathcal{D}$ 
```

Figura 4: Algoritmo para detecção dos pontos de desembarque a partir dos pacotes Wi-Fi.

4. EXPERIMENTOS E RESULTADOS PRELIMINARES

Para avaliar o método foi realizado um estudo preliminar em uma linha do Arco Sul da Rede Metropolitana de Transportes Coletivos de Goiânia RMTc (2018). A linha escolhida (563), ilustrada na Figura 5, possui uma extensão de 10,5 km e 24 pontos de paradas. Essa foi escolhida devido ser do tipo circular, que começa e termina no Terminal Araguaia, passando por diversos tipos de uso e ocupação de solo e vias. Especificamente, a mesma possui trechos que incidem sobre uma rodovia (BR-153), área residencial e área comercial.

4.1. Cenário de Teste

O experimento realizará uma comparação entre as matrizes obtidas com o método frente àquela obtida manualmente. De tal maneira, será utilizado os seguintes critérios para comparar os pares origem-destino, a saber: diferença entre os pontos de paradas nas duas matrizes e distância entre os pontos de parada em metros, $Dist_{origem}$ e $Dist_{destino}$ respectivamente. Além disso, será mensurado o tempo médio T_m de envio dos mesmos ao longo dos trajetos.

A avaliação da reconstrução da matriz O/D a partir do algoritmo proposto será feita através da comparação com a matriz real realizada. Assim, é necessário produzir manualmente a matriz O/D para tal comparação. Para isso, foi feito um estudo de campo prévio para determinar dois cenários de subida e descida de passageiros de forma que o tempo e a distância seja razoável para a detecção dos pacotes. O primeiro trecho, ilustrado pela cor amarela na Figura 5, possui um comprimento de 5.216 km, inicia-se no Terminal Araguaia e após 13 paradas termina em uma área industrial. Já o segundo trecho, ilustrado pela cor vermelha, possui um comprimento de 4.578 km. Inicia-se na área industrial e após 10 paradas termina em um ponto de parada na rodovia BR-153. O tempo de viagem de ambos trajetos é de aproximadamente 10 minutos.

O embarque e desembarque será realizado estritamente no começo e no final do percurso, isto é, o pesquisador irá adentrar no ponto de parada do começo do trecho e sair somente no último ponto. Durante o trajeto, o pesquisador irá carregar uma caixa contendo 4 *smartphones* com Wi-Fi ligado, representando assim a movimentação de 4 passageiros no local de embarque e desembarque. Com intuito de testar os limites da tecnologia Wi-Fi, os dispositivos utilizados



Figura 5: Mapa da linha 563 do Arco Sul da RMTc. Extraído de Google Earth (2019).

nos experimentos permaneceram com a tela travada e sem uso durante o trajeto, representando assim o pior caso de envio de pacotes de sondagem. Os dispositivos utilizados nos experimentos preliminares estão descritos na Tabela 1.

Tabela 1: Lista de *smartphones* utilizados nos testes preliminares.

Fabricante	Sistema Operacional	Versão	MAC
Sony	Android	4.2.2	68:76:4F:0A:60:DA
Samsung	Android	4.1.4	1C:AF:05:BE:79:6D
Samsung	Android	4.2.2	78:A8:73:95:37:A6
Samsung	Andróid	4.1.2	38:2D:D1:C1:80:FE

A fim de aumentar o número de experimentos, utilizou-se um carro de passeio para percorrer e repetir os testes nos percursos. Percorreu-se os dois trechos 3 vezes, totalizando 6 viagens. Assim, considerando o uso de 4 *smartphones* em cada trajeto, o experimento totaliza um total de 12 embarques/desembarques. Os testes ocorreram durante dias úteis no período de entropico de 09:00 às 12:00 e 15:00 às 18:00h dos dias 8, 9 e 10 de Julho de 2019.

Para avaliar o número de pacotes de sonda Wi-Fi capturados com sucesso, também se variou os mecanismos de captura. Utilizou-se duas configurações de *sniffers*. A primeira, com somente um dispositivo, que opera no modo dinâmico, *i.e.*, percorre todos os canais dinamicamente. Já a segunda configuração, utiliza três *sniffers*, cada um deles fixado nas frequências dos canais 1, 6 e 10 respectivamente. Por fim, considerou-se um raio r de 50 metros de influência para os pontos e o tempo máximo Δ de transmissão equivalente ao tempo do trajeto, 10 minutos.

4.2. Resultados Preliminares

Após a execução do método obteve-se os resultados ilustrados na Tabela 2. Note que o tempo médio T_m , requerido para detectar o primeiro pacote de sonda, situou-se no intervalo de 204 a 268 segundos para as diferentes configurações de *sniffers*. Considerando uma velocidade média de 30 km/h para os ônibus, a detecção dos pacotes ocorrerá após aproximadamente 1700 metros, *e.g.*, a 1700 metros do embarque. As métricas $Dist_{origem}$ e $Dist_{destino}$ medem justamente isso. Elas expressam a distância média da distância entre do primeiro e do último pacote emitido pelo passageiro respectivamente ao ponto de embarque e a parada de desembarque na matriz manual. Observe que os valores $Dist_{origem}$ e $Dist_{destino}$ obtidos são menores que os referidos

1700 metros. Uma possível explicação para tal diferença é que a velocidade média do veículo pode ser inferior a 30 km/h, principalmente considerando paradas, o que consequentemente reduziria a distância percorrida até o primeiro e último pacote.

Note que o número de *sniffers* tem influência no resultado. Precisamente, os dados sugerem que o tempo médio de detecção dos pacotes de sonda T_m é um pouco menor na configuração com três *sniffers*. Apesar da diferença ser relativamente pequena, 64 segundos, durante esse intervalo um determinado ônibus pode passar por um ponto de parada e modificar a matriz gerada.

Tabela 2: Lista de *smartphones* utilizados nos testes preliminares.

Configuração	$Dist_{origem}$	$Dist_{destino}$	T_m
<i>Sniffer</i> Dinâmico	1348 m	1347 m	268 s
Três <i>Sniffers</i> Fixos	1010 m	1450 m	204 s

A Figura 6 ilustra a matriz O/D obtida através do algoritmo para o primeiro trecho. Recorde que a matriz real apresenta o embarque de 12 passageiros no primeiro ponto e o desembarque dos mesmos apenas na última parada (14). As curvas da linha de cima da imagem representam a matriz obtida no cenário de um único *sniffer*, enquanto que as linhas abaixo ilustram aquela obtida ao utilizar três *sniffers* fixos.

Primeiramente, note que tanto os pares de origem, quanto os pares de destinos, apontados pelo algoritmo situam aproximadamente a uma distância máxima de 4 paradas referente ao local de embarque e desembarque respectivamente. Precisamente, a primeira matriz possui 6 dos 9 pares de origem-destino dentro do intervalo de 4 paradas. A segunda configuração apresenta 8 de 12 pares no mesmo intervalo. Entretanto, observe que no primeiro caso, o método foi capaz de detectar apenas 9 dos 12 passageiros. A fim de verificar um possível problema no filtro de classificação, analisou-se o registro de todos os pacotes armazenados e verificou-se a ausência daqueles associados aos passageiros que não foram detectados pelo *sniffer* dinâmico, ou seja, o *sniffer* não capturou os pacotes de tais *smartphones*.

Também é notável as curvas de origem-destino que ligam pontos sequenciais, tais como as que ligam a parada 3 para 4 e parada 12 a 13. Esses casos ilustram os cenários em que os dispositivos enviaram apenas um único pacote de sondagem durante o trajeto de 10 minutos. Desta maneira, o algoritmo associa o ponto de parada inicial como origem e o posterior como destino.

Em relação a configuração de três *sniffers*, note que a mesma apresenta um resultado superior àquela de um único *sniffer*. Os dados sugerem que a utilização de três aparelhos diminuiu um pouco o tempo médio T_m para recebimento dos pacotes. Esta redução é ilustrada através da

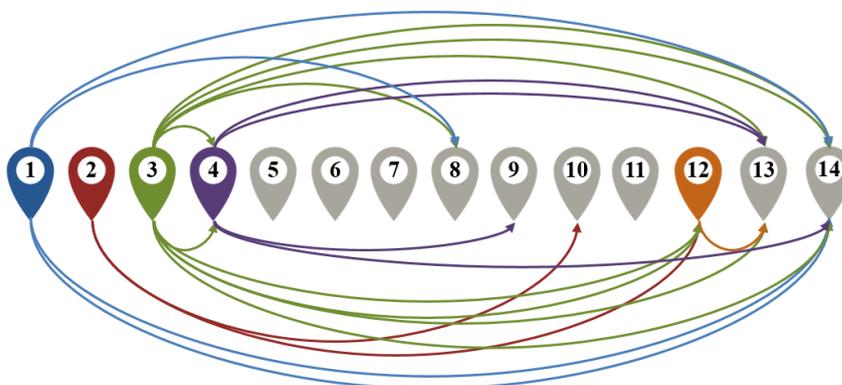


Figura 6: Matriz O/D reconstruída para o primeiro trecho. As linhas acima representam a matriz obtida com um *sniffer*, enquanto que as de baixo a obtida com três *sniffers*.

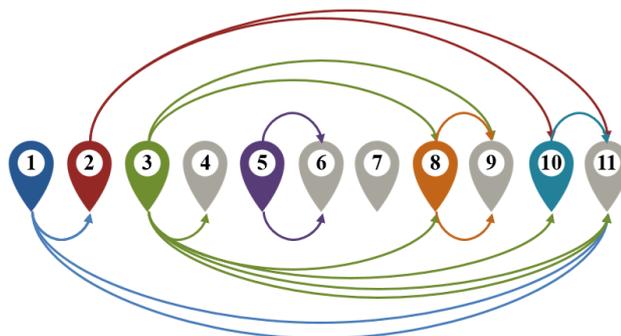


Figura 7: Matriz O/D reconstruída para o segundo trecho. As linhas acima representam a matriz obtida com um *sniffer*, enquanto que as de baixo a obtida com três *sniffers*.

matriz através da associação do embarque de alguns passageiros ao segundo ponto de parada, ao invés do terceiro. Além disso, observa-se uma tendência maior dos destinos dos passageiros estarem atrelado ao ponto final do trajeto (14) ao invés do ponto anterior (13).

A Figura 7 ilustra a matriz obtida para o segundo trajeto. Observe que neste caso houve uma redução em ambas matrizes da quantidade de passageiros detectados. A primeira configuração detectou o par origem-destino de um total de 7 passageiros, sendo que 3 desses estão incorretos por possuírem apenas um pacote de sonda, como os pares 5-6 e 8-9. Um resultado semelhante é visto na matriz com múltiplo *sniffers*. Aqui, observa-se a detecção do deslocamento de 10 de 12 passageiros, sendo que 4 estão incorretos, isto é, são pontos subsequentes.

Uma explicação para esse fato decorre da distância do segundo trajeto ser inferior ao primeiro e que o mesmo atravessa vias de trânsito rápido, como a rodovia BR-153. Assim, rapidamente o veículo cobre o trajeto. Nesse sentido, o tempo de percurso pode ser inferior a transmissão do pacote de sondagem subsequente, o que ocasionaria as falhas apontadas.

5. CONCLUSÃO

Este trabalho investigou a viabilidade de se reconstruir os pares de deslocamento, a matriz O/D, dos passageiros de transporte público urbano a partir de dados Wi-Fi. Abordou-se questões técnicas e limitações inerentes a tecnologia Wi-Fi, como a irregularidade na transmissão dos pacotes de sonda e a respectiva transmissão em diferentes canais de frequência. Para responder essas questões, apresenta-se um algoritmo para reconstrução da matriz O/D a partir das limitações apontadas. O algoritmo utiliza como entrada uma função de classificação de pacotes, de maneira a tornar o trabalho independente do filtro utilizado. Para validação do mesmo foi feito sua codificação na linguagem de programação Python e um estudo de caso em uma linha da Rede Metropolitana de Transporte Coletivo de Goiânia.

Os resultados preliminares sugerem que a matriz reconstruída apresenta similaridade com a real ao considerar um intervalo de pontos de parada frente ao local de embarque e desembarque do passageiro. Especificamente, os pares de origem-destino obtidos não necessariamente serão idênticos e refletirão o ponto de subida e descida do passageiro, entretanto apresenta semelhança ao considerar pontos adjacentes. De fato, os dados ilustram que a maioria dos pares de origem e destino situaram em um intervalo de até 4 pontos da origem e do destino real. Com base nessa característica, pode-se sugerir que a utilização do método venha a ser feita através do uso de regiões maiores, como zonas de tráfego.

Como trabalho futuro, espera-se abordar diversas limitações do trabalho realizado. Dentre elas, destaca-se testes com uma diversidade maior de aparelhos e em linhas de diferentes naturezas, como linhas expressas e alimentadoras. Além disso, deseja-se investigar o comportamento do método ao utilizar zonas de tráfego.

REFERÊNCIAS BIBLIOGRÁFICAS

- Arbex, R. O., e da Cunha, C. B. (2017) Estimaco da matriz origem-destino e da distribuico espacial da lotaco em um sistema de transporte sobre trilhos a partir de dados de bilhetagem eletrnica. *TRANSPORTES*, 25(3), 166.
- Arnott, N. (2014) What's really happening with iOS 8 MAC address randomization? *iMore*. Disponvel em <https://www.imore.com/closer-look-ios-8s-mac-randomization>. Acesso em 19 de janeiro de 2019.
- Chen, C., Ma, J., Susilo, Y., Liu, Y., e Wang, M. (2016) The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies*, 68, 285–299.
- Crow, B. P., Widjaja, I., Kim, J. G., e Sakai, P. T. (1997) IEEE 802.11 Wireless Local Area Networks. *IEEE Communications Magazine*, 35(9), 116–126.
- Danalet, A., Farooq, B., e Bierlaire, M. (2014) A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures. *Transportation Research Part C: Emerging Technologies*, 44, 146–170.
- Ferraz, A. C. P., e Torres, I. G. E. (2004) *Transporte Pblico Urbano - Segunda Edio - Ampliada e Atualizada*. RiMa Editora, Segunda Edio.
- Freudiger, J. (2015) How Talkative is your Mobile Device? An Experimental Study of Wi-Fi Probe Requests. *WiSec Proc. of the 8th ACM Conf. on Security & Privacy in Wireless and Mobile Networks* (p. 1–6). ACM.
- Gast, M. S. (2005) *802.11 Wireless Networks: The Definitive Guide, Second Edition*. O'Reilly Media, Inc.
- Google Earth, Google (2019). Disponvel em <http://earth.google.com/web>. Acesso em 15 de junho de 2019.
- Guerra, A. L., Barbosa, H. M., e Oliveira, L. K. de. (2014) Estimativa de matriz origem/destino utilizando dados do sistema de bilhetagem eletrnica: proposta metodolgica. *TRANSPORTES*, 22(3), 26.
- Hertzog, R., e O'Gorman, J. (2017) *Kali Linux Revealed: Mastering the Pen. Testing Distribution*. Offsec Press.
- LifeHacker. (2019) Which Wifi Channels Should I Use For My Wireless Network? *LifeHacker*. Disponvel em <https://www.lifehacker.com.au/2019/03/which-wifi-channels-should-i-use-for-my-wireless-network/>. Acesso em 22 de maro de 2019.
- Martin, J., Mayberry, T., Donahue, C., Foppe, L., Brown, L., Riggins, C., Rye, E. C., e Brown, D. (2017) A Study of MAC Address Randomization in Mobile Devices and When it Fails. *Proceedings on Privacy Enhancing Technologies*, 2017(4), 365–383.
- Mikkelsen, L., Buchakchiev, R., Madsen, T., e Schwefel, H. P. (2016) Public transport occupancy estimation using WLAN probing. *Proc. of 2016 8th Intl. Wksh on Resilient Net. Design and Mod., RNDM 2016*, 302–308.
- Mohammed, S. L. (2016) Distance Estimation Based on RSSI and Log-Normal Shadowing Models for ZigBee Wireless Sensor Network. *Engineering and Technology Journal*, 34(15), 2950–2959.
- Myrvoll, T. A., Hkegrd, J. E., Matsui, T., e Septier, F. (2017) Counting public transport passenger using WiFi signatures of mobile devices. *2017 IEEE 20th Intl. Conf. on Intelligent Transportation Systems (ITSC)*.
- Oransirikul, T., Nishide, R., Piumarta, I., e Takada, H. (2014) Measuring Bus Passenger Load by Monitoring Wi-Fi Transmissions from Mobile Devices. *Procedia Technology*, 18(September), 120–125.
- Oransirikul, T., Nishide, R., Piumarta, I., e Takada, H. (2016) Feasibility of analyzing Wi-Fi activity to estimate transit passenger population. *Proc. Intl. Conf. on Adv. Info. Networking and Applications, AINAMay*.
- Paradedo Benites, D., Kraus Junior, W., e Carlson, R. C. (2018) ESTIMATIVA DE EMBARQUE E DESEMBARQUE DE PASSAGEIROS POR DETECCO DE SINAL DE Wi-Fi DE DISPOSITIVOS MVEIS. *Anis do XXXII Congresso Nacional de Pesquisa em Transporte*. ANPET, Gramado, RS, Brazil.
- Pelletier, M.-P., Trpanier, M., e Morency, C. (2011) Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4), 557–568.
- Richardson, M., e Wallace, S. (2014) *Getting Started with Raspberry Pi: Electronic Projects with Python, Scratch, and Linux*. (2nd ed). Maker Media, Inc, USA.
- RMTC, Rede Metropolitana de Transporte Coletivo de Goinia (2018). Disponvel em <http://rmtcgoiania.com.br/>. Acesso em 10 de julho de 2019.
- Roriz Junior, M. P., e Medrano, R. A. (2018) UAI-FI: UM MTODO BASEADO EM APRENDIZADO DE MQUINA PARA CONTAGEM AUTOMTICA DE PASSAGEIROS UTILIZANDO SINAIS WI-FI. *Anis do XXXII Congresso Nacional de Pesquisa em Transporte*. ANPET, Gramado, RS, Brazil.
- Soares, R. (2016) Transporte coletivo urbano perdeu 215 milhes de passageiros em um ano e deixou de arrecadar R\$ 680 milhes. *Jornal do Comrcio*. Disponvel em <http://jc.ne10.uol.com.br/blogs/deolhonotransito/2016/05/25/8983/>. Acesso em 20 de fevereiro de 2019.

Paulo de Tarso Gonalves Sousa Jnior (paulodetarsojr@discente.ufg.br)

Marcos Paulino Roriz Junior (marcosroriz@ufg.br)

Kamilla Cruvinel Rodrigues (kcruvinel@discente.ufg.br)

Ronny Marcelo Aliaga Medrano (ronnymarcelo@ufg.br)

Engenharia de Transportes, Faculdade de Cincias e Tecnologia, Universidade Federal de Gois

Rua Mucuri S/N, rea 03, Bairro Conde dos Arcos, Aparecida de Goinia - GO