



UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE AUTOMAÇÃO E
SISTEMAS

Diego Benites Paradedá

**Método probabilístico para estimativa de embarques de passageiros em
transporte coletivo por ônibus utilizando dados de sinalizações Wi-Fi**

Florianópolis

2020

Diego Benites Paradedá

Método probabilístico para estimativa de embarques de passageiros em transporte coletivo por ônibus utilizando dados de sinalizações Wi-Fi

Tese submetida ao Programa de Pós-Graduação em Engenharia de Automação e Sistemas da Universidade Federal de Santa Catarina para a obtenção do título de doutor em Engenharia de Automação e Sistemas.

Orientador: Prof. Werner Kraus Junior, Dr.

Coorientador: Prof. Rodrigo Castelan Carlson, Dr.

Florianópolis

2020

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Paradedda, Diego Benites

Método probabilístico para estimativa de embarques de passageiros em transporte coletivo por ônibus utilizando dados de sinalizações Wi-Fi / Diego Benites Paradedda ; orientador, Werner Kraus Junior, coorientador, Rodrigo Castelan Carlson, 2021.

143 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia de Automação e Sistemas, Florianópolis, 2021.

Inclui referências.

1. Engenharia de Automação e Sistemas. 2. Pesquisa O/D em transporte público. 3. Detecção de dispositivos móveis por meio de Wi-Fi. 4. Estimação bayesiana de passageiros em ônibus. I. Kraus Junior, Werner. II. Castelan Carlson, Rodrigo. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia de Automação e Sistemas. IV. Título.

Diego Benites Paradedda

Método probabilístico para estimativa de embarques de passageiros em transporte coletivo por ônibus utilizando dados de sinalizações Wi-Fi

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Claudio Luiz Marte, Dr.
Universidade de São Paulo - USP

Prof. Marcos Paulino Roriz Junior, Dr.
Universidade Federal de Goiás - UFG

Prof. Laio Oriel Seman, Dr.
Universidade do Vale do Itajaí - UNIVALI

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de doutor em Engenharia de Automação e Sistemas.

Coordenação do Programa de
Pós-Graduação

Prof. Werner Kraus Junior, Dr.
Orientador

Florianópolis, 2020.

AGRADECIMENTOS

Primeiramente agradeço a Deus por me dar força sempre que precisei durante a realização desse trabalho. A minha mãe, que sempre lutou para que eu conseguisse alcançar esse objetivo desde o início dos meus estudos até esse momento e infelizmente não conseguiu estar presente comigo nessa etapa final. Ao meu irmão e meu Pai por sempre me apoiarem nas decisões que tomei ao longo dessa trajetória. Agradeço a minha namorada Daniela por ser paciente durante esse doutorado que consumiu bastante tempo de nós dois. Agradeço muito ao meu orientador Werner, por todo o empenho e dedicação, as várias reuniões durante semanas para encontrarmos o método ideal para o trabalho. Ao meu coorientador Rodrigo pela ajuda na etapa de escrita desta Tese. A Universidade Federal de Santa Catarina por ter me concedido a oportunidade de realizar os estudos nessa universidade de excelência e a CAPES e CNPQ por apoiar financeiramente meus estudos.

*“However bad life may seem,
there is always something
you can do, and succeed at.
While there’s life, there is hope.”
(Stephen Hawking)*

RESUMO

O planejamento e a operação eficientes de sistemas de transporte coletivo dependem de dados atualizados sobre os carregamentos dos ônibus ao longo dos itinerários e as origens e destinos dos passageiros transportados. Métodos tradicionais baseados em contagens manuais de passageiros têm custo elevado e realização onerosa. Nesta tese, propõe-se um método probabilístico para estimar os volumes e os locais de embarque de passageiros a partir de informações coletadas do Wi-Fi de dispositivos móveis dos passageiros embarcados. Para a coleta, foram desenvolvidos um detector que monitora e armazena as buscas de rede e uma aplicação embarcada nos dispositivos que registra com que frequências os dispositivos móveis realizam as buscas de rede. Inicialmente, os dados coletados pelo detector foram usados de forma agregada para estimação de passageiros, com resultados animadores. Porém, a análise individualizada dos dados indicou viés causado pela contagem de falsos positivos, isto é, detecções de pessoas fora do ônibus contabilizadas como passageiros a bordo. A partir desse resultado, e com a aferição do detector por meio do uso da aplicação embarcada, obtiveram-se distribuições de probabilidade dos intervalos entre buscas de rede sucessivas realizadas pelos dispositivos móveis. Um método de estimação bayesiana foi aplicado sobre as distribuições obtidas tanto experimentalmente como por meio de acesso a dados de estudos publicados por terceiros e disponibilizados para acesso público. Simulações em sistema de informações geográficas permitiram espacializar amostras de detecções para aplicação do método. Resultados indicam a adequação da proposta para levantamentos de zonas de tráfego com vários pontos de ônibus, bem como a perda de qualidade na estimação de embarques em pontos específicos quando o tempo de viagem entre pontos é pequeno em face dos intervalos de detecção.

Palavras-chave: Pesquisa O/D em transporte público. Detecção de dispositivos móveis por meio de Wi-Fi. Estimação bayesiana de passageiros em ônibus.

ABSTRACT

The efficient planning and operation of public transport systems depends on up-to-date data about bus loads along routes and the origins and destinations of transported passengers. Traditional methods based on manual passenger counts are expensive and costly. In this thesis, a probabilistic method is proposed to estimate the volumes and the passenger's boarding locations from information collected from the Wi-Fi of mobile devices of the boarded passengers. For the data collection, a detecting device that monitors and stores the network searches and an application embedded in the mobile devices that records how often the mobile devices perform the network searches were developed. Initially, the data collected by the detecting device were used in an aggregate form for passenger estimation, with encouraging results. However, the individual analysis of the data indicated bias caused by the false positives counts, that is, detections of people outside the bus counted as passengers on board. As follow up, and with the gauging of the detector using the embedded mobile application, probability distributions of the intervals between successive network searches performed by mobile devices were obtained. A Bayesian estimation method was applied to distributions obtained experimentally and through access to data from studies published by third parties that were available for public access. Simulations with a geographic information system allowed the spatialization of samples of detections for the application of the method. Results indicate the adequacy of the proposed method for surveying traffic zones with several bus stops, as well as the loss of quality in the estimation of boardings at specific bus stops when the travel time between bus stops is short when compared to the detection intervals.

Keywords: O/D for public transit. Detection of mobile devices through Wi-Fi. Bayesian estimation of bus passengers.

LISTA DE FIGURAS

Figura 1 – Exemplo de usuários portando dispositivos móveis, antes durante e depois da chegada de um ônibus equipado com um MTR.	25
Figura 2 – Exemplo de usuário portando dispositivo móvel desembarcando um ponto de ônibus enquanto um passageiro permanece a bordo.	26
Figura 3 – Formato do quadro MAC.	36
Figura 4 – Controle de quadro.	37
Figura 5 – Representação gráfica dos canais Wi-Fi na frequência de 2.4 GHz.	39
Figura 6 – Componentes do dispositivo MTR.	44
Figura 7 – Ilustração de como o MTR age quando há troca de pacotes entre um roteador e um dispositivo móvel.	44
Figura 8 – Potência do sinal registrada com dispositivo móvel em repouso em marcações espaçadas de 1 metro entre si.	49
Figura 9 – 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.	50
Figura 10 – Histograma (a) tempo da primeira detecção e (b) segunda detecção agrupados.	53
Figura 13 – Trajeto realizado pela linha. Os pontos de paradas de ônibus estão identificados com as letras O, A, B e C. Os segmentos do trajeto nas linhas tracejadas representam o deslocamento do veículo entre dois pontos de ônibus consecutivos.	56
Figura 11 – Tempo para a primeira detecção por dispositivo no experimento controlado (0 s significa que o dispositivo não foi detectado): (a) Experimento 1; (b) Experimento 2; (c) Experimento 3.	61
Figura 12 – Tempo para a segunda detecção por dispositivo no experimento controlado (0 s significa que o dispositivo não foi detectado): (a) Experimento 1; (b) Experimento 2; (c) Experimento 3.	62
Figura 14 – Contagem de dispositivos móveis acumulados durante o trajeto no experimento com cenário não controlado, com contagem manual de passageiros embarcados: (a) Primeiro rodada, (b) Segunda rodada, e (c) Terceira rodada.	63
Figura 15 – Fluxograma de execução da aplicação de buscas de rede para dispositivos móveis.	66
Figura 16 – Número de buscas realizadas pelo dispositivo móvel de controle registradas pela aplicação.	68

Figura 17 – (a) Buscas de rede realizadas pelos seis voluntários com dados agrupados em 30 minutos para sete períodos. (b) Distribuição de probabilidade do intervalo de detecções.	71
Figura 18 – (a) Número de detecções de buscas de rede com dispositivos distintos registrados no experimento Embarcado (Capítulo 4.3); (b) média do intervalo de buscas de rede.	72
Figura 19 – (a) Número de detecções de buscas de rede com dispositivos distintos registrados no experimento realizado por Nunes (2018); (b) média do intervalo de buscas de rede.	73
Figura 20 – Passageiro embarcando no ponto l onde na representação sequencial é o ponto de ônibus $k - 2$ e sendo detectado após o ponto k	75
Figura 21 – Exemplo de um dispositivo móvel $i = 3$, detectado após o ponto de embarque PT-3. As informações de distância entre os pontos são conhecidas. Sendo Δt_{il} , 497 s (ponto PT - 1 até a detecção), 173 s (ponto PT - 2), 38 s (ponto PT - 3).	76
Figura 22 – Probabilidade desejadas da distribuição Embarcado considerada como estudo de caso. (a) Probabilidade para o intervalo onde $l = 1$. (b) para $l = 2$. (c) para $l = 3$	77
Figura 23 – Passageiros que possivelmente embarcaram no ponto 3.	80
Figura 24 – Comparação do resultado das probabilidades entre o método por combinação e por classes.	89
Figura 25 – Ponto agrupados por Zonas de embarque.	90
Figura 26 – Passageiro desembarcando no ponto l onde na representação sequencial é o ponto de ônibus $k + 2$ e sua última detecção ocorreu após o ponto k	91
Figura 27 – Trajeto da linha 185-UFSC SEMIDIRETO com 17 pontos de paradas de ônibus, onde EMB é o número de embarques conhecidos de cada ponto.	95
Figura 28 – (a,c,e) Número de detecções de buscas de rede com dispositivos distintos registrados; (b,d,f) média do intervalo de buscas de rede. (a,b) Experimento embarcado (Capítulo 4.3); (c,d) Experimento aplicativo (Capítulo 5); (e,f) Experimento realizado por Nunes (2018)	98
Figura 29 – Exemplo de detecções gerados no software Transcad (CALIPER, 2020) através de uma distribuição uniforme, onde o mínimo é igual a 1 e o máximo é igual a duas vezes a média da distribuição escolhida através da Tabela 22.	100
Figura 30 – Descarte de detecção quando a geração ocorre em um intervalo de tempo maior que o trajeto até o ponto 1.	101
Figura 31 – Detecções geradas para 25 embarques no ponto PT - 1. (a) Distribuição Embarcado, (b) Distribuição Aplicativo e (c) Distribuição Nunes (2018).	103

Figura 32 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 1 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	104
Figura 33 – Detecções geradas para 2 embarques no ponto PT - 2. (a) Distribuição Embarcado, (b) Distribuição Aplicativo e (c) Distribuição Nunes (2018).	105
Figura 34 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 2 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	106
Figura 35 – Detecções geradas para 4 embarques no ponto PT - 3. (a) Distribuição Embarcado, (b) Distribuição Aplicativo e (c) Distribuição Nunes (2018).	107
Figura 36 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 3 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	108
Figura 37 – Detecções geradas para 1 embarque no ponto PT - 4. (a) Distribuição Embarcado, (b) Distribuição Aplicativo e (c) Distribuição Nunes (2018)	109
Figura 38 – Estimativa de embarques realizados no ponto de parada de ônibus P4 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	110
Figura 39 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 5 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	112
Figura 40 – Estimativa de embarques realizados no ponto de embarque PT - 6 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	113
Figura 41 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 7 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	114
Figura 42 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 8 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	115

Figura 43 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 9 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	116
Figura 44 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 10 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	117
Figura 45 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 11 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	118
Figura 46 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 12 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	119
Figura 47 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 13 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	120
Figura 48 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 14 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	121
Figura 49 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 15 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	122
Figura 50 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 16 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	123
Figura 51 – Estimativa de embarques realizados no ponto de parada de ônibus PT - 17 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	124

Figura 52 – Estimativa de embarques realizados na zona de embarque Z1 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	126
Figura 53 – Estimativa de embarques realizados na zona de embarque Z2 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	127
Figura 54 – Estimativa de embarques realizados na zona de embarque Z3 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	128
Figura 55 – Estimativa de embarques realizados na zona de embarque Z4 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	129
Figura 56 – Estimativa de embarques realizados na zona de embarque Z5 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	130
Figura 57 – Estimativa de embarques realizados na zona de embarque Z6 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).	131

LISTA DE TABELAS

Tabela 1 – Países com o maior número de usuários de <i>smartphones</i>	32
Tabela 2 – Utilização de sistema operacional por países com maior uso de <i>smartphone</i>	33
Tabela 3 – Market Share de Fabricantes no Brasil.	34
Tabela 4 – Estrutura do EUI-48.	35
Tabela 5 – Tabela com os campos de tipos e subtipos do Controle de Quadro. . .	37
Tabela 6 – Fabricantes com registros de <i>Organizationally Unique Identifier</i> (OUI). .	39
Tabela 7 – Detecções coletadas por um MTR e quatro MTR's durante 2 horas. . .	47
Tabela 8 – Intervalos entre detecções sucessivas de um mesmo dispositivo e detec- ções acumuladas em todas as rodadas de todos os experimentos.	53
Tabela 9 – Resultados de quatro rodadas de cada experimento; são esperadas 48 detecções por experimento (12 dispositivos, 4 rodadas).	54
Tabela 10 – Estimativa de embarcados pelo MTR x Embarcados contagem manual; Estimativa de embarque MTR x Embarques contagem manual; Estima- tiva de desembarques pelo MTR x Desembarques contagem manual.	60
Tabela 11 – Identificação da busca de rede perdida pelo MTR e o intervalo de tempo em relação a busca de rede anterior	67
Tabela 12 – Número de buscas de rede realizadas por dispositivos durante um pe- ríodo de 24 horas.	69
Tabela 13 – Tabela com resumo dos dados obtidos na Figura 21	76
Tabela 14 – Tabela com a probabilidade de que a detecção S_i tenha ocorrido, dado que o evento B_{il} em que um passageiro portando um dispositivo móvel embarcou no exemplo da Figura 21.	77
Tabela 15 – Tabela com o exemplo do cálculo da probabilidade total para o estudo de caso.	78
Tabela 16 – Tabela com o cálculo da Equação 9, onde $\sum_{l=1}^k P(B_{il} S_i) = 1$	79
Tabela 17 – Tabela com valores de probabilidade para B_9 aplicando divisão de classes. Indicação com “o” qual classe cada probabilidade pertence ou com “x” se a probabilidade não entrar em alguma classe é descartada.	84
Tabela 18 – Tabela com o cálculo da distribuição binomial para as classes formadas por dispositivos móveis de passageiros com probabilidades de embarques no ponto PT - 9.	85
Tabela 19 – Cálculo de T_9 para as classes c_1, c_2, c_3, c_7 e c_8 , através do método da convolução.	87
Tabela 20 – Tempo de execução e cálculo da probabilidade utilizando modelo de combinação e Heurística de Classificação.	88
Tabela 21 – Tabela com as características da Linha 185-UFSC SEMIDIRETO	94

Tabela 22 – Tabela resumida das informações sobre buscas de rede obtidas de diferentes fontes de dados.	97
Tabela 23 – Dados do ponto de parada de ônibus PT - 1	102
Tabela 24 – Dados do ponto de parada de ônibus PT - 2	104
Tabela 25 – Dados do ponto de parada de ônibus PT - 3	107
Tabela 26 – Dados do ponto de parada de ônibus PT - 4	109
Tabela 27 – Dados do ponto de parada de ônibus PT - 5	112
Tabela 28 – Dados do ponto de parada de ônibus PT - 6	113
Tabela 29 – Dados do ponto de parada de ônibus PT - 7	114
Tabela 30 – Dados do ponto de parada de ônibus PT - 8	115
Tabela 31 – Dados do ponto de parada de ônibus PT - 9	116
Tabela 32 – Dados do ponto de parada de ônibus PT - 10	117
Tabela 33 – Dados do ponto de parada de ônibus PT - 11	118
Tabela 34 – Dados do ponto de parada de ônibus PT - 12	119
Tabela 35 – Dados do ponto de parada de ônibus PT - 13	120
Tabela 36 – Dados do ponto de parada de ônibus PT - 14	121
Tabela 37 – Dados do ponto de parada de ônibus PT - 15	122
Tabela 38 – Dados do ponto de parada de ônibus PT - 16	123
Tabela 39 – Dados do ponto de parada de ônibus PT - 17	124
Tabela 40 – Tabela com as médias das estimativas de embarques de cada ponto, com erro relativo aos embarques conhecidos.	125
Tabela 41 – Tabela com as médias das estimativas de embarques de cada zona. . .	128

LISTA DE ABREVIATURAS E SIGLAS

AP	<i>Access Point</i>
APC	<i>Automatic Passenger Counting</i>
BSS	<i>Basic Service Set</i>
CID	<i>Company ID</i>
CPF	Cadastro de Pessoa Física
EUI	<i>Extended Unique Identifier</i>
IEEE	<i>Institute of Electrical and Electronic Engineers</i>
IEEE RA	<i>IEEE Registration Authority</i>
MAC	<i>Media Access Control</i>
MTR	Monitor de Tráfego na Rede
NTM	<i>Network Traffic Monitoring</i>
OD	Origem/Destino
OUI	<i>Organizationally Unique Identifier</i>
SSID	<i>Service Set Identifier</i>
SVM	<i>Support Vector Machine</i>

LISTA DE SÍMBOLOS

GHz	Gigahertz
MHz	Megahertz
dB	Decibel
dBm	Decibel miliwatt
dB _{mW}	Decibel miliwatt

SUMÁRIO

1	INTRODUÇÃO	20
1.1	OBJETIVOS	22
1.1.1	Objetivo Geral	22
1.1.2	Objetivos Específicos	22
1.2	CONTRIBUIÇÃO	22
1.3	PRODUÇÃO CIENTÍFICA	23
1.4	ORGANIZAÇÃO DO DOCUMENTO	24
2	CONTEXTUALIZAÇÃO E METODOLOGIA	25
2.1	CONTEXTUALIZAÇÃO	25
2.2	REVISÃO DA LITERATURA	26
2.3	ASPECTOS REGULATÓRIOS E TECNOLÓGICOS DA PESQUISA	30
2.3.1	Lei, diretrizes e normas para pesquisas com dispositivos Wi-Fi no Brasil	30
2.3.2	<i>Market Share</i> de dispositivos	31
2.3.3	Padrão IEEE 802.11	34
2.3.4	Canais e Frequências	38
2.3.5	Fabricantes	39
2.4	METODOLOGIA	40
2.5	POSSIBILIDADE DE APLICAÇÃO PRÁTICA	42
3	IMPLEMENTAÇÃO E AFERIÇÃO DO MONITOR DE TRÁFEGO NA REDE (MTR)	43
3.1	REQUISITOS A SEREM PREENCHIDOS PELO MTR	43
3.2	ARQUITETURA PROJETADA PARA O MTR	43
3.3	AFERIÇÃO DO MTR	45
3.3.1	Teste de confiabilidade do MTR	46
3.3.2	Raio de detecção do MTR	47
4	EXPERIMENTOS DE CAMPO	50
4.1	RESULTADOS AGREGADOS	52
4.2	RESULTADOS DESAGREGADOS	54
4.3	EXPERIMENTO DE CAMPO COM DISPOSITIVO A BORDO DE UM ÔNIBUS	55
4.4	PROCESSAMENTO DOS DADOS COLETADOS A BORDO DE UM ÔNIBUS	57
4.5	RESULTADOS DO EXPERIMENTO A BORDO DE UM ÔNIBUS	58
4.6	SÍNTESE DOS RESULTADOS DO EXPERIMENTO A BORDO DE UM ÔNIBUS	59
5	APLICAÇÃO PARA REGISTROS DE BUSCAS DE REDES EM DISPOSITIVOS MÓVEIS	64

5.1	APLICAÇÃO E FLUXOGRAMA DE EXECUÇÃO	65
5.2	EXPERIMENTOS COM COLETA DE DADOS PELA APLICAÇÃO	66
5.2.1	Resultados do experimento controlado	67
5.2.2	Resultados do experimento não controlado	68
5.2.3	Eficácia do NTM e regularidade nas buscas de redes	69
6	ANÁLISE DO COMPORTAMENTO DE DISPOSITIVOS MÓ- VEIS	70
6.1	AJUSTE DO INTERVALO DE BUSCAS DE REDE PARA COMPA- TIBILIZAÇÃO DE DETECÇÃO COM COLETA A BORDO	70
6.2	EVENTO BUSCA DE REDE X_i	71
6.3	DISTRIBUIÇÃO DE PROBABILIDADE DE BUSCAS DE REDE	71
6.3.1	Utilização de dados de coleta Wi-Fi	72
7	ESTIMATIVA DE EMBARQUES DE PASSAGEIRO UTILI- ZANDO DADOS DE COLETA WI-FI E MÉTODOS ESTA- TÍSTICOS	74
7.1	MODELO DE EMBARQUES DESAGREGADOS POR PONTO DE ÔNIBUS	74
7.1.1	Estimador Bayesiano para embarques em ponto	75
7.1.2	Estimativa do Total de Embarques em pontos de parada de ônibus	79
7.2	CÁLCULO DE EMBARQUES APROXIMADO COM HEURÍSTICA DE CLASSIFICAÇÃO	82
7.3	CÁLCULO POR CLASSIFICAÇÃO VS COMBINAÇÃO	87
7.4	ESTIMATIVA DO TOTAL DE EMBARQUES EM PONTOS DE ÔNI- BUS AGREGADOS POR ZONAS	89
7.5	ESTIMATIVA DE DESEMBARQUES DE PASSAGEIROS	91
8	ESTUDO DE CASO DA ESTIMATIVA DE EMBARQUES	94
8.1	MODELO DE GERAÇÃO ARTIFICIAL DE DETECÇÕES	95
8.2	RESULTADO DO CÁLCULO DA ESTIMATIVA DO TOTAL DE EMBARQUES EM PONTOS DE PARADA DE ÔNIBUS	101
8.2.1	PONTO DE PARADA DE ÔNIBUS PT - 1	101
8.2.2	PONTO DE PARADA DE ÔNIBUS PT - 2	103
8.2.3	PONTO DE PARADA DE ÔNIBUS PT - 3	106
8.2.4	PONTO DE PARADA DE ÔNIBUS PT - 4	108
8.2.5	DEMAIS PONTOS DE PARADAS DE ÔNIBUS	111
8.2.6	Síntese dos resultados	125
8.3	RESULTADOS DO CÁLCULO DA ESTIMATIVA DO TOTAL DE EMBARQUES EM ZONAS DE EMBARQUE	125
9	CONCLUSÃO	132
	REFERÊNCIAS	135

1 INTRODUÇÃO

Conhecer a demanda é fundamental para o planejamento da oferta de ônibus adequada para um sistema de transporte coletivo. Com uma oferta baixa a qualidade do serviço fica ruim com a capacidade dos ônibus lotadas acima do permitido em algumas vezes. Com uma oferta alta pode ocorrer ociosidade e veículos com ocupação muito baixa (FERRAZ; TORRES, 2001).

Para Ceder (2016), a análise e estimativa demanda de sistemas de transporte público são problemas complexos porque envolve o comportamento dos indivíduos. Quando o planejamento de transporte está sendo realizado, o perfil da demanda de uma linha específica irá afetar diretamente no número de veículos necessários para satisfazer a demanda na hora pico (WRIGHT; HOOK *et al.*, 2007).

Dessa forma a informação do índice de renovação de passageiros, que é o volume total de passageiros transportados na viagem pelo volume de passageiros na seção crítica da viagem, se torna essencial para um planejamento adequado. Um exemplo é o planejamento do sistema de transporte coletivo da região metropolitana de Florianópolis. No planejamento do sistema, a demanda era calculada considerando a linha toda como um trecho crítico, dessa forma o fator de renovação era 1. Essa premissa de projeto implica em superdimensionamento da frota pois indica que só haveria desembarques no ponto final, o que não é razoável. Na operação real de uma linha, desembarques ao longo do itinerário reduzem a necessidade de mais capacidade para embarques posteriores, visto que foi liberado espaço para novos passageiros. Assim, o conhecimento dos movimentos de embarque e desembarque ao longo do itinerário permitem o dimensionamento correto da frota operante.

De acordo com Vuchic (2017), 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 de ônibus ao longo do trajeto. Maneiras 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, 2018). Tal método é caro e sujeito a erros, motivando a busca por métodos automáticos de contagem e identificação de passageiros.

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.

A ideia de usar sinais de Wi-Fi para detectar dispositivos móveis portáteis transportados por passageiros no transporte público surgiu como uma promessa de métodos que não precisem obter informações manuais por pesquisadores para estimar uma variedade de

dados sobre viagens dos passageiros (EL-TAWAB *et al.*, 2016; JI *et al.*, 2017; HÅKEGÅRD *et al.*, 2018; NUNES, 2018; JUNIOR; MEDRANO, 2018; RYU *et al.*, 2020). O problema encontrado na utilização do sinal Wi-Fi como forma de obter estimativas de embarques e desembarques de passageiros, está nos intervalos de tempo com a qual os dispositivos móveis realizam as buscas de redes. Os estudos atuais não trazem informações com relação ao comportamento desses dispositivos, e é um fator importante, porque fornece indícios se um dispositivo móvel será detectado em um tempo hábil ou não.

Semelhante ao Bluetooth e outras tecnologias de comunicação de curto alcance, o Wi-Fi usa strings de ID (conhecidas como endereços MAC) que identificam exclusivamente um dispositivo (IEEE, 2007). Juntamente com esquemas para assegurar que o MAC detectado se refere a um determinado dispositivo em um determinado ônibus, podem ser obtidas informações sobre a origem-destino das viagens, incluindo viagens conectadas e carregamento de ônibus entre as paradas. Assim, usando um dispositivo combinado com software especializado de código livre com capacidade para coletar sinais Wi-Fi, pode ser possível explorar os benefícios da detecção de MAC para fins de gestão e operação de transporte público eficiente e com grande quantidade de dados.

O uso generalizado de dispositivos móveis com interfaces Wi-Fi é o principal motivo para considerar esquemas de contagem automática baseados na detecção de endereços MAC. No Brasil, por exemplo, existem cerca de 230 milhões de *smartphones* em uso pela população (MEIRELLES, 2019), com aproximadamente 210 milhões de habitantes, isso significa que existe um smartphone por habitante em uso no Brasil; Barbosa (2018) cita que 86% das pessoas usam Wi-Fi fora de casa no país. O predomínio do uso da conexão Wi-Fi se deve principalmente ao acesso gratuito em vários locais e ao menor consumo de energia de Wi-Fi em comparação com o acesso alternativo à Internet 3G / 4G. Portanto, os *smartphones* fornecem um meio potencial para estimativa de localização pessoal, inclusive dentro de um ônibus.

Embora muitas pesquisas estão sendo realizadas utilizando o Wi-Fi, por exemplo, Dunlap *et al.* (2016), Ji *et al.* (2017), Håkegård *et al.* (2018), Junior e Medrano (2018), Jiayu Liu *et al.* (2019), Paolo *et al.* (2019) e Nitti *et al.* (2020), é necessário uma cautela nos resultados apresentados. Isso porque o protocolo de comunicação utilizado para realizar a conexão Wi-Fi não realiza as buscas de redes com um intervalo de tempo regular, o que causa problemas na estimação dos embarques e desembarques e as pesquisas recentes não tratam dessa limitação. Assim, é preciso ter conhecimento do comportamento das buscas de redes realizadas pelos dispositivos móveis.

Através dos dados coletados investiga-se um método estatístico baseado no teorema de Bayes, que permite realizar inferências baseado no conhecimento de um embarque ou desembarque que pode estar relacionado ao evento do dispositivo móvel realizar uma busca de rede.

Dessa forma, essa pesquisa busca através de dados coletados do Wi-Fi de dispo-

sitivos móveis portados por passageiros embarcados em ônibus, investigar a viabilidade estimar os embarques e desembarques de passageiros, e assim obter dados de deslocamento dos mesmos, bem como o carregamentos das linhas sem a necessidade de pesquisas manuais utilizando métodos estatísticos.

1.1 OBJETIVOS

1.1.1 OBJETIVO GERAL

O objetivo deste trabalho é obter estimativas de embarques e desembarques de passageiros, utilizando a sinalização emitidas por dispositivos Wi-Fi portados por passageiros embarcados em linhas de ônibus, sem a necessidade de contagem manual por pesquisadores a bordo ou em pontos ao longo do itinerário.

1.1.2 OBJETIVOS ESPECÍFICOS

- Obter um dispositivo de baixo custo, com a finalidade de coletar informações de sinais Wi-Fi emitidos por dispositivos móveis portados por passageiros que estão embarcados em ônibus;
- Confirmar a rastreabilidade de dispositivos móveis com Wi-Fi atuante a partir do uso de tecnologias consolidadas de detecção;
- Confirmar a possibilidade de associar um dispositivo móvel a uma pessoa;
- Estabelecer os limites de operação do dispositivo de detecção em relação a acurácia da detecção direta de sinais de Wi-Fi na contagem de passageiros a bordo;
- Encontrar os intervalos de sinalizações emitidas por dispositivos móveis com Wi-Fi através do desenvolvimento de uma aplicação para dispositivos móveis que registre as sinalizações;
- Obter um modelo probabilístico para estimar onde os embarques e desembarques ocorreram;
- Obter um modelo de geração de detecção para um cenário onde não é possível realizar experimentos de campo, como no caso de uma pandemia.

1.2 CONTRIBUIÇÃO

A principal contribuição do trabalho é uma análise crítica às técnicas de estimação de embarques e desembarques de passageiros que usam dados coletados de dispositivos móveis Wi-Fi, estabelecendo limites para que essa técnica possa ser aplicada, e assim foi

proposto um método probabilístico de estimação de embarques que permite determinar a exatidão das estimativas realizadas com a detecção de Wi-Fi.

Trabalhos recentes que estão sendo apresentados levam em consideração o uso de dados agregados, enquanto este trabalho busca trazer além dos dados agregados os dados desagregados e as implicações de só utilizar dados agregados.

Essa tese também traz uma metodologia baseada em probabilidade para estimar onde ocorreram os embarques dos passageiros considerando todos os pontos de paradas de uma linha.

Outra contribuição é um olhar cauteloso sobre a estimativa de embarques e desembarques de passageiros somente utilizando a coleta de informações do Wi-Fi de dispositivos móveis embarcados. Em muitos artigos e congressos foram apresentadas informações de que há possibilidade de estimativa de embarques e desembarques com até 85% de certeza, esse dado não é confiável porque novamente os dados apresentados são agregados. O trabalho realizado nesta tese permite que seja realizado a estimativa de embarque ponto a ponto e zona a zona.

Dessa forma esse trabalho contribui para realizar a estimativa de embarques em cada um dos pontos e não somente da linha como um todo, o que auxilia o operador de transporte coletivo nos carregamentos.

Foi desenvolvido um dispositivo capaz de detectar sinais de Wi-Fi de dispositivos móveis que possam auxiliar empresas de transporte coletivo em tomadas de decisões com relação ao carregamento de suas linhas.

Foi desenvolvido uma aplicação para dispositivos móveis com a finalidade de obter a taxa de sinalizações realizadas por dispositivos móveis para serem utilizadas como o comportamento populacional dos dispositivos móveis. Na literatura não há muitas informações sobre o comportamento dessas sinalizações.

1.3 PRODUÇÃO CIENTÍFICA

Ao longo do trabalho foram realizadas duas publicações em formato de artigo. O primeiro artigo *Estimativa de embarque e desembarque de passageiros por detecção de sinal de Wi-Fi de dispositivos móveis* (PARADEDA *et al.*, 2018) foi apresentado no XXXII Congresso de Pesquisa e Ensino em Transportes (ANPET) no ano de 2018, e apresentava os primeiros dados da pesquisa. Esse artigo recebeu o Prêmio ANPET 2018 de Produção Científica, na área de Gestão de Transportes. O segundo artigo intitulado *Bus passenger counts using Wi-Fi signals: some cautionary findings* (PARADEDA *et al.*, 2019), foi publicado na revista *transportes* no ano de 2019. Esse artigo aprofundou a pesquisa e apresentou resultados indicando que é preciso um cuidado na utilização de sinalizações de Wi-Fi para realizar as estimativas de embarques e desembarques de passageiros. Atualmente está sendo desenvolvido um artigo onde os dados coletados nos experimentos são aplicados em modelo probabilístico, estimando os embarques e desembarques.

1.4 ORGANIZAÇÃO DO DOCUMENTO

No Capítulo 2 é realizada a contextualização e metodologia empregada na Tese. No Capítulo 3 são apresentados o dispositivo de detecção e os testes controlados realizados para sua aferição e validação. Os experimentos de campo realizados são apresentados no Capítulo 4. O desenvolvimento da aplicação capaz de registrar buscas de redes realizadas por dispositivos móveis é descrito no Capítulo 5. É apresentado no Capítulo 6 o comportamento dos dispositivos móveis. O método de estimativa de embarques é apresentado no Capítulo 7, enquanto que os resultados obtidos são apresentados no Capítulo 8. Por fim, a conclusão desta tese é apresentada no Capítulo 9.

2 CONTEXTUALIZAÇÃO E METODOLOGIA

Nesse capítulo são abordados a contextualização da pesquisa; a revisão da literatura com as tecnologias disponíveis; os aspectos regulatórios para realizar pesquisas com Wi-Fi no Brasil e o padrão de tecnologia Wi-Fi.

2.1 CONTEXTUALIZAÇÃO

Supondo que haja interesse em realizar pesquisas sobre e-desce para futuramente aplicar na atualização da demanda de um sistema de transporte coletivo, é necessário que essas pesquisas sejam realizadas em cada uma das linhas do sistema. É muito difícil e caro colocar pesquisadores em todos os ônibus para realizar a contagem de sobre-e-desce de passageiros. Através dos dispositivos móveis portados por passageiros, há a possibilidade de estimar os embarques e desembarques através das sinalizações Wi-Fi emitidas pelos aparelhos.

Por exemplo, no cenário da Figura 1 há três usuários do sistema de transporte aguardando a chegada de um ônibus em um ponto de ônibus **A**. A princípio não é possível saber em qual linha eles desejam embarcar, e nem em qual ponto irão desembarcar. Cada usuário está portando um dispositivo capaz de realizar conexões Wi-Fi, com o mesmo ativo. Como não temos certeza do tipo de dispositivo que o usuário está portando, vamos chamá-lo de dispositivo móvel.

O ônibus que está se aproximando do ponto de ônibus **A**, é um ônibus com um dispositivo detector de sinais Wi-Fi embarcado; vamos chamar esse detector de *Network Traffic Monitoring* (NTM) (BAUER, 2004) ou Monitor de Tráfego na Rede (MTR). Quando o ônibus se aproxima do ponto de ônibus **A**, o detector começa a captar os sinais Wi-Fi emitidos pelos dispositivos móveis dos usuários no ponto. Apenas dois usuários realizam o embarque e seguem viagem com o ônibus enquanto um usuário fica esperando no ponto de ônibus **A** até o veículo que faz sua linha chegue.



Figura 1: Exemplo de usuários portando dispositivos móveis, antes durante e depois da chegada de um ônibus equipado com um MTR.

Os dispositivos móveis dos usuários que embarcaram no ônibus continuam emitindo sinais Wi-Fi ao longo da viagem, sinais esses sendo detectados pelo MTR. Como os dispositivos móveis dos usuários foram detectados a primeira vez no momento que o ônibus chegou no ponto de ônibus **A** e continuaram sendo detectados ao longo da viagem,

considera-se que esses dois usuários embarcaram no ponto **A**. Já o usuário que ficou no ponto **A**, e foi detectado pelo MTR quando o ônibus chegou no ponto a primeira vez, é descartado pois seu dispositivo móvel não foi detectado novamente ao longo da viagem.

No caso ilustrado na Figura 2, quando o usuário desembarca e se afasta do ônibus, o MTR para de captar os sinais Wi-Fi emitidos pelo dispositivo móvel portado pelo usuário que desembarcou, mas ainda não é possível saber se o usuário realmente desembarcou. Quando o ônibus parte do ponto de ônibus **B**, seguindo sua viagem, é confirmado que realmente aquele usuário desembarcou no ponto B porque o seu sinal de Wi-Fi não foi mais coletado ao longo da viagem. Já o usuário que estava embarcado continuou sendo detectado ao longo do trajeto até o momento em que desembarcou em um ponto qualquer.

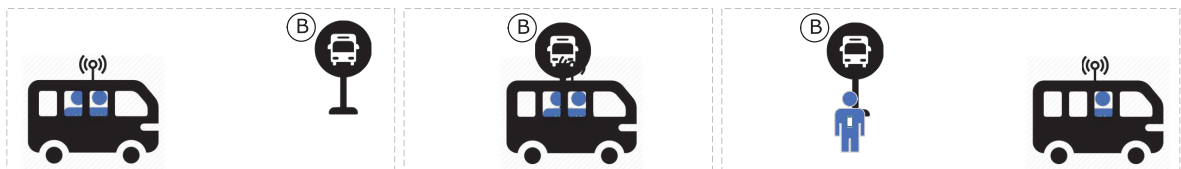


Figura 2: Exemplo de usuário portando dispositivo móvel desembarcando um ponto de ônibus enquanto um passageiro permanece a bordo.

Dado que um número significativo de usuários portam dispositivo móvel com Wi-Fi ativado, e estão sinalizando frequentemente, dessa forma é possível desenvolver uma matriz origem/destino do transporte coletivo por pontos de ônibus, saber onde cada passageiro embarcou ou desembarcou, em qual linha ele estava, duração da sua viagem.

Essa contextualização não leva em consideração os instantes de tempo em que os eventos ocorrem, somente que eles realmente ocorrem. E também não leva em consideração passageiros que já estão embarcados, é somente para contextualizar o comportamento do detector no momento de embarque e desembarque. Ao longo desta tese serão abordados os aspectos que indicam se é possível realizar essa detecção no trajeto de uma linha real.

2.2 REVISÃO DA LITERATURA

A escolha pelo método de coleta de dados se dá baseado na facilidade, velocidade e confiança de obter a informação que se deseja coletar. Quando se trata de dispositivos onde há necessidade da interação do usuário para o acionamento de alguma função específica, como no caso do *smartphone*, essa escolha se torna ainda mais complexa por ter o fator humano envolvido.

Nos *smartphones* existem quatro possíveis formas para coleta de informações dos usuários. A primeira que fornece a posição geográfica aproximada na forma de latitude e longitude do usuário é o GPS, mas essa necessita da interação dos usuários para acionar o seu funcionamento. O consumo de bateria aumenta consideravelmente quando o GPS

é acionado, fazendo com que essa opção seja descartada pelo fato de que os usuários tipicamente não mantêm essa função habilitada nos seus *smartphones*.

A segunda é através do próprio sinal da operadora de telefonia que se mantém sempre em funcionamento sem necessitar da interação do usuário. Diversos autores (ÇOLAK *et al.*, 2015; JIANG *et al.*, 2017; BEKHOR *et al.*, 2013; ZHANG *et al.*, 2010; LIND *et al.*, 2017; ALEXANDER *et al.*, 2015), conseguiram com sucesso determinar trajetórias realizadas pelos usuários e modelos capazes de identificar origem e destino através informação dos dados coletados das operadoras de telefonia. Esses dados consistem de horário que um celular foi registrado em uma antena, horário de ligação, horário que recebeu uma mensagem, horário que acessou a internet através de 3G ou 4G. Porém esses dados só estão disponíveis em países que armazenam esse tipo de informação. No Brasil esta informação não é armazenada, exceto em alguns casos como durante as olimpíadas do Rio de Janeiro em 2016. Xu e González (2017) apresentaram uma estimativa de demanda de viagens baseado em dados de celulares para a cidade do Rio de Janeiro durante os jogos olímpicos com os dados de ligações por celulares. Note, porém, que ligações por celulares já não ocorrem com tanta frequência. De acordo com Barbosa (2019), há indicação de que ocorreu um crescimento de 73% nas chamadas de voz ou vídeo pela internet. As operadoras de telefonia no Brasil são bem restritivas e as informações que coletam, como no caso das olimpíadas, não são fornecidas gratuitamente, somente são vendidas por preços muito elevados. Isso leva a crer que uma estimativa de demanda baseada em dados Wi-Fi ou Bluetooth sejam mais interessantes.

A terceira e quarta forma são Bluetooth e Wi-Fi, apesar de necessitarem do acionamento do usuário no seu *smartphone*, costumam ser utilizados com uma frequência maior. A diferença entre a escolha da tecnologia Wi-Fi se comparada com a Bluetooth se dá pela velocidade com que se coleta os dados, o que para aplicações em tempo real é extremamente importante. Abedi *et al.* (2013) apresentam a coleta de dados através de dispositivo Bluetooth e Wi-Fi focando no aumento do raio de detecção com a troca de antena. Além disso, realizaram testes para determinar a velocidade de detecção das duas tecnologias. Através dos testes realizados determinaram que através da tecnologia Wi-Fi é possível detectar um usuário oito segundos mais rápido que com Bluetooth. Embora a diferença de tempo pareça ser muito pequena, quando se trata de detecções sucessivas de um mesmo dispositivo, essa diferença pode ser impactante.

Existem alguns trabalhos na literatura que lidam com detecções de sinais de Wi-Fi, mas os estudos da utilização desse método focando em pesquisas de embarques e desembarques em ônibus são recentes. Um sistema de captura de sinais Wi-Fi foi utilizado por El-Tawab *et al.* (2016) com o objetivo de determinar o tempo de espera dos passageiros em dois pontos de ônibus utilizando dispositivos previamente conhecidos. Os resultados indicaram que é possível estimar o tempo de espera, no entanto não foi apresentado como o dispositivo de coleta foi testado. O estudo discute a possibilidade de haver falsos positivos,

resultado esse que causa grande impacto levando a indicações de que passageiros foram contados e não deveriam ter sido.

Dunlap *et al.* (2016) combinou Bluetooth, Wi-Fi e GPS em um período longo de coleta de dados. Foram utilizadas técnicas básicas de filtragem para processar a coleta de dados. No entanto os resultados apresentados estão distantes, já que não é apresentado um modelo de calibração do sistema de detecção de forma que possa ser relacionado com dados reais. Como os próprios autores concluem, no seu método há possibilidade de detectar dispositivos Wi-Fi e utilizá-lo como pesquisas em sistemas de transporte público, mas não indicam se os níveis de precisão são adequados para tal.

Através de métodos estatísticos Håkegård *et al.* (2018) obtiveram resultados satisfatórios na estimação de embarques e desembarques de passageiros. Foram utilizados três tipos de algoritmos baseado no tempo de detecção e na potência do sinal. Parâmetros das funções da distribuição de probabilidade são estimados pela média da contagem automática de passageiras desenvolvida pelos autores (*Automatic Passenger Counting (APC)*) com dados que vieram da empresa de ônibus que opera no local dos estudos. A conclusão é que para um conjunto relativamente grande de dados com uma margem grande de observação (semanas a meses), o algoritmo é capaz de estimar com precisão uma matriz Origem/Destino (OD). No entanto, o estudo não aborda a questão de pesquisas em espaços de tempo menores, nem a capacidade da própria detecção de dispositivos Wi-Fi. Isso por que pode haver dificuldade em detectar dispositivos Wi-Fi mesmo que estejam com o Wi-Fi ligado (resultando em falsos negativos) e a detecção de dispositivos que estão atualmente fora do ônibus (resultando em falsos positivos), isso pode ocorrer em vias muito congestionadas, ou em semáforos com tempo de vermelho muito longo. Ao longo desta tese, é discutido que esses pontos estão relacionados ao intervalo das detecções, e se mostrou uma abordagem necessária para seguir os estudos com cautela.

Nunes (2018) apresentou um método de classificação de carregamento de ônibus em seis categorias (partindo de vazio até lotado) utilizando detecções por Wi-Fi. A classificação utilizando $\chi^2 - test$ e um método de análise geográfica baseado no itinerário foram utilizados para o processamento de dados. Os resultados indicam uma capacidade de estimar categoricamente na maioria dos casos. O autor apresenta alguns pontos que apresentaram problemas quando a ocupação subitamente muda, o que ocorre frequentemente nos pontos de parada. Uma categorização mais precisa e uma estimação dos pares de origem e destino não foram tentados.

Utilizando técnicas de aprendizado de máquina, Junior e Medrano (2018) estimaram embarques e desembarques de passageiros ao longo da rota de uma linha de ônibus. O primeiro experimento de campo, apenas com dispositivos conhecidos gerou os dados do grupo de controle. Posteriormente um experimento de campo com dispositivos não conhecidos foi realizado. Com os dados coletados, *Support Vector Machine (SVM)* foi usado para classificar dispositivos móveis de acordo com as classes de embarque e desembarque.

Embora os autores indiquem que é possível estimar com uma certeza os embarques com 83% e os desembarques com 88,7%, os resultados refletem o processamento dos dados de forma agregada de uma viagem inteira. Quando os dados são desagregados por pontos de ônibus, os resultados no artigo revelam uma subestimação e uma superestimação em cada ponto que quando agregados tendem a compensar ao longo da rota. Assim, os dados agregados parecem ter mais chance de gerar dúvidas quanto à precisão de estimar os embarques/desembarques.

Com um sistema de detecção embarcado dentro do ônibus Ji *et al.* (2017) apresentaram um método com o objetivo de estimar o carregamento dos veículos ao longo da rota de ônibus. Através de um Modelo Bayesiano Hierárquico, a proposta captura informações combinando dados observados de diferentes viagens de ônibus. Quando é realizada a estimativa dos fluxos de OD de uma viagem de ônibus, os fluxos de outras viagens também são levados em conta. Os resultados dos autores indicam que há viabilidade de combinar detectores com os dados de sinais Wi-Fi para obter um resultado satisfatório e detalhado da demanda de passageiros de uma linha. No entanto os autores indicam que mesmo com todo o cuidado na filtragem dos dados ainda podem aparecer falsos positivos, entre outros problemas, como passageiros que transportam mais de um *smartphone* ou dispositivos que não emitem sinais Wi-Fi.

Em Ryu *et al.* (2020) foi proposto um método de um sistema de coleta de dados Wi-Fi de baixo custo. O estudo mostrou a possibilidade de estimar a matriz Origem/Destino de passageiros de transporte coletivo e o tempo de espera em pontos validado através de sistema de vídeo. A potência do sinal é um método utilizado para melhorar a acurácia do dispositivo e foi utilizado o algoritmo de janela deslizante com a finalidade de mitigar a aleatoriedade dos sinais Wi-Fi. O estudo foi realizado em quatro pontos de ônibus ao longo de um corredor em Charlottesville cidade do estado da Virgínia nos Estados Unidos, e indicou que foi possível estimar 91% dos passageiros dos ônibus e que o erro no tempo de espera foi de 7 segundos. No entanto a aleatoriedade nas buscas de redes realizadas por dispositivos móveis pode levar a passageiros que deveriam ser contados e não foram. O método proposto considera uma identificação dos passageiros além do sinal Wi-Fi com um sistema de vídeo, os autores indicam um custo de US\$ 200,00. Para os padrões brasileiros implantar esse sistema em todos os ônibus e todos os pontos de embarque e desembarque o valor se torna muito elevado se comparado com o sistema proposto nessa tese, com um custo de aproximado de US\$ 70,00 tendo somente que ser instalado em veículos e não em pontos, e ainda não seria necessário a instalação em toda a frota, caso a análise seja feita por linhas, seria necessário a implantação somente da frota que opera determinada linha.

Em resumo, apesar de haver formas de detectar passageiros através de seus *smartphones* utilizando o sinal da rede de celular, a tecnologia atual empregada pelas operadoras não gravando as informações dos usuários não permitem que esse campo seja explorado. Parece evidente que esse seria o cenário ideal, já que basta o passageiro estar portando

um *smartphone*. Por outro lado, trabalhos na literatura sugerem que a implantação da detecção Wi-Fi como um meio para obter dados de transporte de passageiros é alcançada com sucesso razoável. Porém existem muitas variáveis que geram dúvidas com relação aos resultados obtidos, seja na forma de utilizar dados agregados, os falsos positivos e falsos negativos, entre outros. Contrariando esses trabalhos, nossos resultados experimentais fornecem conclusões cautelosas sobre o uso da tecnologia em pesquisas para sistema de transporte. Nessa tese são apresentados experimentos controlados e experimentos de campo, além de um experimento mais detalhado sobre o comportamento dos dispositivos Wi-Fi, sem semelhantes na literatura e um modelo proposto para estimar os embarques e desembarques.

2.3 ASPECTOS REGULATÓRIOS E TECNOLÓGICOS DA PESQUISA

Esta seção apresenta a lei que regulamenta o uso da internet no Brasil e como pode afetar a pesquisa, e a tecnologia utilizada no Wi-Fi com os subsídios necessários para que seja realizada a pesquisa.

2.3.1 LEI, DIRETRIZES E NORMAS PARA PESQUISAS COM DISPOSITIVOS WI-FI NO BRASIL

No Brasil a legislação está baseada na lei de proteção de dados, onde exige consentimento explícito para coleta e uso dos dados, então mesmo para pesquisas na área acadêmica provavelmente vai ser necessário passar por algum comitê de ética para realizar as pesquisas, a lei 13709/18 de 14 de agosto de 2018 entrou em vigor no ano de 2020, e o Art.1º diz:

Esta Lei dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.

E embora o Art.4º, Inc. II, onde essa lei não se aplica ao tratamento de dados pessoais realizado para fins exclusivamente acadêmicos, desde que seja aplicada a está hipótese os Art.7º e Art.11º dessa Lei. E os Art.7º e Art.11º são explícitos que há necessidade do consentimento do titular dos dados.

Com base nessas informações, seria quase impossível de realizar essa pesquisa com dados Wi-Fi no Brasil no ano de 2020. Quando esse trabalho estava sendo realizado, ele tinha como base a a lei 12965/14 de 23 de abril de 2014 (CARVALHO, 2014). Lei essa que estabelece os princípios, garantias, direitos e deveres para o uso da internet no Brasil. Ela é mais conhecida como o Marco Civil da internet e no seu texto nhá algo que impeça a

realização de coletas Wi-Fi. Embora não haja no texto da lei algo que impeça a realização de coletas Wi-Fi, podemos nos apoiar no Art.6º que diz:

Na interpretação desta Lei serão levados em conta, além dos fundamentos, princípios e objetivos previstos, a natureza da internet, seus usos e costumes particulares e sua importância para a promoção do desenvolvimento humano, econômico, social e cultural.

Entende-se que este trabalho esteja contemplado no contexto social. Ainda mais, o texto da Lei foca basicamente em conexões, onde a responsabilidade pelos dados de tráfego é da operadora. Não há informações referentes a captura dos dados que estão ao ar livre. Não havendo uma legislação no país referente a essa captura de dados a pesquisa se baseia na legalidade observada nos Estados Unidos.

Muitas leis nos Estados Unidos proíbem ou restringem o monitoramento de redes e o compartilhamento de registros das atividades da rede. As leis foram criadas para proteger a privacidade online. As leis cobrem tanto o setor privado como o setor público, no entanto os detalhes das leis variam de acordo com quem está realizando o monitoramento. Como não há um consenso geral, Garfinkel e McCarrin (2014) aconselha que coletar dados deva ser usado com prudência. Nenhuma lei americana específica se a coleta pode ou não ser realizada para fins acadêmicos. Embora não haja informações referente a isso, aconselha-se que todas as informações sensíveis sejam descaracterizadas para que não seja possível a identificação do usuário.

Embora, como já citado anteriormente, não haja nada no Brasil que proíba a captura de dados ao ar livre, é necessário se preocupar com a integridade dos dados dos usuários que possam ser obtidos. Neste trabalho, as informações sobre a navegação online dos usuários não são consideradas, considera-se assim que a integridade dos usuários está mantida. Além disso, somente um pesquisador tem acesso aos dados que foram coletados, e os dados não estão disponíveis para visualização online, somente off-line em uma partição do disco rígido com criptografia para acesso.

Conclui-se que pela sensibilidade dos dados que possam ser obtidos, torna-se necessário uma lei específica sobre esse tipo de informação e quem poderia coletar e analisar esses dados.

2.3.2 MARKET SHARE DE DISPOSITIVOS

Segundo Kotler *et al.* (2017), *Market Share* é a participação de uma empresa e seus concorrentes no mercado. Ela representa o quanto cada empresa detém do mercado. Para este trabalho, o conhecimento do *Market Share* de dispositivos móveis se torna necessário para entender o comportamento dos consumidores através da representatividade de uma marca de *smartphone* em um determinado país, e até do sistema operacional mais utilizado.

Tabela 1: Países com o maior número de usuários de *smartphones*.

Rank	País	População Total	Utilização do <i>Smartphone</i>	Usuários de <i>Smartphone</i>
1	China	1 415 046	55,3%	782 848
2	India	1 354 052	27,3%	374 893
3	Estados Unidos	326 767	77,0%	251 688
4	Federação Russa	143 965	63,8%	91 865
5	Brasil	210 868	41,3%	87 172
6	Indonésia	266 795	27,4%	73 155
7	Japão	127 185	55,3%	70 327
8	Alemanha	82 293	78,8%	64 830
9	México	130 759	45,6%	59 597
10	Reino Unido	66 574	82,8%	54 713

Fonte: newzoo (2018).

O conhecimento sobre os dispositivos mais utilizados, permitem ações futuras e entendimentos sobre as características dos dispositivos quando se está realizando uma pesquisa. Na área da telefonia são dois sistemas operacionais dominantes, Android da Google LLC e iOS da Apple inc. Existe uma gama de aparelhos que utilizam esses sistemas operacionais.

Android é um sistema operacional voltado para *smartphones*, *Tablets* e uma série de outros dispositivos desde relógios, centrais multimídias para veículos até *SmartTVs*. Lançado em 2003, atualmente é o sistema operacional voltado para *smartphones* mais popular do mundo. É um projeto de código livre, liderado pela Google que, utiliza o projeto como base para criar suas próprias versões de Android que eventualmente serão utilizadas pelas empresas de dispositivo móveis.

iOS é um sistema operacional voltado para dispositivos exclusivamente da empresa Apple Inc. O sistema operacional é encontrado em *smartphones* iPhone, *Tablets* iPads e no dispositivo reproduzidor de músicas e vídeos iPod Touch.

Primeiro será comparada a utilização de *smartphones* nos 10 países com maior uso de *smartphones* em 2018. Um usuário de *smartphone* é qualquer um que utilizou um *smartphone* pelo menos uma vez ao mês. Esses números vêm do Global Mobile Market Report (KORTEKAAS, 2018) e é baseado em um modelo que leva em conta a progressão econômica, demografia, população online e desigualdade. A Tabela 1 ilustra esses 10 países.

É preciso levar em consideração que os dados obtidos pelo newzoo (2018) levam em consideração acessos a sítios eletrônicos específicos, e por isso os números de dispositivos registrados no Brasil é inferior ao que é registrado por Meirelles (2019) que é uma pesquisa mais abrangente e leva em consideração a posse do dispositivo móvel. Com base nessa

tabela de utilização de *smartphones* será analisado o *Market Share* baseado em sistema operacional dos 10 países listados na Tabela 1.

Os dados para análise de *Market Share* vêm do StatCounter (2020), serviço de análise *web*. O *Statcounter* possui um código de rastreamento em mais de 2 milhões de sítios eletrônicos globais. Os dados coletados abrangem várias atividades e a posição geográfica dos dispositivos. Entre as atividades, estão sendo coletadas as informações do navegador utilizado para acessar o sítio eletrônico e o sistema operacional do dispositivo. Focando basicamente em *smartphones* a Tabela 2 ilustra como se dá o *Market Share* mundial e o *Market Share* de cada um dos 10 países com maior uso de *smartphone* baseado no sistema operacional dos dispositivos.

Tabela 2: Utilização de sistema operacional por países com maior uso de *smartphone*.

País	Android (%)	IOS (%)
Mundo	74,9	22,2
China	73,2	25,1
India	90,3	2,7
Estados Unidos	42,8	56,9
Federação Russa	70,4	28,0
Brasil	84,1	14,4
Indonésia	92,2	5,9
Japão	23,2	76,5
Alemanha	67,7	31,2
México	80,9	18,8
Reino Unido	47,1	52,0

Fonte: StatCounter (2020)

Como já citado anteriormente, pela Tabela 2 podemos comprovar que o sistema operacional Android é o mais encontrado atualmente no mercado dos *smartphones*. No Brasil sua representação é de 84% dos *smartphones*. Dessa forma chega-se à conclusão de que uma pesquisa baseada na utilização de *smartphones* no Brasil, há muitas chances de os *smartphones* que estão sendo pesquisados possuírem sistema operacional Android. Para a análise dos dados esse fato é muito importante, ainda mais por que há uma limitação com a real informação do Wi-Fi dos dispositivos que podem ser mascaradas, caso não estejam além do Wi-Fi ligado o GPS também precisa estar ligado. Dispositivos Android não necessitam estar com o GPS do dispositivo ativado para enviar as informações reais do aparelho enquanto não está conectado. Isso difere dos dispositivos com iOS que se não estiverem com o GPS ativado sofrem um mascaramento das informações, fazendo com que seja necessário outros procedimentos para obter as informações reais (VANHOEF *et al.*, 2016).

Expandindo um pouco mais esse mercado, é possível analisar os 10 fabricantes dos dispositivos mais utilizados em 2018. E nesse caso vamos analisar somente para o

Tabela 3: Market Share de Fabricantes no Brasil.

Fabricante	Média de 12 meses (2018) (%)
Samsung	41,4
Motorola	22,8
Apple	12,2
LG	11,2
Asus	2,2
Lenovo	1,9
Nokia	1,3
Alcatel	1,3
Xiaomi	0,8
Sony Ericsson	0,7
Outros	4,2

Fonte: StatCounter (2020)

Brasil. Com base na Tabela 3, é possível indicar que a maior disseminação de *smartphones* no Brasil está entre Samsung e Motorola que utilizam nos seus dispositivos o sistema operacional Android, com 64,2% do mercado, se combinarmos as outras fabricantes que utilizam Android essa fatia do mercado passa para 80% contra os 12,2% da iOS, isso significa que há 80% de chance de que as informações coletadas não serão mascaradas e o dado obtido será informação real do dispositivo.

2.3.3 PADRÃO IEEE 802.11

Sabe-se que foram estimados pela Wi-Fi.org (SUTHERS, 2018) que em 2019 existiriam mais de 10 bilhões de dispositivos com a capacidade de se conectar em um roteador doméstico. Isso inclui Smart TV, Video Games, Tablets, *Smartphones*, eletrodomésticos em geral, entre outros. Grande parte desses dispositivos domésticos utilizam como conexão o protocolo IEEE 802.11. Não vamos entrar a fundo na parte do comportamento e das características das redes sem fio. Portanto nessa seção o foco se dará na estrutura do quadro 802.11 e quais partes foram importantes para o desenvolvimento deste trabalho, além de uma análise rápida sobre o protocolo em si.

A arquitetura do protocolo IEEE 802.11, conta com um conjunto básico de serviços, *Basic Service Set* (BSS), que é o bloco fundamental da arquitetura 802.11. Um BSS possui uma ou mais estações sem fio e uma estação-base central, conhecida como ponto de acesso, ou *Access Point* (AP), na terminologia 802.11. Podemos imaginar uma rede residencial, onde normalmente há somente um roteador fazendo o serviço de conexão com a internet e ainda servindo como ponto de acesso sem fio para os dispositivos (estações) se conectarem. Cada dispositivo (estação) sem fio 802.11 possui um endereço *Media Access Control* (MAC) de 6 bytes armazenado no *firmware* do hardware. Esse endereço é administrado pelo

Institute of Electrical and Electronic Engineers (IEEE) e a autoridade que regulamenta os registros de endereços MAC's é a *IEEE Registration Authority* (IEEE RA), e teoricamente ele é globalmente exclusivo. Podemos fazer uma analogia do MAC com o Cadastro de Pessoa Física (CPF) no Brasil, em que cada pessoa tem um CPF que é único e cada dispositivo tem um MAC que é único. Embora um dispositivo possa ter mais de um MAC desde que tenha mais de uma placa de rede, para cada placa o MAC é único.

São definidos dois tipos de identificador um de 24-bits que define a identificação única da organização OUI e um de 24-bits da identificação da companhia *Company ID* (CID). Um OUI ou CID pode ser usado para identificar uma companhia, organização, fabricante, fornecedor. A IEEE RA também reconhece e atribui um OUI-36 de 36 bits como um identificador globalmente exclusivo de uma organização. Os primeiros 24 bits do OUI-36 atribuído não duplica nenhuma atribuição OUI ou CID.

O *Extended Unique Identifier* (EUI) é um identificador de 48 ou 67-bits único. Com algumas exceções, cada EUI é globalmente único e aplicado diretamente no *hardware* ou em algum outro objeto que requer uma identificação única. Eles são utilizados globalmente como endereços de rede únicos, mais conhecidos como endereço MAC (RA, 2018).

A Tabela 4 ilustra a representação do EUI-48, onde basicamente os primeiros três identificadores do octeto se referem à identificação do fabricante ou quem adquiriu aquela faixa de endereços e é atribuído pelo órgão regulador, enquanto os três últimos é definido pelo próprio fabricante, formando assim o endereço MAC do dispositivo. Nesse caso o endereço MAC da Tabela 4, AC-DE-48-23-45-67 é alocado a um dispositivo único no mundo.

Tabela 4: Estrutura do EUI-48.

Identificador Octeto	0	1	2	3	4	5
Componente MAC	24-bit OUI			24-bit extensão		
Exemplo Valor (hex)	AC	DE	48	23	45	67
Exemplo Valor (binário)	1010 1100	1101 1110	0100 1000	0010 0011	0100 0101	0110 0111

Fonte: RA (2018). Traduzido pelo autor.

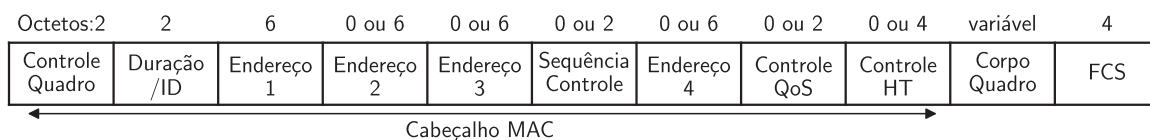
Sabe-se que hoje em dia há formas de randomizar esse MAC, mas essa randomização é falha, e de acordo com Vanhoef *et al.* (2016), Matte *et al.* (2016), Martin *et al.* (2017) e Vasilevski *et al.* (2019) isso pode ser contornado.

Cada dispositivo móvel para enviar e receber dados precisa se associar com um AP para poder receber e enviar dados da camada de rede (KUROSE; ROSS, 2006). Mas sabe-se que mesmo sem uma associação há trocas de informações (KUMKAR *et al.*, 2012; SEPEHRDAD *et al.*, 2013; WALIULLAH *et al.*, 2015; ACER; WALTARI, 2017; LÓPEZ *et al.*, 2018); isso será discutido mais para frente. Um dispositivo AP é configurado com um Identificador de Conjunto de Serviços, ou *Service Set Identifier* (SSID), que é nada

mais que o nome da rede que está disponível para conectar. Além disso é preciso designar um canal para conexão. O padrão 802.11b opera na faixa de frequência de 2,4 GHz a 2,485 GHz e dentro dessa faixa de 85 MHz são definidos 11 canais com sobreposição parcial. Não há sobreposição de canais se eles estiverem separados por quatro ou mais canais. Particularmente os canais 1, 6 e 11 são os únicos três canais que não são sobrepostos entre si. Com o entendimento de que para realizar a conexão é preciso se conectar a um roteador em um canal determinado pelo mesmo. Supondo que você esteja com seu *smartphone* e deseje se conectar na internet em um grande centro onde há muitos AP's disponíveis para a conexão, o seu dispositivo começa a executar funções internas para estabelecer essa conexão. Quando o seu *smartphone* entra no raio desses AP's, há envio de quadros de sinalização que incluem o SSID e o endereço MAC do AP e o canal disponível para realizar a conexão. O seu dispositivo é projetado para realizar a varredura dos 11 canais em busca de sinalização de qualquer AP que possa estar por perto (alguns transmitindo no mesmo canal). Escolhendo o AP através da sinalização emitida, o *smartphone* pode se associar enviando uma solicitação de associação, sendo respondido através de uma resposta de associação.

Este trabalho considera as informações do protocolo de comunicação até o momento da associação, não considerando as informações quando a associação é realizada. Desta forma-se preserva-se a privacidade ao não serem obtidos os dados de navegação dos usuários. As outras etapas do protocolo de comunicação após o momento da associação podem ser encontradas em (JAMES; KEITH, 2005). O interesse nesse momento é analisar os quadros de sinalização e os quadros de solicitação de associação.

A Figura 3 ilustra o formato geral do quadro, ou *Frame* do protocolo 802.11. O quadro é um conjunto de campos de ordem fixa em todos os quadros. Os primeiros três campos do quadro (Controle de Quadro, Duração e Endereço 1) e o último campo (FCS), constituem o mínimo formato de quadro presente em todos os quadros. Todos os campos possuem subcampos.



Fonte: IEEE (2007). Traduzido pelo autor.

Figura 3: Formato do quadro MAC.

Olhando um pouco mais a fundo no quadro, vamos analisar os subcampos do Controle de Quadro.

Controle de quadro

Os primeiros subcampos do campo Controle de Quadro são Versão do protocolo, Tipo e subtipo. O restante dos subcampos do Quadro de Controle, dependem das configu-

rações do Tipo e Subtipo. A Figura 4 ilustra os subcampos do Controle de Quadro. Para o trabalho foram investigados mais a fundo os subcampos de Tipo e Subtipo por conterem as informações referentes a requisições e respostas dos dispositivos. Para os outros campos, consultar James e Keith (2005).

B0	B1	B2	B3	B4	B7	B8	B9	B10	B11	B12	B13	B14	B15
Versão Protocolo	2	2	4	1	1	1	1	1	1	1	1	1	1
	Tipo	Subtipo	Para DS	De DS	Fragmento	Repetir	Gerenciamento Energia	Dados	Quadro Protegido	+HTC/Ordem			
Bits:	2	2	4	1	1	1	1	1	1	1	1	1	1

Fonte: IEEE (2007). Traduzido pelo autor.

Figura 4: Controle de quadro.

Subcampos Tipo e Subtipo

Os subcampos Tipo e Subtipo juntos são responsáveis por identificar a função de um quadro. O subcampo tipo possui um comprimento de 2 bits e o Subtipo um comprimento de 4 bits. Existem três tipos de quadros, são eles: controle, dados e gerenciamento. E cada um dos Tipos possuem seus subtipos. A Tabela 5 ilustra as combinações possíveis de tipo e subtipo.

Tabela 5: Tabela com os campos de tipos e subtipos do Controle de Quadro.

Tipo B3 B2	Tipo Descrição	Subtipo B7 B6 B5 B4	Subtipo descrição
00	Gerenciamento	0000	Requisição de Associação
00	Gerenciamento	0001	Resposta de Associação
00	Gerenciamento	0010	Requisição de Reassociação
00	Gerenciamento	0011	Resposta de Reassociação
00	Gerenciamento	0100	Requisição de sondagem
00	Gerenciamento	0101	Resposta de sondagem
00	Gerenciamento	0110	Anúncio de temporização
00	Gerenciamento	0111	Reservado
00	Gerenciamento	1000	Beacon
00	Gerenciamento	1001	ATM
00	Gerenciamento	1010	Dessassociação
00	Gerenciamento	1011	Autenticação
00	Gerenciamento	1100	Desautenticação
00	Gerenciamento	1101	Ação
00	Gerenciamento	1110	Ação No Ack
00	Gerenciamento	1111	Reservado

Fonte: IEEE (2007). Adaptado pelo autor.

O interesse nessa tabela é buscar campos que possam fornecer qualquer informação trocada entre um AP e um dispositivo que queira se conectar a ele. Para o caso deste trabalho, analisamos a composição dos quadros de Requisição de Sondagem.

Requisição de sondagem, ou *Probe Request*

O corpo do quadro Requisição de sondagem possuem informações importantes com relação a troca de pacotes entre o AP e o dispositivo que queira se conectar. Algumas das informações mais importantes para este trabalho são:

- *Timestamp* - É o campo que representa o tempo da função de sincronização de uma fonte de um quadro. Podemos a grosso modo dizer que é o momento em que a informação foi recebida e sincronizada.
- *Beacon Interval* - Para entender o campo Beacon Interval, primeiro é preciso compreender o significado do *Beacon*. *Beacon* é um campo de gerenciamento que contém as informações relativas à rede, ou seja, ele serve para anunciar a presença de uma rede Wi-Fi no local. Dessa forma o *Beacon Interval*, é o intervalo com que os *Beacons* são transmitidos periodicamente.
- SSID - É o campo que identifica o nome da rede a qual foi trocada a informação. Normalmente o nome dessa rede é em uma linguagem natural.

Radiotap

O *Radiotap* não é uma parte da estrutura do quadro do protocolo 802.11. Ele é um *container* para metadados mais flexível para gerar as características dos quadros. No *Radiotap* é possível receber a importante informação de potência do sinal. A potência do sinal é relativa ao dispositivo de captura do pacote, e não do pacote em si. Por exemplo, se foi recebido um pacote onde consta a potência do sinal em -40 dB, significa que o dispositivo de captura captou esse pacote a -40 dB, mas não significa que o pacote foi enviado com uma potência de sinal de -40 dB. Essa questão é abordada na Seção 3.3.2

2.3.4 CANAIS E FREQUÊNCIAS

A rádio frequência é vital para a infraestrutura da comunicação *wireless* (TSE; VISWANATH, 2005). Os equipamentos que pretendem realizar uma comunicação *wireless* utilizando o protocolo 802.11 precisam respeitar as frequências definidas no padrão. Existem faixas de frequências que vão desde 900 MHz até 60 GHz (IEEE, 2007). Para o caso da comunicação *wireless* de dispositivos móveis atualmente considera-se frequências de 2.4 GHz. Cada faixa de frequência é dividida em canais. Nas frequências de 2.4 GHz são designados 14 canais em toda sua faixa, espaçados em 5 MHz um do outro com exceção para 12 MHz antes do canal 14. Alguns canais como exemplo o canal 14 no Brasil não são utilizados e são proibidos pelo órgão regulador Anatel (2017). A Figura 5 ilustra a representação dos canais Wi-Fi na frequências de 2.4 GHz, isso significa que no Brasil qualquer dispositivo como roteadores ou pontos de acesso que está fornecendo um sinal para conexão, pode optar por operar em alguma das 13 frequências permitidas. E esse tipo de configuração pode ser realizada manualmente no próprio dispositivo, ou deixar que o próprio dispositivo escolha o canal de forma automática. Por outro lado, dispositivos

móveis quando tentam se conectar a um roteador ou ponto de acesso, realizam buscas de rede por esses canais. Nesse momento está ocorrendo troca de informações entre os dispositivos através do protocolo 802.11, envolvendo as informações da Tabela 5.

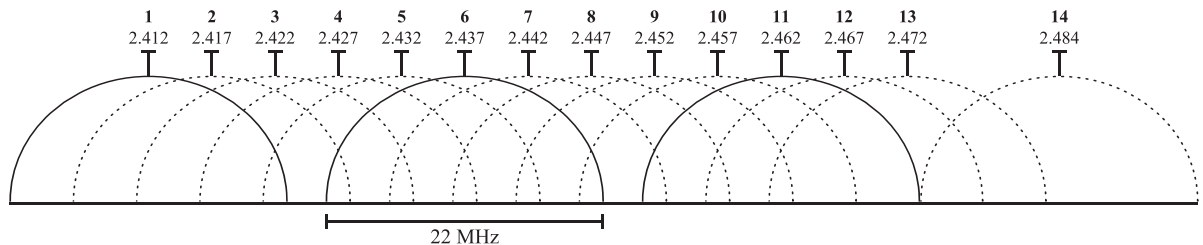


Figura 5: Representação gráfica dos canais Wi-Fi na frequência de 2.4 GHz.

2.3.5 FABRICANTES

Os endereços MAC são atribuídos às placas de rede dos dispositivos pelos fabricantes. Através desse endereço é possível descobrir quem foi o fabricante do dispositivo. Os primeiros três bytes do endereço MAC identificam o fabricante. Utilizando os fabricantes da Tabela 3 como exemplo, foi obtido alguns registros de identificação únicos, ou OUI, cada fabricante realiza a compra de inúmeros registros, isso por que não seria possível registrar todos os dispositivos produzidos somente com esses três bytes da tabela. A Tabela 6 ilustra alguns desses registros. Para verificar os fabricantes, basta copiar os valores e acessar Vendors (2020).

Tabela 6: Fabricantes com registros de OUI.

Fabricante	OUI
Samsung	00:23:C2
Motorola	00:E0:0C
Apple	F0:76:6F
LG	2C:54:CF
Asus	00:04:0F
Lenovo	74:04:2B
Nokia	A4:81:EE
Alcatel	00:C0:BE
Xiaomi	F8:A4:5F
Sony Ericsson	30:39:26

Fonte: Vendors (2020). Adaptado pelo autor.

2.4 METODOLOGIA

A pesquisa busca formas de estimar os embarques e desembarques de passageiros usando novas técnicas de medição para obter essa informação. Dessa forma buscou-se tecnologias passivas e de baixo custo que pudessem fornecer informações sobre os passageiros do sistema de transporte coletivo. Após uma revisão da literatura, algumas opções para obter as informações dos usuários foram elencadas. Utilizar o sinal da operadora de telefonia celular esbarrou na tecnologia adotada pelas operadoras do Brasil que não registram as informações sobre os dados. Já a utilização do Bluetooth precisaria estar sempre ativa no dispositivo móvel do passageiro, o que pode não ocorrer. Já a tecnologia Wi-Fi mostrou indícios de que seria uma opção adequada para coleta de dados de forma passiva (KUMKAR *et al.*, 2012). Assim, o Wi-Fi foi a tecnologia escolhida para esta pesquisa.

Antes de avançar na pesquisa, foram realizados testes com computadores portáteis, com capacidade de detectar sinais Wi-Fi de dispositivos móveis, assim verificando a viabilidade de realizar a detecção de passageiros através dos sinais emitidos pelos seus dispositivos móveis, e após alguns testes foi confirmada que essa possibilidade era real. Os computadores portáteis são dispositivos que permitem uma certa mobilidade, mas com um custo muito elevado, e embora permitam uma certa mobilidade eles ocupam um espaço considerável. Como um dos objetivos da pesquisa é utilizar um dispositivo de baixo custo para a detecção, o computador portátil não preenche esse requisito. Dessa forma definiu-se o projeto de um dispositivo de baixo custo.

Através de pesquisas realizadas por Muniz e Lakhani (2015), Casciati *et al.* (2015), Herrera-Quintero *et al.* (2016) e Yang Liu *et al.* (2016), foi verificado que existem dispositivos de baixo custo e que podem ser utilizados como detectores de sinais Wi-Fi, mas necessitam que sejam feitas modificações nos mesmos para este fim. Assim, foi utilizado um computador Raspberry Pi, e realizadas as devidas modificações para que seja capaz de coletar informações de dispositivos móveis através do sinal Wi-Fi. O fato desse dispositivo ser pequeno permite que seja instalado em pontos de ônibus ou no próprio veículo.

Os testes realizados com o detector de sinal Wi-Fi, consideram os indicadores de outros dispositivos detectores encontrados na literatura. Os testes consistem em verificar se o detector realmente está detectando sinais Wi-Fi de dispositivos móveis. Diferentemente de outros trabalhos, os primeiros testes foram realizados em um campo de teste onde todos os dispositivos móveis que serão detectados são conhecidos. Assim foi verificado se o detector de sinal Wi-Fi conseguia detectar os sinais Wi-Fi emitidos pelos dispositivos móveis conhecidos. Dessa forma o dispositivo foi testado e validado.

Após os testes o detector de sinais Wi-Fi foi testado em experimentos de campo dessa vez em um cenário real, com dispositivos móveis não conhecidos, com o objetivo de obter dados do funcionamento do dispositivo desenvolvido em laboratório. O primeiro experimento foi realizado em um local onde todos os dispositivos móveis a serem detectados

eram conhecidos, enquanto o segundo foi realizado a bordo de um ônibus e contava como uma contagem manual de passageiros. Através do primeiro experimento foram obtidos os limites da precisão com que o dispositivo consegue coletar os dados, bem como se os instantes que os dados são coletados são coerentes. Já o segundo experimento forneceu uma base de dados com os dados coletados, e foi realizada uma primeira estimativa dos embarques de passageiros. Os resultados obtidos no segundo experimento foram muito similares aos da literatura como em Ji *et al.* (2017), Håkegård *et al.* (2018), Junior e Medrano (2018) e Ryu *et al.* (2020), e da mesma forma que na literatura mostraram um certo otimismo nas estimativas. Esse otimismo levou à uma análise mais profunda dos dados coletados. Assim foi proposta uma análise com os dados desagregados, análise não encontrada na literatura.

A partir da análise dos dados desagregados foi necessário realizar a caracterização da sinalização Wi-Fi, ou seja, com que frequência os dispositivos móveis emitem essa sinalização. Na literatura não há informações concretas sobre como isso ocorre. O estudo de Freudiger (2015), traz algumas informações, mas com sistemas operacionais ultrapassados. Nos sistemas operacionais atuais o comportamento difere muito da época da sua pesquisa. Assim, foi desenvolvida uma aplicação para dispositivos móveis que registra as sinalizações emitidas por um dispositivo móvel. A aplicação foi testada e validada através de experimentos com usuários. Todos os experimentos e testes foram realizados antes do atual cenário da pandemia de Covid-19. Novas coletas de dados foram inviabilizadas pela pandemia.

Em virtude do cenário onde não há possibilidade de realizar coleta de dados de detecções de passageiros e de GPS dos ônibus, como no caso da atual pandemia (WIKIPEDIA, 2020), é proposto um método para gerar as detecções de forma sintética, mas respeitando os intervalos de buscas de redes a partir dos dados coletados anteriormente. Como não havia formas de gerar os sinais de GPS para determinar onde essa detecção gerada ocorreu, foi utilizado o software Transcad para mostrar as informações geográficas. O transcad é um Sistema de Informação Geográfica (GIS) projetado especificamente para uso de profissionais de transporte para armazenar, exibir, gerenciar e analisar dados de transporte (CALIPER, 2020). Assim os dados gerados são mostrados visualmente como ocorreria em um experimento de campo.

O modelo de geração permite que outros dados possam ser usados para gerar a detecção. Dessa forma buscou-se novas fontes de dados disponíveis, para serem utilizadas para alimentar o modelo de geração de detecções.

Utilizar somente os dados brutos dos experimentos e das novas fontes de informações obtidas, agregando os dados e estimando o embarque dos passageiros não é eficaz. Dessa forma foi proposto para o problema um método baseado em probabilidades para estimar o embarque de cada passageiro individual. Esse método permite que seja atribuída uma probabilidade de que cada passageiro possa ter embarcado em um ponto de ônibus.

Os testes do modelo de geração de detecções e do método probabilístico foram aplicados em um trajeto real de uma linha de ônibus da cidade de Florianópolis no estado de Santa Catarina, onde foram obtidas as estimativas de embarques dos passageiros. Embora o cenário seja um trajeto real, os dados utilizados para essa estimativa são sintéticos, gerados a partir do modelo de geração de detecção.

2.5 POSSIBILIDADE DE APLICAÇÃO PRÁTICA

Embora a legislação na época dessa pesquisa não era específica com relação à coleta de dados, era necessário um cuidado com as informações coletadas, pois elas continham dados sensíveis a possíveis invasões. Hoje a legislação Brasileira é mais restritiva, o que dificulta muito pesquisas com a coleta de dados Wi-Fi.

Com a disseminação de dispositivos móveis e fabricantes espalhados, há uma margem muito grande para realizar pesquisas com aparelhos de várias marcas. Como já citado anteriormente a randomização de endereços MAC poderia ser um problema, mas ela pode ser superada, ainda mais que dispositivos com sistema operacional Android que são a maioria no Brasil ainda não fazem a randomização com muita frequência.

O protocolo 802.11 permite que sejam coletadas essas informações que são sensíveis para invasões e permitem que estudos através dos sinais Wi-Fi sejam realizados. Sem essas informações seria impossível realizar estudos nessa área porque não seria possível identificar cada dispositivo móvel.

A literatura sobre coleta de dados Wi-Fi é muito vasta. Alguns artigos recentes (SMIRNOV *et al.*, 2020; LIU, J. *et al.*, 2019; PAOLO *et al.*, 2019; URAS *et al.*, 2020), a coleta de dados Wi-Fi é utilizada para muitas aplicações. Recentemente está sendo utilizado com foco no transporte coletivo como em Ji *et al.* (2017), Junior e Medrano (2018), Ryu *et al.* (2020) e Nitti *et al.* (2020). Embora as informações mostrem um otimismo considerável, muitos aspectos nas pesquisas não são considerados e precisam ser analisados com mais cautela.

3 IMPLEMENTAÇÃO E AFERIÇÃO DO MONITOR DE TRÁFEGO NA REDE (MTR)

Com o objetivo de capturar sinais Wi-Fi emitidos por dispositivos móveis (ex. *smartphones*) portados por usuários do sistema de transporte coletivo, foi desenvolvido um dispositivo monitor do tráfego de requisições e de pacotes em redes sem fio Wi-Fi conhecido como NTM (*Network Traffic Monitoring*) ou MTR (Monitor de Tráfego na Rede). Neste capítulo são descritos os requisitos e a arquitetura do dispositivo. Além disso foram realizados os testes e aferição.

3.1 REQUISITOS A SEREM PREENCHIDOS PELO MTR

O principal requisito que o MTR precisa atender é capturar em tempo real toda e qualquer requisição de conexão de rede ou troca de pacote emitida por dispositivos móveis. Além desse, o MTR precisa preencher os requisitos da lista abaixo.

- Tenha capacidade de armazenar grande quantidade de dados;
- Consiga enviar os dados armazenados em tempo real para um servidor, caso haja necessidade de análise do carregamento de uma linha em tempo real;
- Seja portátil;
- Seja desenvolvido com um custo baixo.

3.2 ARQUITETURA PROJETADA PARA O MTR

O MTR ilustrado na Figura 6 foi desenvolvido com base no microcomputador Raspberry Pi 3 modelo B ilustrado na figura pelo número 1, equipado com placa Wi-Fi da intelbras modelo WBN-300 ilustrado na figura pelo número 2 e com uma bateria portátil da TP-Link, modelo TL-PB10400, ilustrado na figura pelo número 3. O sistema operacional do MTR é baseado na arquitetura Debian utilizando a distribuição Kali Linux (ALLEN *et al.*, 2014). Esse sistema operacional é próprio para aplicações que envolvam ferramentas de teste de penetração em redes Wi-Fi e campos de segurança forense. A distribuição Kali conta com o software Aircrack-ng (KUMKAR *et al.*, 2012; BEGGS, 2014; HUTCHENS, 2014; NAJERA-GUTIERREZ; ANSARI, 2018), sendo altamente indicado para coleta de informações em redes Wi-Fi por possuir ferramentas próprias para o mesmo.

Quando se trata de coletar informações de dispositivos móveis sempre há interação entre dois dispositivos. O primeiro é responsável por permitir que uma informação seja enviada caso ele seja requisitado, seja um roteador ou um ponto de acesso. O segundo, um dispositivo móvel, está tentando se conectar pedindo informações do roteador ou do ponto de acesso. O MTR figurativamente está no meio desses dois. Enquanto está sendo

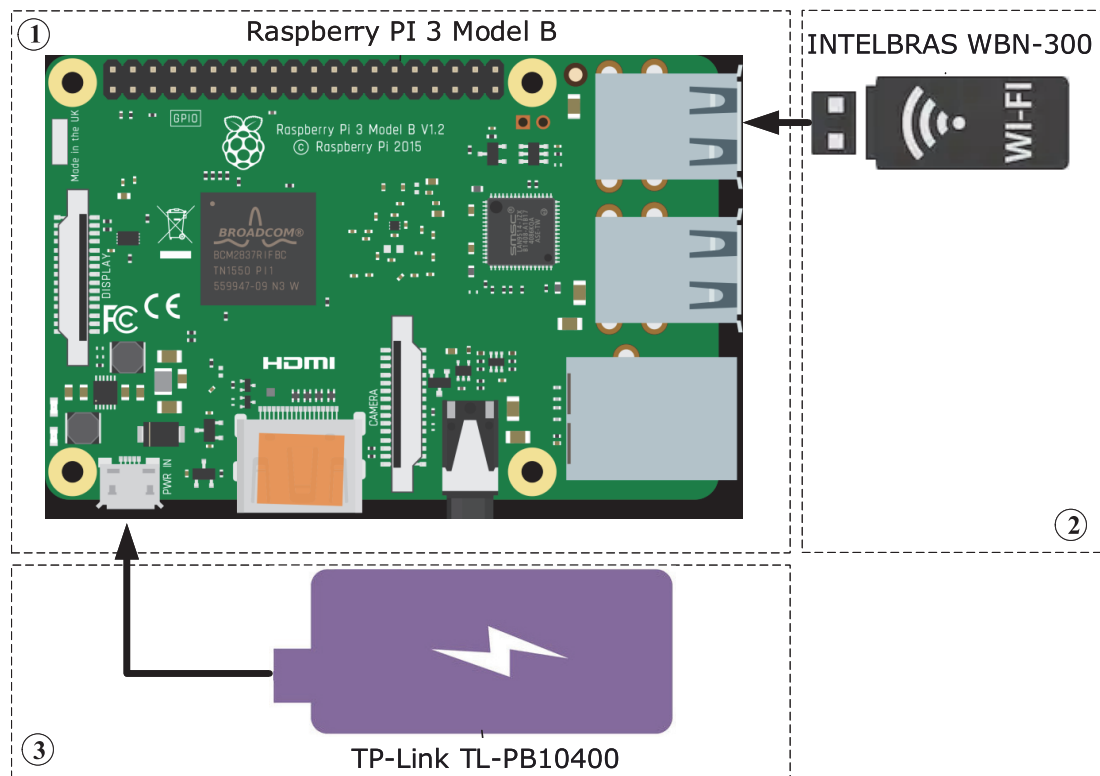


Figura 6: Componentes do dispositivo MTR.

realizada essa troca de informações o MTR captura o dado que está sendo trocado, dentro do seu raio de detecção.

A Figura 7 ilustra a representação de um dispositivo móvel realizando uma requisição de rede a qualquer dispositivo como roteadores ou ponto de acesso, que responde a requisição. A informação trocada entre dispositivos pode ser qualquer uma da Tabela 5. Essa informação é capturada pelo dispositivo MTR ilustrado pelo retângulo tracejado.

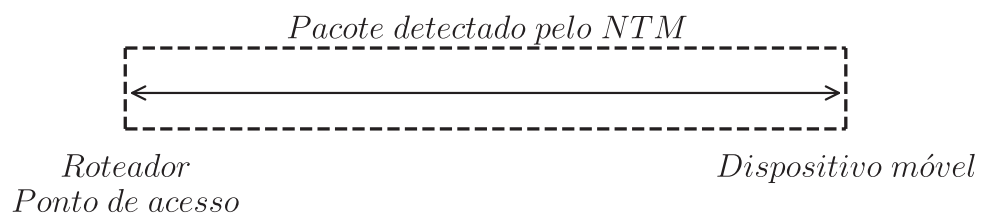


Figura 7: Ilustração de como o MTR age quando há troca de pacotes entre um roteador e um dispositivo móvel.

O dispositivo MTR desenvolvido quando captura uma informação como ilustrado na Figura 7, tem a capacidade de registrar informações de cada dispositivo móvel que está tentando realizar uma conexão Wi-Fi ou que já está conectado e está realizando troca de pacotes entre roteador ou ponto de acesso. As informações registradas estão representadas

abaixo, e são individuais para cada dispositivo móvel:

- Endereço MAC: endereço físico único de cada dispositivo móvel que teve seu sinal detectado pelo MTR;
- Primeira vez visto: primeira vez que um endereço MAC de um dispositivo móvel foi registrado pelo MTR;
- Última vez visto: última vez que um endereço MAC de um dispositivo móvel foi registrado MTR;
- Potência do sinal: valor de potência do sinal da troca de informações entre do dispositivo móvel e o MTR;
- Pacotes: número de pacotes reportado entre a troca de informações entre roteador e dispositivos móveis;
- BSSID: endereço MAC de um ponto de acesso mais conhecido como roteador ao qual o dispositivo móvel está conectado ou tentando se conectar; e
- *Probed* ESSIDS: endereço MAC do ponto de acesso ao qual um dispositivo Wi-Fi está tentando se conectar.

Essas informações são coletadas constantemente de todos os dispositivos móveis que estiverem no raio de alcance do MTR e são armazenadas para futuras análises. Quando um dispositivo móvel sai do raio de detecção do MTR essas informações deixam de ser registradas. Todas essas informações coletadas são necessárias para conseguir realizar a identificação de cada dispositivo móvel conseguindo distinguir cada um caracterizando o momento da primeira e última vez que o dispositivo móvel foi registrado.

Após o desenvolvimento do MTR, testes e experimentos em campo foram realizados para determinar a eficácia do dispositivo.

3.3 AFERIÇÃO DO MTR

Para determinar a eficácia do MTR dois testes foram realizados. O primeiro teste de confiabilidade, é baseado na capacidade de detecção, isto é, o MTR é realmente capaz de detectar dispositivos que estejam realizando uma busca de rede para se conectar. Se isso for possível, verifica-se com que confiabilidade isso ocorre, ou seja, se há possibilidade de perder informações ou de não detectar algum dispositivo móvel. O segundo teste, de raio de detecção, avalia qual a distância mínima com que o MTR é capaz de detectar um dispositivo móvel.

3.3.1 TESTE DE CONFIABILIDADE DO MTR

O teste de confiabilidade foi dividido em duas etapas. Na primeira etapa o MTR foi configurado com uma placa Wi-Fi, e através de codificação foi capaz de comutar entre as 13 frequências (Capítulo 2, Seção 2.3.4) em intervalos de milissegundos. O teste foi realizado por 24 horas com um dispositivo Wi-Fi de controle (com MAC conhecido) realizando buscas de rede.

O resultado indicou que o MTR é capaz de detectar os dispositivos móveis. Porém o dado mais importante observado foi a presença de um espaçamento longo entre as detecções, o que pode ter duas causas. A primeira causa é a falta de sincronia entre o dispositivo Wi-Fi e o MTR. Ocorre quando o dispositivo de controle está realizando uma busca de rede em uma frequência diferente. Por exemplo, o MTR está realizando a busca de rede na frequência 1 e o dispositivo de controle está realizando a busca de rede na frequência 5. Isso faz com que o MTR perca o registro dessa busca de rede. A segunda causa é o intervalo entre as buscas de rede do dispositivo de controle, que será discutida no Capítulo 5.

Na segunda etapa foram utilizados quatro MTR's cada um configurado com uma placa Wi-Fi. Através de codificação a comutação foi realizada entre grupos de frequências, são eles: $F1 = [1, 2, 3]$, $F2 = [4, 5, 6]$, $F3 = [7, 8, 9]$, $F4 = [10, 11, 12, 13]$. Onde o MTR 1 é responsável por comutar entre as frequências do grupo $F1$, MTR 2 pelo grupo $F2$, MTR 3 pelo grupo $F3$ e o MTR 4 pelo grupo $F4$. Os grupos foram ajustados devido a restrições de projeto.

Embora haja espaço físico no MTR para colocar quatro placas de rede, há somente um processador em cada MTR para processar informações. Assim o uso de quatro placas Wi-Fi, poderia levar à lentidão e sobrecarga no sistema e eventualmente à perda de informações. Dessa forma optou-se pela utilização de quatro MTR's cada um realizando o processamento individualmente de cada grupo. O dispositivo Wi-Fi de controle foi mantido o mesmo da etapa anterior.

O resultado dessa etapa é apresentado na Tabela 7. Como pode ser observado na Tabela 7, o número de detecções aumenta quase 50% quando a detecção ocorre por agrupamento de frequências, já que o agrupamento faz com que o intervalo entre comutação de frequências seja mais ágil. Esse resultado mostra que foi coletado um número maior de informações, mas ainda não garante que toda busca por rede que o dispositivo de controle realiza é detectado. Isso ocorre por não sabermos quantas buscas de rede o dispositivo de controle realmente realizou, essa aferição é apresentada no Capítulo 5.

O MTR foi projetado para locais onde os dispositivos móveis Wi-Fi que serão detectados permanecerão por um certo intervalo de tempo. Utilizar quatro MTR's traz um benefício para detecções imediatas, quando o parâmetro buscado são detecções ocorrendo em intervalos que indiquem que a detecção não seja esporádica.

Os testes realizados garantem que dispositivos móveis Wi-Fi podem ser detectados

Tabela 7: Detecções coletadas por um MTR e quatro MTR's durante 2 horas.

Hora	Número de placas Wi-Fi		Diferença	%
	1	4		
1	91	154	+63	69
2	83	158	+75	90
Total	174	312	+138	79

mesmo sem estar conectados na rede. Porém não permitem saber a real eficácia do MTR, porque não se sabe se as buscas de redes por dispositivos detectados foram todas as realizadas pelo dispositivo Wi-Fi.

O teste que comprova a eficácia do MTR é apresentado no Capítulo 5. O resultado desse teste indica que menos de 1% das buscas realizadas por dispositivos Wi-Fi são perdidas pelo MTR. Esse número indica que se um dispositivo móvel Wi-Fi realizar uma busca por rede há 99% de chance dele ser detectado. Para manter a estrutura do documento que é baseado na ordem cronológica dos eventos, esse teste é apresentado na Seção 5.2.

3.3.2 RAI0 DE DETECÇÃO DO MTR

O segundo teste é baseado na distância em que o MTR consegue coletar um dado. Como o projeto pretende inserir um MTR dentro de um veículo, podem haver detecções fora do raio desejado. Dessa forma um ajuste na potência do MTR para limitar o seu raio de detecção pode ajudar a não detectar dispositivos móveis muito distantes. Isso reduz o volume de dados coletados, além de facilitar na filtragem de dispositivos móveis que podem ser considerados como possíveis passageiros em um embarque ou desembarque de um ônibus. Testes iniciais indicaram que em campo aberto a distância de detecção do MTR pode chegar até 100 metros o que indica que muitos dispositivos móveis que estão muito distantes podem ser detectados.

À medida que descreve a potência do sinal e define o tamanho do raio de detecção é representada em decibéis por miliwatts (dBm ou dB_{mW}). O ajuste dessa configuração na placa Wi-Fi do MTR irá determinar o alcance de detecção do sinal.

Por padrão muitos dispositivos Wi-Fi no Brasil vêm configurado para 20 dBm, o que equivale a 100 mW de acordo com a seguinte relação:

$$P = 1 \times 10^{\frac{x}{10}} \quad (1)$$

em que P é a potência em dBm.

O cálculo da distância envolve muitos parâmetros, como perda de transmissão, sensibilidade do receptor, ganho, perda, margem de desvanecimento¹, o que torna esse cálculo complexo pela dificuldade em obter alguns desses parâmetros. O teste em campo se tornou mais prático para obter a distância desejada.

¹ Diferença entre o nível de recepção nominal e o limiar de recepção (MANNING, 2009).

Como referido anteriormente distâncias de detecção muito grandes podem levar ao aumento do volume de dados coletados e um número elevado de dispositivos móveis que futuramente serão descartados na filtragem. Dessa forma o objetivo do teste é obter a menor distância com a qual o MTR consegue detectar algum dispositivo Wi-Fi e assim ajustar de acordo com a necessidade, como por exemplo, para o tamanho de um ônibus de 18 metros.

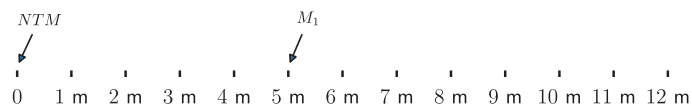
O teste consiste em manter o MTR em uma posição fixa, com um dispositivo Wi-Fi de controle realizar buscas por rede, e variar a posição do dispositivo de controle afastando-o gradativamente do MTR. Registra-se com um medidor a distância entre o MTR e o dispositivo de controle assim como a potência do sinal registrada pelo MTR. Os testes foram realizados em campo aberto sem quaisquer outros dispositivos e interferências. O menor valor de potência registrado foi 1 mW, sendo a potência mínima de operação da placa Wi-Fi do MTR. Com essa potência a distância máxima de detecção registrada foi 12 metros de raio, isso significa que, nenhum dispositivo que estiver a mais de 12 metros de raio do MTR será detectado, permitindo uma detecção de 24 metros de diâmetro.

Com o teste do raio de detecção, também foi possível realizar um teste para análise da sensibilidade da potência do sinal entre o MTR e os dispositivos móveis Wi-Fi. A potência do sinal do MTR é RX (Recepção), enquanto a dos dispositivos Wi-Fi é TX (Transmissão). A potência registrada pelo MTR significa que o dado de potência registrado não é a potência do dispositivo Wi-Fi e sim a potência que a placa Wi-Fi do MTR captou essa informação.

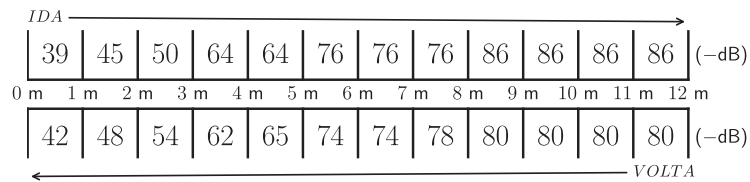
Para esse teste foram colocadas 12 marcações espaçadas de um metro entre si como ilustrado na Figura 8(a). O MTR foi colocado na posição 0 e, como exemplo, o dispositivo de controle *M1* na marcação 5. O dispositivo Wi-Fi de controle inicialmente posicionado na marcação 0 mais próximo junto ao MTR foi deslocado aos poucos até a marcação 12.

Na Figura 8(b) está representado as potências registradas no momento em que o dispositivo *M1* estava em repouso em cada uma das marcas. Quanto maior a potência do sinal em dB, mais próximo do MTR o dispositivo Wi-Fi está. Nesse caso pode ser observado que partindo da posição 0 até a posição 1, em média a potência do sinal registrada é em -39 dB e quanto mais afastado do MTR a potência vai sendo registrada mais baixa. Com base nessas informações, é garantido afirmar que quanto mais distante do MTR, menor é a chance de um dispositivo móvel Wi-Fi ser detectado.

Com base em todos os testes realizados, sabe-se que o MTR realmente é capaz de detectar um dispositivo móvel Wi-Fi se o mesmo estiver buscando rede, desde que o mesmo fique tempo suficiente para ser detectado. Esse tempo é relativo pois depende da configuração e aplicativos instalados de cada aparelho. No Capítulo 5 é discutido o comportamento de cada aparelho, fornecendo um tempo médio para que ocorra essa detecção. A potência do sinal para detecção do MTR foi fixada em 1 mW. Como o maior ônibus em operação na área de aplicação dos testes possui 18 metros, procura-se dessa



(a)



(b)

Figura 8: Potência do sinal registrada com dispositivo móvel em repouso em marcações espaçadas de 1 metro entre si.

forma evitar a detecção de dispositivos que estejam fora do ônibus que possam gerar falsos positivos. Após todos os testes foram realizados experimentos em campo com cenários controlados.

4 EXPERIMENTOS DE CAMPO

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 9 mostra a configuração do experimento no campo. Um dispositivo MTR 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 MTR, foi posicionado a 12 m de um dos marcadores com o objetivo de detectar os dispositivos que saiam do raio de detecção do MTR central por aquela extremidade.

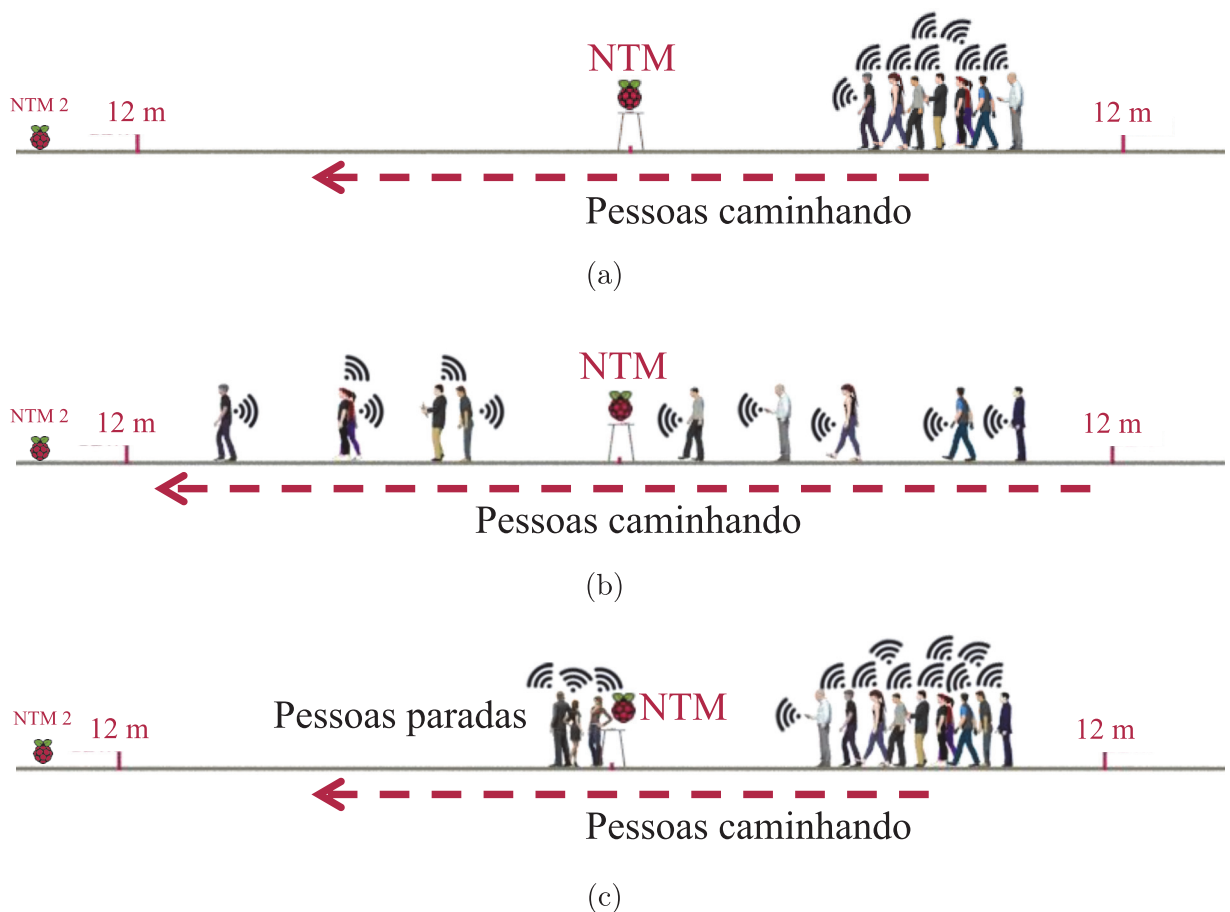


Figura 9: Configuração e etapas do Experimento 1: (a) Etapa 1- pessoas passaram em grupo em frente ao detector; (b) Etapa 2 - pessoas passaram em intervalos diferentes em frente ao detector; (c) Etapa 3 - pessoas paradas enquanto outras passaram em frente ao detector.

O experimento foi dividido em três etapas. A primeira etapa está representada na Figura 9(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 com relação à quantidade de dados processados, quando da travessia de um grupo de pessoas simultaneamente pela área de captura.

A segunda etapa está representada na Figura 9(b). Consistiu no deslocamento dos participantes em momentos diferentes de uma extremidade à outra da área do experimento passando em frente ao MTR. Esta etapa foi realizada para verificar a capacidade de detectar dispositivos em momentos diferentes.

A terceira etapa está representada na Figura 9(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.

A primeira detecção de um usuário é importante para determinar a latência do MTR. Isto é, verificar quanto tempo é necessário para que o usuário seja detectado pela primeira vez depois de entrar no raio de detecção. Muitos fatores podem influenciar essa detecção. Uma delas é o dispositivo estar procurando uma rede Wi-Fi para se conectar no momento que ele entra no raio de detecção. Se isso ocorrer o usuário será detectado em um intervalo rápido, mas caso ele não procure uma rede Wi-Fi no momento de entrada no raio, esse intervalo será maior. Isso varia de dispositivo para dispositivo. O experimento controlado nos forneceu uma ideia do comportamento de alguns dispositivos quando da busca por rede Wi-Fi.

Por se tratar de um experimento controlado foi possível observar os comportamentos de cada um dos indivíduos do experimento. A dinâmica do experimento consiste em um participante obrigatoriamente estar com o dispositivo Wi-Fi ligado, entrar no raio de detecção caminhando e após 24 metros sair do raio de detecção. Os participantes ficaram encarregados de registrar os instantes de entrada e saída do raio de detecção através de um aplicativo instalado em seus dispositivos móveis, para esse experimento foi utilizado o aplicativo Counter (TSUKANOV, 2018) para Android e o Tally Counter (PIXEL RESEARCH LABS, 2018) para iOS.

Algumas observações precisam ser feitas com relação a esse experimento. Como se trata de um experimento com pessoas, há um atraso na percepção do participante que ele entrou e saiu do raio de detecção, dessa forma subtrai-se e soma-se 3 segundos do tempo marcado pelo participante, na entrada e saída do raio de detecção respectivamente. Isso é feito porque, caso o participante seja detectado 3 segundos antes da marcação pelo mesmo, esse intervalo será considerado, caso isso não seja feito, a primeira detecção pode ocorrer muito mais tarde do que efetivamente ocorreu.

A segunda observação é com relação à característica dos experimentos. Como os experimentos possuem duas características diferentes, com participantes caminhando e participantes parados, a primeira detecção pode ser afetada. Dessa forma optou-se por analisar os resultados de duas formas. A primeira isolando os dados dos usuários caminhando ao longo dos experimentos e em seguida os dados dos usuários parados ao longo dos experimentos, e extrair as estatísticas de ambos.

4.1 RESULTADOS AGREGADOS

Todos os experimentos foram analisados individualmente, mas os resultados serão apresentados de forma agregada, devido à possibilidade de isolar a única diferença entre os experimentos, sendo o participante parado ou em movimento.

Na análise dos dados busca-se informações da primeira detecção, a qual deve indicar o momento em que o dispositivo móvel ficou visível pela primeira vez para o MTR. No caso do MTR embarcado no ônibus a primeira detecção dá indícios de que o dispositivo móvel está dentro do ônibus. A segunda detecção é uma forma de confirmação de que o dispositivo móvel realmente está embarcado dentro do ônibus.

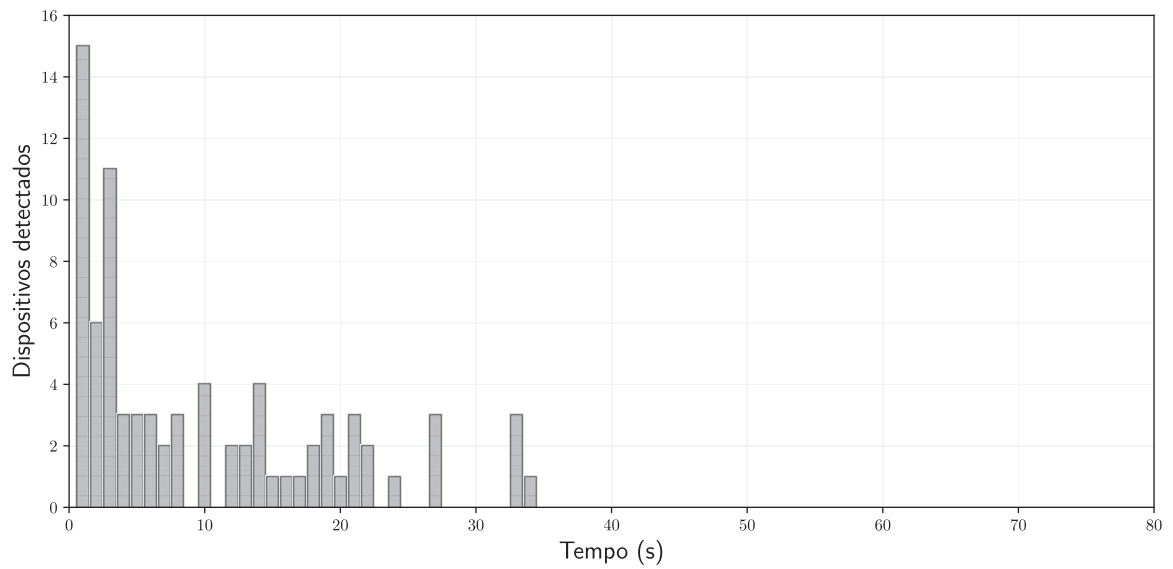
Do ponto de vista agregado, os dados detectados apresentam sucesso na detecção e estimação de passageiros. A Tabela 8 indica que nos primeiros 40 segundos da coleta de dados, 86% dos dispositivos móveis dos participantes foram detectados em um experimento ao menos uma vez; os intervalos para detecções subsequentes também estão incluídos nesta taxa de detecção. Como esse intervalo é consistente com o tempo de viagem entre dois pontos de ônibus, pode indicar com um nível alto de confiança que o sistema está conseguindo estimar os passageiros somente considerando esse resultado.

A representação para o melhor entendimento desses resultados é ilustrada pela Figura 10. A Figura 10(a) ilustra o histograma dos tempos da primeira detecção com dados agregados em classes com intervalos de 1 segundo de todas as rodadas de todos os experimentos. Nesse caso, todas as detecções ocorreram com um intervalo de até 35 segundos. Mas será discutido nos dados desagregados que nem todos os dispositivos móveis foram detectados em cada rodada de experimento.

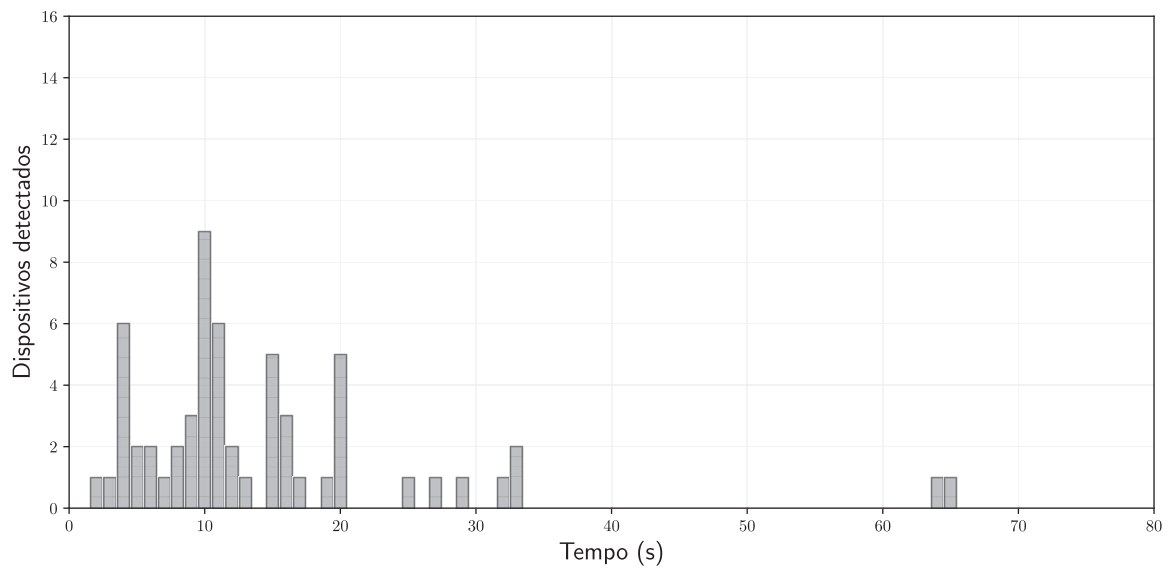
A segunda detecção é ilustrada na Figura 10(b). Novamente, há muito otimismo quando observado que a maioria das detecções ocorreram em um intervalo de até 35 segundos. No entanto como no caso da primeira detecção, nem todos os dispositivos móveis participantes foram detectados uma segunda vez na mesma rodada dos experimentos. Se isso ocorresse confirmaria que o dispositivo MTR seria capaz de detectar múltiplas vezes um mesmo dispositivo ao longo do itinerário. No entanto sabe-se que isso não é problema de detecção pelo MTR e sim do intervalo entre buscas de rede realizados por dispositivos móveis, que será discutido no Capítulo 5.

Tabela 8: Intervalos entre detecções sucessivas de um mesmo dispositivo e detecções acumuladas em todas as rodadas de todos os experimentos.

Intervalo (s)	Taxa de detecção cumulativa (%)
40	86
50	89
60	93
70	95
80	96
90	97



(a)



(b)

Figura 10: Histograma (a) tempo da primeira detecção e (b) segunda detecção agrupados.

4.2 RESULTADOS DESAGREGADOS

Com o objetivo de obter as características intrínsecas do conjunto de dados, é necessário desagregar os dados em cada uma das rodadas dos experimentos. O resultado ilustrado na Tabela 9 indica que a taxa de sucesso na primeira detecção fica em média em 56%; isso significa que ao invés de obter 48 detecções, somente metade delas realmente ocorreram.

Em termos de dispositivos detectados em todas as quatro passagens, o cenário é melhor, indicando que eventualmente a maioria dos dispositivos no final acabam sendo detectados em alguma passagem. A taxa de sucesso de detecção é de 92% para a primeira detecção, indicando que há 92% de chance que um dispositivo será detectado a primeira vez, e de 70% em média para a segunda detecção, indicando que há 70% de chance de um dispositivo ser detectado a segunda vez após ter sido detectado a primeira vez, isso apesar da alta taxa de falha nas detecções em cada passagem.

A Figura 11 ilustra os resultados para as ocorrências da primeira detecção em cada uma das passagens dos participantes pelo detector. A aleatoriedade se manifesta nas diferenças entre os tempos em que o usuário foi detectado e a ocorrência de detecções. Em cada experimento o comportamento é diferente para um mesmo usuário.

A Figura 12 ilustra os resultados da segunda detecção dos dispositivos móveis. Os resultados indicam que na maioria dos casos nem todos os participantes tiveram uma segunda detecção durante o experimento comparado com o caso da primeira detecção, onde a maioria teve sua primeira detecção. Isso leva a crer que o tempo de busca de rede de cada aparelho mesmo sendo da mesma marca e mesmo modelo possui um comportamento diferente em cada experimento, esse detalhe gera um alerta de que talvez seja necessário analisar os dispositivos móveis de cada participante. Além disso, quando a segunda detecção ocorre, no geral ela leva um tempo relativamente longo para ocorrer, quando ocorre, como pode ser observado em todas as figuras.

Tabela 9: Resultados de quatro rodadas de cada experimento; são esperadas 48 detecções por experimento (12 dispositivos, 4 rodadas).

Exp.	Detecção	Dados Desagregados		Dados Agregados		
		Detecções sucessivas	Sucesso (%)	Dispositivos detectados	Sucesso (%)	Tempo máximo até detecção (s)
1	1 ^o	27	57	11	92	20
	2 ^o	16	34	7	58	19
2	1 ^o	24	50	11	92	34
	2 ^o	22	46	10	83	81
3	1 ^o	29	61	11	92	33
	2 ^o	22	46	8	67	65

Os resultados apresentados sugerem que as detecções de dispositivos através do Wi-Fi são erráticas, sem garantia de que todo dispositivo será detectado em cada passagem do

experimento. A segunda detecção, importante para a rastreabilidade de um dispositivo, leva muito tempo para ocorrer e ocorre com uma taxa de sucesso muito baixa, diferentemente das primeiras detecções em cada experimento.

Com base nesses resultados não é indicado inserir o dispositivo de coleta em pontos de ônibus, isso porque o intervalo de passagem de um ônibus pode ser muito curto a ponto de não conseguir detectar um passageiro caso o veículo somente passe pelo ponto e não pare, ou até mesmo se o tempo de embarque e desembarque for muito rápido a ponto de não conseguir detectar novamente o dispositivo.

Dessa forma novos experimentos foram considerados dessa vez com o dispositivo de detecção embarcado dentro do veículo, e contando ainda com os dispositivos de detecção em pontos de ônibus.

4.3 EXPERIMENTO DE CAMPO COM DISPOSITIVO A BORDO DE UM ÔNIBUS

Os resultados apresentados nas seções anteriores são divergentes. Enquanto os dados agregados dão uma falsa impressão de que é possível estimar os embarques e desembarques baseado somente nessas informações, os dados desagregados mostram que nem todos os dispositivos móveis são detectados adequadamente. Com o objetivo de aprofundar ainda mais a compreensão sobre o comportamento do tipo de dado que está sendo coletado, foi desenvolvido um experimento combinando quatro dispositivos MTR, um embarcado e três em pontos de ônibus, e três pesquisadores com dispositivos móveis de controle. O cruzamento dos dados coletados em pontos de ônibus e os coletados embarcados fornecem informações sobre passageiros que estavam dentro do ônibus e foram detectados fora do ônibus, e sobre os passageiros detectados anteriormente no ponto de parada de ônibus fora do ônibus, se também foram detectados dentro do ônibus.

Para o experimento, foi escolhida uma seção da linha de ônibus 185 UFSC Semi-Direto nas proximidades da Universidade Federal de Santa Catarina. Esta é a linha mais frequente e com maior demanda na região. Os pontos de ônibus estão identificados na Figura 13 como O, A, B e C. O trajeto é dividido em três segmentos, OA (548 m), AB (370 m), e BC (330 m), com o início e fim de cada segmento sendo um ponto de ônibus. Ao todo foram utilizados quatro dispositivos de detecção, nos pontos A, B e C, e o quarto com um dos pesquisadores embarcado no veículo. Os dados de GPS com informações do veículo foram obtidos com o operador. Os embarques dos pesquisadores ocorrem no ponto O e os desembarques no ponto C. Foram realizadas três viagens pelo trajeto ilustrado na Figura 13. O tempo de trajeto é de aproximadamente 15 minutos, totalizando cerca de 45 minutos de experimento.

No sistema de transporte coletivo da cidade de Florianópolis são utilizados ônibus do tipo Padron de 15 metros com portas na frente no meio e no fundo do veículo e também veículos articulados de 18 metros com a mesma configuração de portas do veículo Padron. Os embarques de todos os passageiros ocorrem somente pela porta dianteira, exceto em

casos de embarques de portadores de necessidades especiais, onde geralmente o embarque e desembarque ocorre pela porta do meio do veículo. Já os desembarques dos demais passageiros podem ocorrer nas portas centrais e do fundo do veículo.

Em cada uma das três viagens cada pesquisador embarcou com um dispositivo móvel com Wi-Fi e tendo seu endereço MAC conhecido. Os três dispositivos móveis dos pesquisadores formaram um conjunto de dispositivos de controle. Um dos pesquisadores também embarcou com um dispositivo MTR. Esse pesquisador ficou localizado no centro do veículo e também foi o responsável por contar os desembarques que ocorriam nas portas centrais do veículo. O segundo pesquisador ficou localizado nos fundos do veículo e foi o responsável por contar os desembarques realizados pela porta traseira, enquanto o terceiro pesquisador ficou localizado na frente do veículo antes do sistema de pagamento e foi o responsável por contar os embarques realizados pela porta dianteira.

As contagens de embarque e desembarque de passageiros foram realizadas a cada procedimento de parada do veículo. A contagem de passageiros já embarcados no ponto O foi realizada imediatamente após o embarque dos pesquisadores. Quando o veículo chegou no ponto C, o último do trajeto, os três pesquisadores desembarcaram e se deslocaram até o ponto O para realizar um novo embarque.

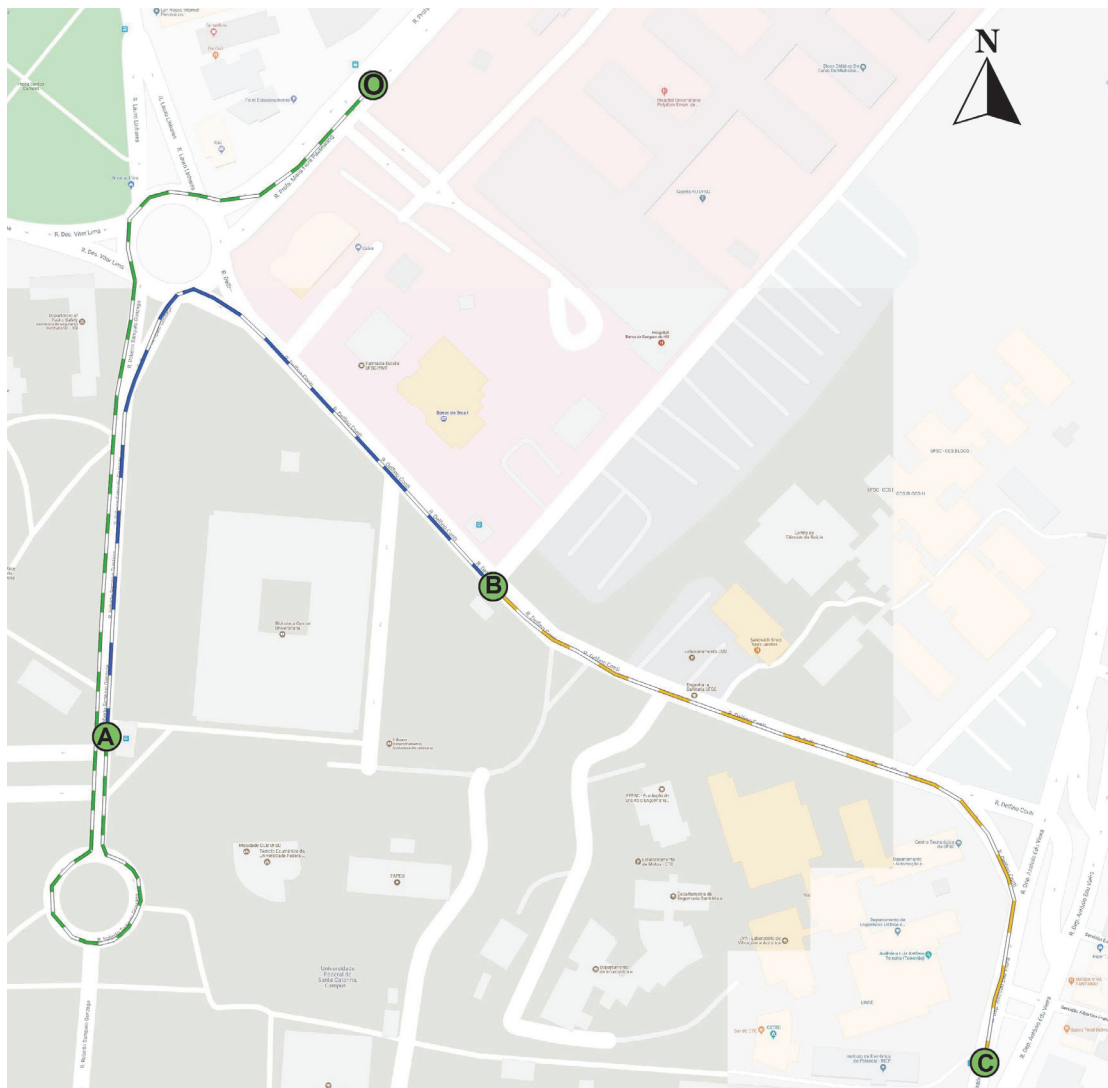


Figura 13: Trajeto realizado pela linha. Os pontos de paradas de ônibus estão identificados com as letras O, A, B e C. Os segmentos do trajeto nas linhas tracejadas representam o deslocamento do veículo entre dois pontos de ônibus consecutivos.

4.4 PROCESSAMENTO DOS DADOS COLETADOS A BORDO DE UM ÔNIBUS

O processamento dos dados consiste em analisar os dados obtidos e aplicar os filtros necessários para qualificar melhor as informações. O processo é realizado em cinco etapas.

- Remoção de qualquer dado corrompido ou com problemas de informação (ex. MAC incompleto, divergência em data ou horário);
- Contagem do número de vezes que um dispositivo móvel foi detectado, descartando aqueles que foram detectados somente uma vez;
- Identificar a potência do sinal com relação ao dispositivo de coleta, devendo obedecer a valores entre -79 dB e -1 dB;
- Classificação dos segmentos nos quais o passageiro foi detectado;
- Cruzamento entre o tempo de detecção em que o passageiro foi detectado e a distância do ônibus do ponto de ônibus.

Um dispositivo móvel é considerado válido se e somente se foi detectado no mínimo após 20 s da partida do ônibus do ponto ou se a detecção ocorreu após o ônibus estar afastado no mínimo 10 m do ponto de parada e que o tempo de viagem de um usuário portando um dispositivo móvel seja maior que 3 minutos e menor que 60 minutos. Essa premissa foi determinada para evitar detecções de dispositivos móveis que estejam fora do veículo. Assim, um dispositivo móvel que foi detectado fora de ônibus se for detectado novamente após 20 segundos ou 10 metros depois do ônibus partir do ponto, tem grande chance de estar dentro do veículo. Para o tempo de viagem assume-se que o usuário não irá realizar viagens menores que 3 minutos e maior que 60 minutos.

Após esse procedimento os passageiros são classificados como embarque ou com desembarque de acordo com as premissas:

- Um dispositivo móvel é considerado embarque se e somente se o dispositivo foi detectado pelo dispositivo de detecção externo pela primeira vez antes de o ônibus chegar no ponto de parada de ônibus; se a potência média do sinal de detecção em relação ao dispositivo de detecção no ponto está entre -40 dB e -79 dB; se o número de detecções sucessivas do dispositivo móvel pelo MTR é maior que três dentro do ônibus.

- Um dispositivo móvel é considerado como desembarque se e somente se o dispositivo móvel foi detectado a primeira vez pelo MTR do ponto de parada de ônibus no momento em que o ônibus chega no ponto; se a potência média do sinal em relação ao MTR registrada entre -70 dB e -79 dB; se o número de detecções sucessivas pelo MTR do dispositivo móvel estava entre dois e três dentro do veículo e o dispositivo móvel não foi detectado pelo MTR a bordo ao longo do trajeto restante. A segunda detecção não pode ter ocorrido em menos que 18 s para o caso de duas detecções, isso porque o tempo médio para o desembarque é de 19 segundos. Caso a primeira detecção ocorra exatamente no momento em que o passageiro portando o dispositivo móvel esteja descendo do veículo no ponto de parada de ônibus, ele levaria no mínimo até 19 segundos para sair do raio de detecção do MTR ocorrendo sua segunda detecção nesse instante.

4.5 RESULTADOS DO EXPERIMENTO A BORDO DE UM ÔNIBUS

Após o processamento, dois resultados foram obtidos. O primeiro é baseado nas detecções acumuladas ao longo do trajeto comparadas com a contagem manual embarcada, ilustrado na Figura 14. Na Figura 14 as barras vermelhas hachuradas com x indicam as contagens manuais de passageiros embarcados realizadas pelos pesquisadores a bordo do ônibus. A identificação da detecção dos dispositivos de cada um dos pesquisadores são representados pelos símbolos X , \bullet , \star . Eles são representados no gráfico no momento que foram detectados pelo MTR. A linha azul ilustra os dispositivos móveis detectados pelo MTR que após a filtragem foram considerados válidos. Por fim o hachurado laranja representa o momento em que o ônibus para no ponto de parada de ônibus para realizar o momento de embarque e desembarque de passageiros, o comprimento da barra representa o tempo que o veículo ficou parado no ponto.

Na rodada do experimento da Figura 14(a), é possível identificar que há um número maior de passageiros detectados a bordo no segmento OA (de 0 a 190 s) se comparado com a contagem manual. Esse resultado mostra o caso de falsos positivos, isso é, um número maior de dispositivos móveis foram detectados do que deveriam ter sido. Além disso o momento das detecções dos dispositivos móveis dos pesquisadores também não ocorre no instante do embarque, mas sim ao longo do trajeto. É possível observar que o dispositivo móvel do pesquisador \star não foi identificado no segmento OA. O problema de falsos positivos não ocorre no segmento AB (de 270 a 330 s) e BC (de 390 a 430 s). De fato, o número de detecções é muito menor do que a contagem manual, possivelmente pelo curto intervalo do tempo de viagem entre os pontos e o tempo necessário para que um dispositivo móvel seja detectado como visto anteriormente na Seção 4. O problema de não detectar os pesquisadores ocorre novamente em ambos os segmentos, no segmento AB os dispositivos móveis dos pesquisadores X e \star não foram detectados, e no segmento BC somente foi detectado o dispositivo móvel do pesquisador \star .

Nas rodadas seguintes do experimento, ilustradas nas Figuras 14(b) e 14(c), as contagens foram subestimadas em todos os segmentos. Os problemas de detecção dos dispositivos móveis dos pesquisadores não serem detectados em algum momento do trajeto ocorreu em ao menos um segmento do trajeto em ambas as rodadas.

Uma razão que pode justificar os falsos positivos é o congestionamento de veículos no trajeto do ônibus. Nesse caso em especial a região de estudo é uma área com bastante congestionamento o que faz com que o ônibus trafegue em uma velocidade muito baixa. Além de ser uma área com bastantes pedestres fazendo com que o ônibus pare muito para travessia de pedestres. Isso faz com que pedestres que estejam fora do ônibus possam ser detectados mais de duas vezes pelo dispositivo MTR embarcado. Os falsos negativos em potenciais são dispositivos que deveriam ser detectados e não foram. Isso pode ocorrer em dispositivos móveis portados por passageiros embarcados que estão com o Wi-Fi desligado fazendo com que não haja busca de rede, nesse caso não há como fazer o passageiro ser detectado.

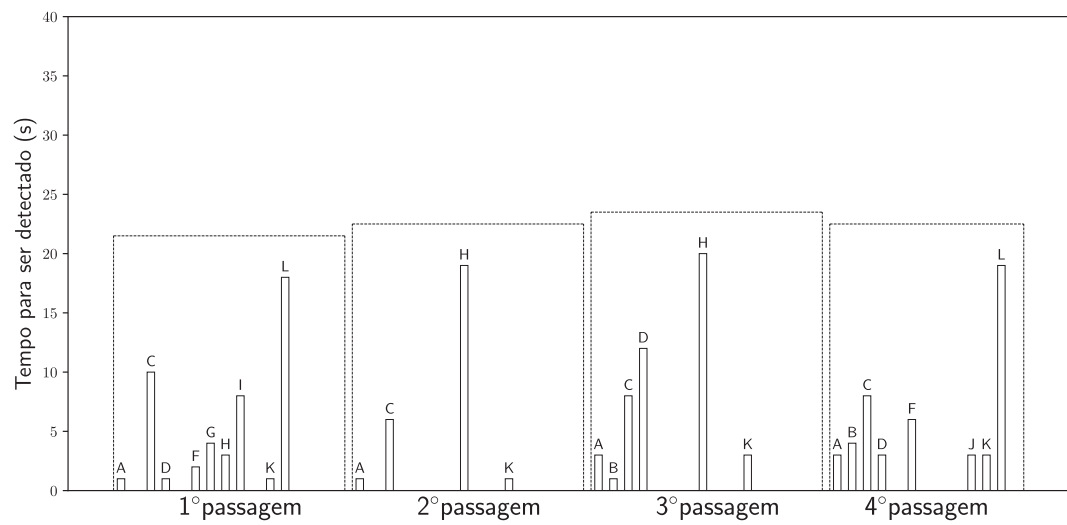
4.6 SÍNTESE DOS RESULTADOS DO EXPERIMENTO A BORDO DE ÔNIBUS

Analisando os resultados agregados em cada uma das rodadas do experimento ilustrado na Tabela 10, pode ser observado que os falsos positivos parecem desaparecer. Isso dá a impressão de que a contagem automática será sempre menor do que a contagem manual, o que não é verdade para o segmento OA na primeira rodada do experimento. A análise desse tipo de resultado de forma agregada, pode ser observado em vários artigos na literatura (DUNLAP *et al.*, 2016; JI *et al.*, 2017; HÅKEGÅRD *et al.*, 2018; NUNES, 2018; JUNIOR; MEDRANO, 2018; RYU *et al.*, 2020).

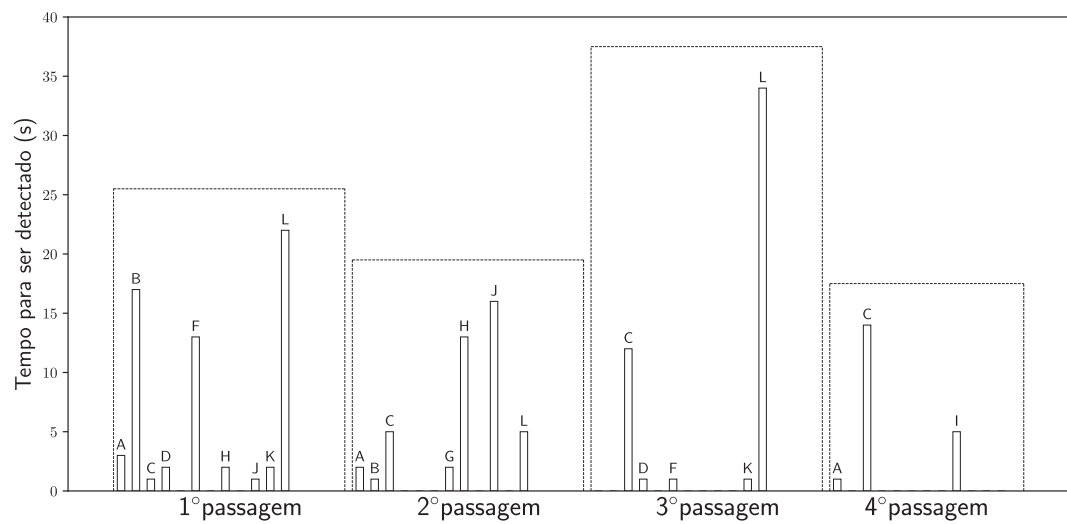
No entanto, analisando os dados de forma desagregada pode levar a um entendimento melhor do comportamento do sistema de detecção. Como pode ser observado na Tabela 10 as contagens estimadas através da detecção pelo MTR e as contagens manuais realizadas pelos pesquisadores mostram uma disparidade significativa. Novamente a presença de falsos positivos pode ser explicada pela velocidade operacional do ônibus como já citado anteriormente, os falsos negativos é que levantam a dúvida da eficiência do MTR, o que levou ao desenvolvimento do aplicativo para dispositivos móveis que registram as buscas de redes realizadas. O Capítulo 5 apresenta os experimentos e resultados da utilização desse aplicativo.

Tabela 10: Estimativa de embarcados pelo MTR x Embarcados contagem manual; Estimativa de embarque MTR x Embarques contagem manual; Estimativa de desembarques pelo MTR x Desembarques contagem manual.

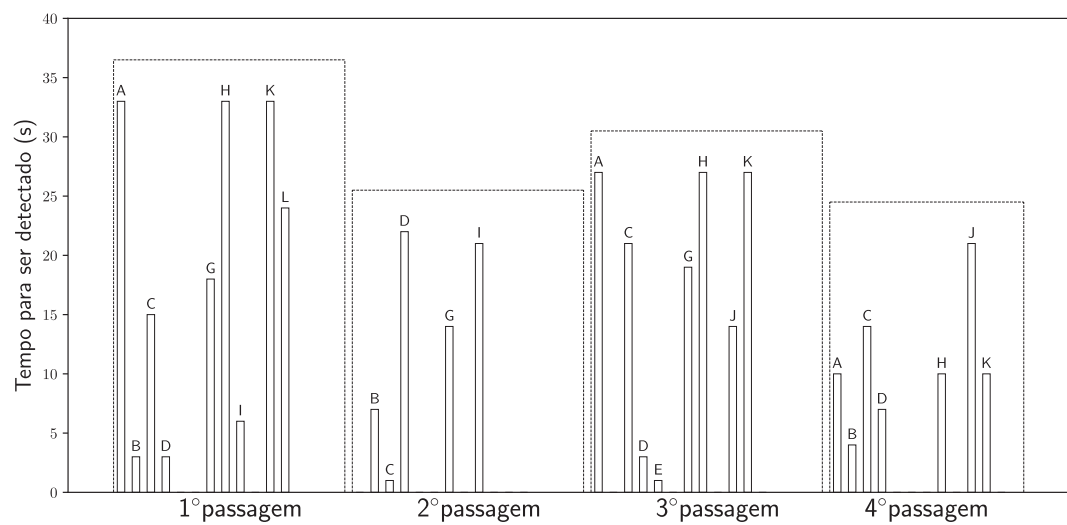
Tipo	Rodada	Trajeto	Contagem		Diferença (%)
			Manual	MTR	
Embarcados (entre trajetos)	1	OA	46	60	30
		AB	57	34	-40
		BC	72	19	-74
	2	OA	50	41	-18
		AB	63	37	-41
		BC	67	26	-61
	3	OA	46	31	-33
		AB	61	34	-44
		BC	74	34	-54
Embarques	1	OA	20	31	55
		AB	15	10	-33
		BC	17	3	-82
	2	OA	34	27	-21
		AB	13	12	-8
		BC	9	3	-68
	3	OA	24	30	-25
		AB	15	7	-53
		BC	17	0	-100
Desembarques	1	OA	12	16	33
		AB	4	2	-50
		BC	2	2	0
	2	OA	12	22	83
		AB	0	5	-
		BC	5	2	-60
	3	OA	15	24	60
		AB	0	3	-
		BC	4	1	-75



(a)

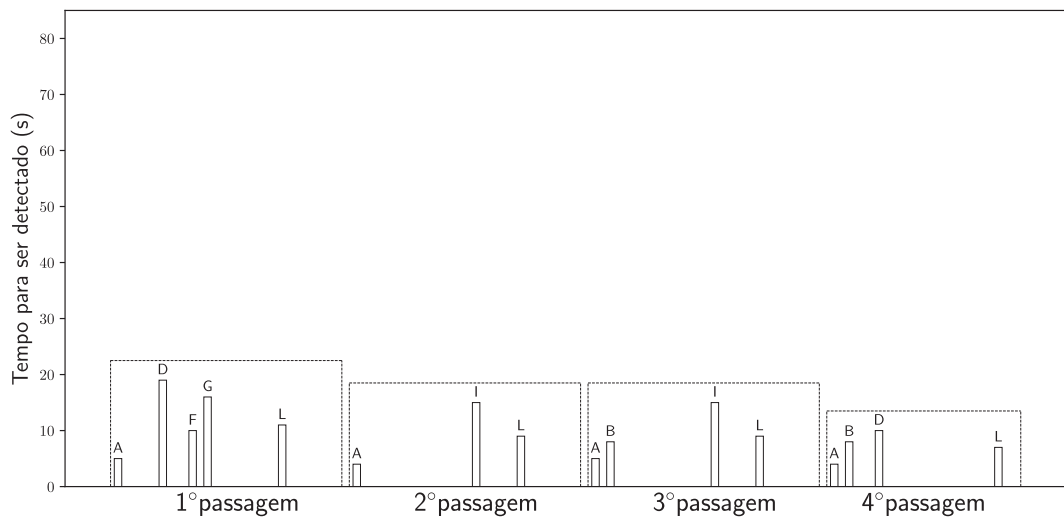


(b)

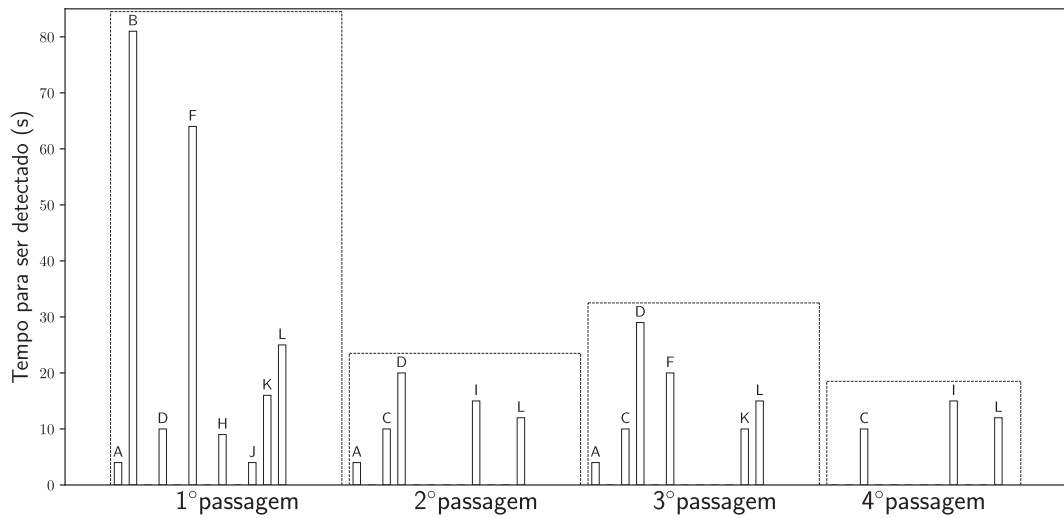


(c)

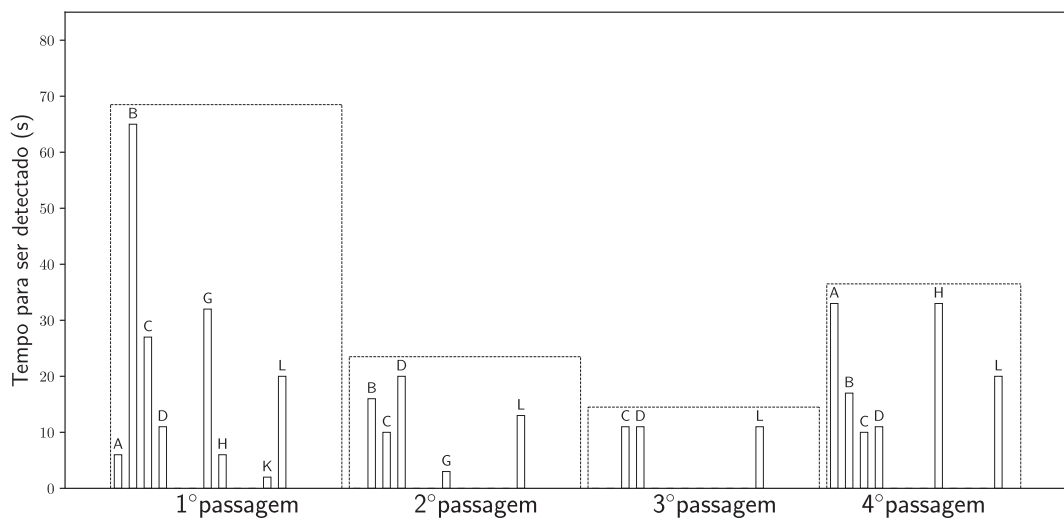
Figura 11: Tempo para a primeira detecção por dispositivo no experimento controlado (0 s significa que o dispositivo não foi detectado): (a) Experimento 1; (b) Experimento 2; (c) Experimento 3.



(a)

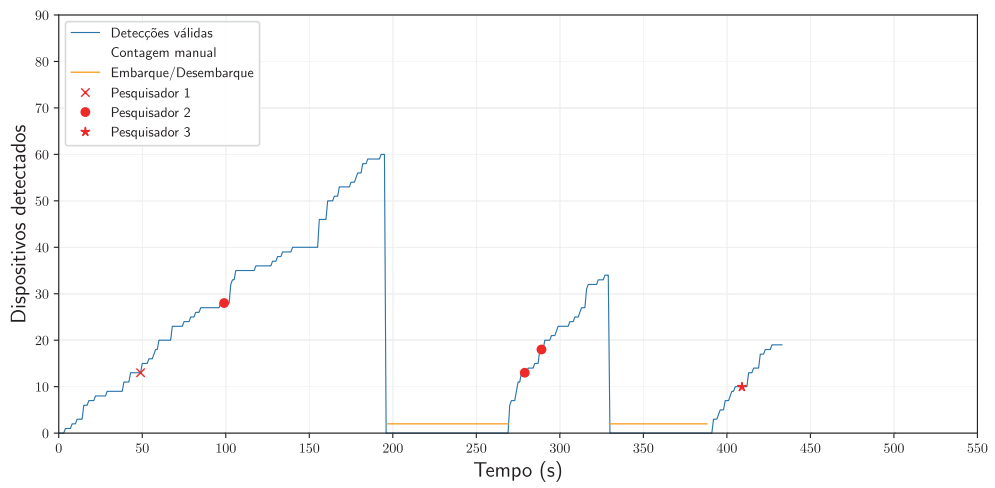


(b)

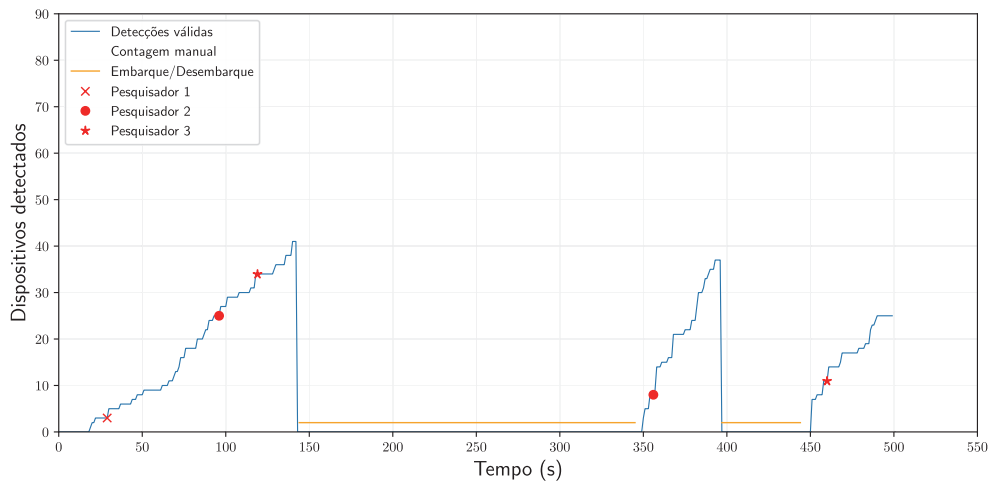


(c)

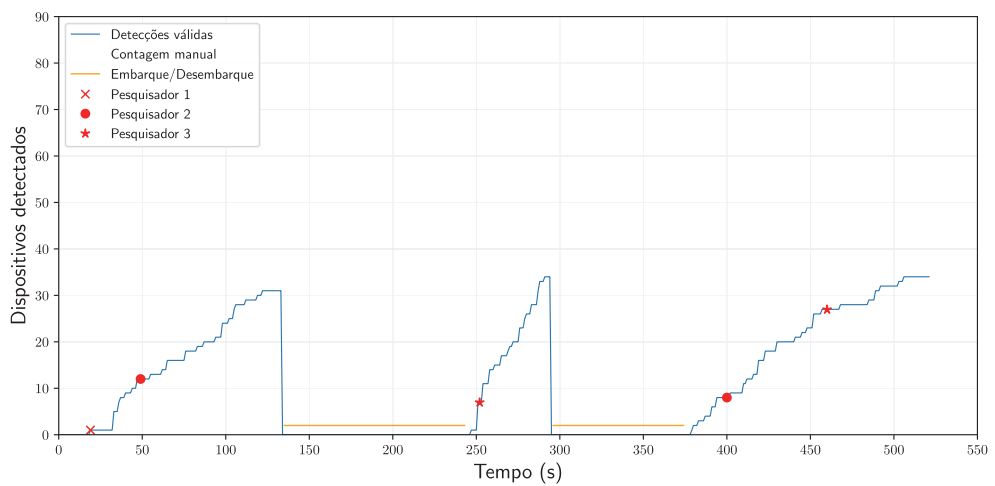
Figura 12: Tempo para a segunda detecção por dispositivo no experimento controlado (0 s significa que o dispositivo não foi detectado): (a) Experimento 1; (b) Experimento 2; (c) Experimento 3.



(a)



(b)



(c)

Figura 14: Contagem de dispositivos móveis acumulados durante o trajeto no experimento com cenário não controlado, com contagem manual de passageiros embarcados: (a) Primeiro rodada, (b) Segunda rodada, e (c) Terceira rodada.

5 APLICAÇÃO PARA REGISTROS DE BUSCAS DE REDES EM DISPOSITIVOS MÓVEIS

Este capítulo apresenta uma aplicação para dispositivos móveis que possui a capacidade de registrar cada busca que um dispositivo móvel realiza em uma rede para tentar uma conexão Wi-Fi. O objetivo da utilização dessa aplicação é confirmar a eficácia do MTR.

Os testes realizados anteriormente com o MTR demonstraram que há possibilidade de detectar dispositivos móveis mesmo sem estarem conectados em uma rede Wi-Fi. A partir dessas detecções foram obtidos dados que fornecem intervalos de tempos entre detecções, permitindo que seja analisado cada dispositivo móvel individualmente.

Durante a análise individual foi observado que o intervalo entre detecções obtidas pelo MTR para um mesmo dispositivo móvel em média apresentou intervalos um tanto elevados (ex. 10 minutos de intervalo entre uma detecção e outra), e com padrões de detecção aleatórios. Além disso em experimentos realizados em campo os dados apresentaram a presença de falsos negativos, ou seja, dispositivos móveis que deveriam ser detectados, mas por algum motivo não foram.

Com esses pontos levantados surge a hipótese de que o problema está na detecção das buscas de rede realizada por dispositivos móveis pelo MTR, ou o problema está na geração dessas buscas de rede realizadas pelos dispositivos móveis, ou o usuário está com o Wi-Fi do dispositivo móvel desligado. Os dois primeiros pontos foram possíveis de medir, já o último não. Mas para o último, ainda é possível se basear que 86% dos usuários mantêm o Wi-Fi dos dispositivos móveis ativado quando estão fora de suas residências (BARBOSA, 2018).

Sabe-se atualmente que no mercado há muitos modelos de dispositivos móveis (Seção 2.3.2) com uma variedade de versões do sistema operacional Android e iOS. Além disso, há inúmeros aplicativos disponíveis nas lojas de aplicativos que para funcionar precisam estar com uma conexão Wi-Fi estabelecida ou realizando uma busca de rede para estabelecer a conexão. Na literatura não há informações concretas com relação ao intervalo de tempo que estes dispositivos móveis realizam buscas de rede. Além dos aplicativos, algumas funções dos dispositivos móveis e interação do usuário, são configurações pré-definidas e gatilhos permitindo que a busca de rede seja realizada.

Quando a interação do usuário para utilizar um aplicativo que necessite estar conectado é acionado um gatilho e realizada uma requisição de rede. Essa conexão permite que o aplicativo atualize ou busque alguma informação requisitada pelo usuário, é nesse momento em que o dispositivo móvel realiza a busca de rede. Outro fator é a própria busca de rede sendo ativada pelo sistema operacional. Essa busca de rede normalmente é realizada quando o dispositivo móvel não está conectado, ou quando há conexão na tentativa de buscar redes com uma intensidade de sinal maior melhorando assim a velocidade da conexão.

Funções do dispositivo móvel pré-configuradas são: a redução da busca de rede para economia de energia do dispositivo móvel; busca de rede Wi-Fi quando a conexão 3G/4G do dispositivo móvel está ativa. A conexão 3G/4G é uma opção de conexão que consome muita bateria do dispositivo móvel, dessa forma o sistema operacional pode buscar redes com uma frequência maior para realizar a troca pela conexão Wi-Fi.

Nas versões obsoletas do sistema operacional Android para dispositivos móveis não havia restrições para realização de buscas de rede. A partir da versão 8.0 surgiram restrições permitindo que aplicativos em segundo plano façam buscas de rede a cada 30 minutos. Já na versão 9.0, os aplicativos em primeiro plano podem realizar quatro buscas de rede a cada 2 minutos, mas no geral todos os aplicativos em segundo plano combinados podem realizar uma busca de rede a cada 30 minutos (GOOGLE, 2018). Para o caso de aplicativos em primeiro plano, dependendo como foi programado, a busca de rede é realizada a cada 30 segundos aproximadamente. Todavia, não se sabe ao certo quanto tempo a aplicação deve ficar em *standby* após realizar as quatro buscas de rede.

Já pelo sistema operacional, as buscas de rede variam entre 60 a 180 segundos. Com base nessas informações foi desenvolvido uma aplicação para dispositivo móvel baseado em Android capaz de registrar as buscas de rede que um dispositivo móvel realiza independente de quem ativou a busca de rede, se foi um aplicativo ou se foi o sistema operacional.

5.1 APLICAÇÃO E FLUXOGRAMA DE EXECUÇÃO

A aplicação desenvolvida utiliza a linguagem Java para programação através do software Android Studio. Todas as funções utilizadas são para realizar a coleta de informações do *status* do dispositivo móvel sem invadir funções que comprometessem a segurança do mesmo.

O algoritmo de execução da aplicação é ilustrado na Figura 15. Na primeira etapa a aplicação identifica se o dispositivo móvel está com o Wi-Fi ligado, essa etapa é um pré-requisito para que a aplicação funcione, a ativação é necessária, mas a conexão não é obrigatória. Caso o Wi-Fi não esteja ativado, fica aguardando até que o usuário ative. Após a verificação da ativação do Wi-Fi é realizada a checagem da versão do sistema operacional. Em versões do sistema operacional Android igual ou acima da 9.0, a Google implementou uma obrigatoriedade de ativar o sinal do GPS do dispositivo móvel para utilizar as funções do monitoramento do Wi-Fi. Para isso, caso a versão sistema operacional seja igual ou maior que a 9.0 a aplicação requisita que o usuário ative o GPS do dispositivo móvel.

Realizada as verificações iniciais, é checado se ocorreu algum erro, ou se há uma requisição por parte do usuário para fechar a aplicação. Caso não ocorra, o processo de análise de buscas de rede começa. A cada 1 ms é realizada uma análise para verificar se foi realizada uma busca de rede, caso tenha sido realizada uma busca de rede as informações são registradas em um arquivo de texto. Ao mesmo tempo que o dado é registrado, já recomeça uma nova análise para verificar se mais buscas estão ocorrendo. O

código desenvolvido para a aplicação pode ser encontrado em Paradedda (2018b)

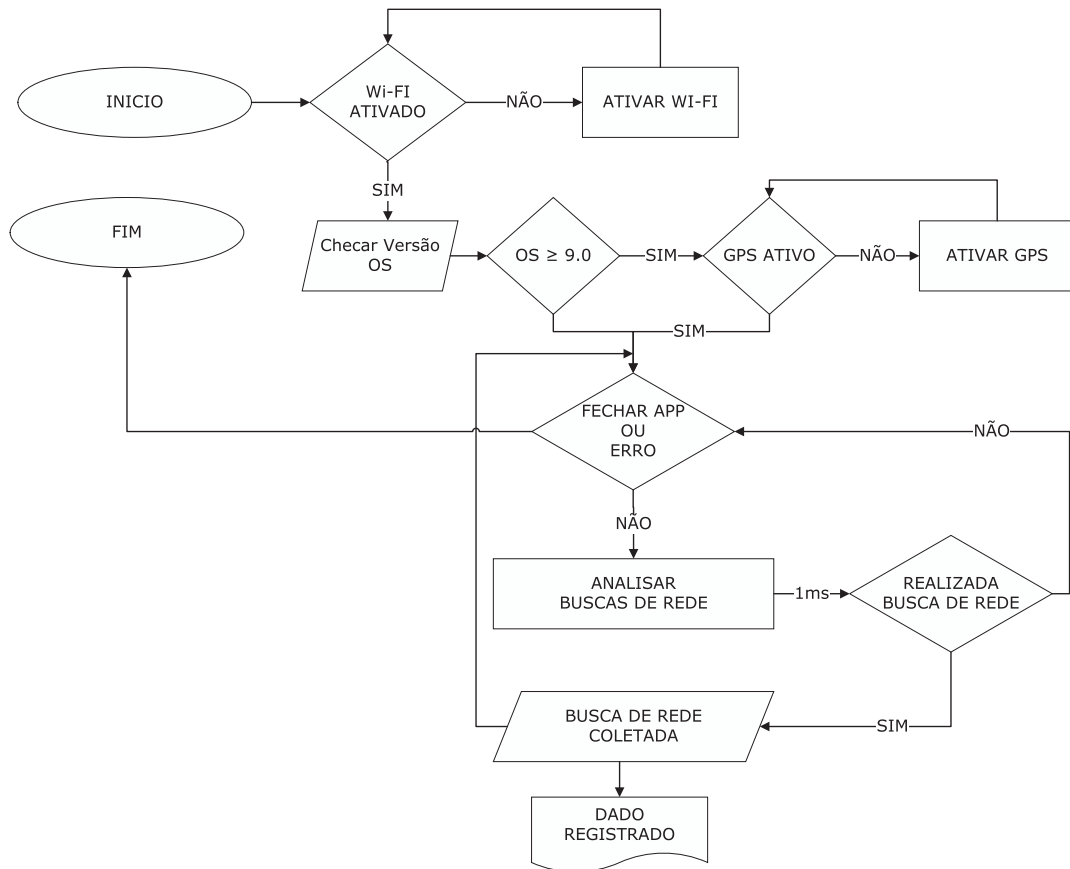


Figura 15: Fluxograma de execução da aplicação de buscas de rede para dispositivos móveis.

A aplicação capturou informações como: status do 3G/4G, nível de bateria, canal de busca, nome da rede, potência do sinal, status Wi-Fi se conectado ou desconectado. Esse tipo informação coletada permite realizar análises do impacto do status 3G/4G na busca de rede Wi-Fi, assim como, se o nível de bateria diminui a taxa com que as buscas de rede são realizadas. No contexto do projeto desenvolvido para esta tese, essas análises não foram realizadas. Somente foram analisadas a quantidade de buscas de rede realizada para aferição do MTR e o intervalo médio de tempo entre as buscas de rede para compreender o comportamento dos dispositivos móveis.

5.2 EXPERIMENTOS COM COLETA DE DADOS PELA APLICAÇÃO

O experimento controlado consistia em utilizar um dispositivo móvel de controle sem nenhum aplicativo instalado com os padrões de fábrica para que as buscas de rede sejam realizadas somente pelo sistema operacional e ao mesmo tempo utilizar um MTR para detectar as buscas de rede realizadas. Foram cruzadas as informações do MTR para identificar se toda busca de rede realizada pelo dispositivo móvel foi detectada pelo MTR, verificando assim a eficácia do MTR.

O experimento não controlado consistia em voluntários instalarem a aplicação em seus dispositivos móveis mantendo suas rotinas normais de uso, tanto conectado como desconectado de redes Wi-Fi. O objetivo desse teste foi buscar dados que possam indicar o intervalo médio em que as buscas de rede ocorrem em dispositivo móveis com diferentes aplicativos instalados. Ao todo seis voluntários se candidataram para realizar os testes. Embora haja dispositivos móveis de mesma marca e modelo no experimento, a maior frequência de buscas de rede ocorrem por buscas realizadas por aplicativos instalados, sendo esses diferentes para cada dispositivo móvel dos voluntários, dessa forma não há problema utilizar dispositivos móveis de mesmo modelo e marca.

5.2.1 RESULTADOS DO EXPERIMENTO CONTROLADO

No primeiro teste com o dispositivo de controle os resultados indicam que há uma perda de menos de 1% das buscas de rede realizadas pelo dispositivo móvel de controle. A Figura 16 ilustra o resultado de 37 horas consecutivas de experimento. O dispositivo de controle não estava conectado em uma rede Wi-Fi e somente realizava as buscas de rede através das requisições do sistema operacional. Ao todo foram realizadas 989 buscas de rede pelo dispositivo móvel de controle das quais 979 foram detectadas pelo MTR.

Na Figura 16 é ilustrado o histograma das buscas de redes que ocorreram e foram registradas pela aplicação.

A Tabela 11 representa um resumo dos dados. Nessa tabela são apresentados o número das buscas de rede que foram perdidas e o intervalo de tempo desde a busca de rede anterior. Por exemplo, a busca de rede S_2 ocorreu 11 segundos após ter ocorrido o S_1 , onde a busca de rede S_1 foi detectado pelo MTR mas a S_2 não.

Tabela 11: Identificação da busca de rede perdida pelo MTR e o intervalo de tempo em relação a busca de rede anterior

Busca de rede	Intervalo
S_2	11
S_3	12
S_{61}	125
S_{62}	4
S_{63}	12
S_{168}	133
S_{378}	118
S_{825}	129
S_{828}	195
S_{990}	193

Estes resultados indicam que o MTR é capaz de detectar dispositivos móveis sem perdas significativas de buscas de redes.

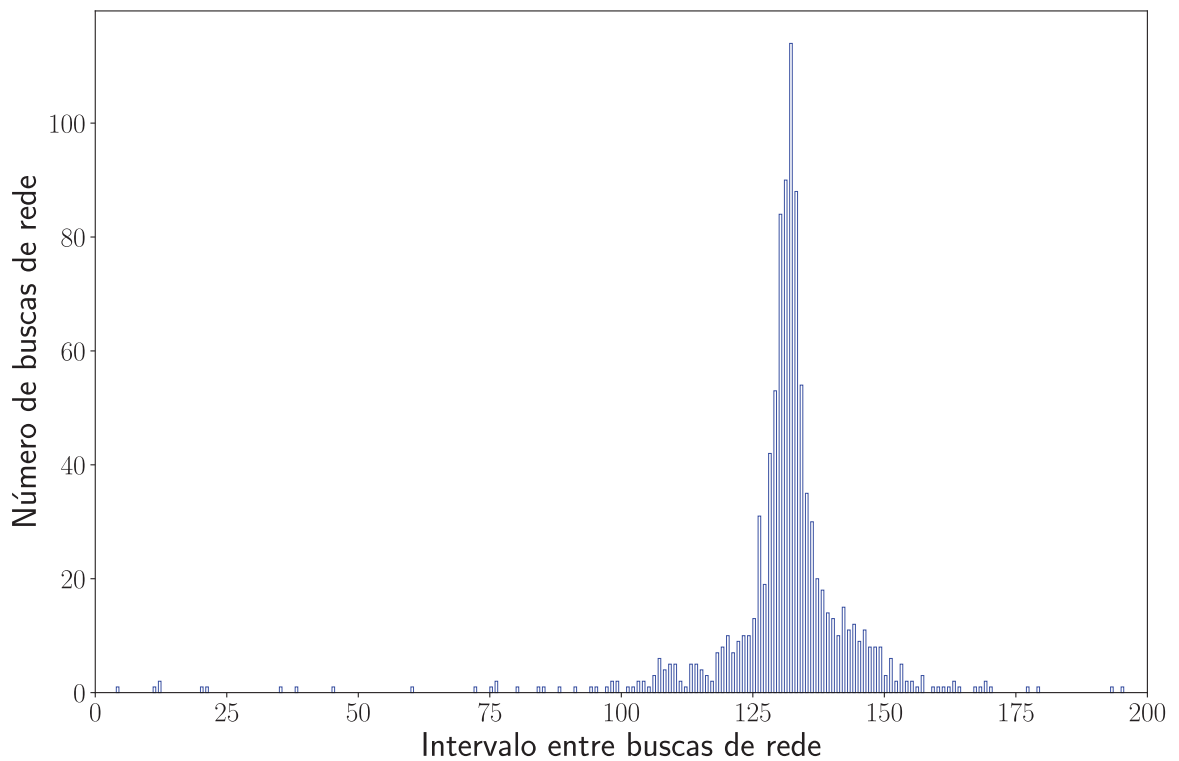


Figura 16: Número de buscas realizadas pelo dispositivo móvel de controle registradas pela aplicação.

5.2.2 RESULTADOS DO EXPERIMENTO NÃO CONTROLADO

Para o segundo experimento, a aplicação de buscas de rede foi distribuída para os seis voluntários. Após três dias de experimentos, foram analisados os intervalos entre as buscas de rede de dispositivos móveis coletados com a aplicação.

Durante a execução desse experimento os voluntários deveriam manter a aplicação de registro de buscas de rede sempre aberto, mas muitas vezes os voluntários acabavam fechando a aplicação involuntariamente, sendo necessário abrir novamente. Em outros momentos a bateria dos dispositivos móveis acabavam e era necessário recarregar os mesmos. Isso faz com que a coleta de dados não seja constante. Embora esses problemas tenham ocorrido, os três dias de experimento foram suficientes para obter dados para análise.

Após a realização dos experimentos e processamento dos dados, foram obtidas informações para um período de 24 horas de seis dispositivos de voluntários diferentes. A Tabela 12 indica que há diferenças entre as buscas de redes realizadas pelos dispositivos móveis dos voluntários. Essa variação ocorre devido aos aplicativos instalados em cada dispositivo móvel. É possível observar também que, mesmo dispositivos móveis da mesma marca a diferença de busca de rede também ocorre, novamente é devido aos aplicativos instalados em cada dispositivo móvel.

Tabela 12: Número de buscas de rede realizadas por dispositivos durante um período de 24 horas.

Vendor	OS Versão	Buscas de rede	Buscas de rede/h
Huawei	9.0	280	12
Motorola	9.0	187	8
Samsung	8.0	217	9
Samsung	8.0	167	7
Samsung	8.0	228	10
Xiaomi	9.0	211	9

5.2.3 EFICÁCIA DO NTM E REGULARIDADE NAS BUSCAS DE REDES

O desenvolvimento do aplicativo para coletar buscas de redes realizadas pelos dispositivos móveis permitiu que fosse possível verificar a eficácia do NTM. Com os experimentos realizados foi possível identificar que o NTM perde aproximadamente 1% das buscas de redes que são realizadas pelos dispositivos móveis. Isso permite que possamos afirmar que se um dispositivo móvel realizar buscas de rede, 99% dessas buscas serão detectadas.

Outro ponto importante é a verificação da regularidade com que os dispositivos móveis realizam buscas de redes. No experimento controlado há a indicação que o dispositivo de controle realiza em média uma busca de rede a cada 130 segundos. Mas esse comportamento é de um dispositivo com padrões de fábrica, sem aplicativos gatilhos para acionar uma nova busca de rede. Na comparação do dispositivo de controle com os dispositivos do experimento não controlado, o número de buscas de rede dos dispositivos do experimento não controlado é menor. Esse fato acaba chamando a atenção e indicando que a frequência com que as buscas de redes ocorrem precisa ser mais aprofundada como será visto no Capítulo 6.

6 ANÁLISE DO COMPORTAMENTO DE DISPOSITIVOS MÓVEIS

Neste capítulo é apresentado a forma de se obter uma distribuição que represente o comportamento médio das buscas de redes realizadas por dispositivos móveis. Essa distribuição será chamada de distribuição de probabilidade do intervalo de detecções. Os dados utilizados foram obtidos através dos experimentos realizados no Capítulo 5.

6.1 AJUSTE DO INTERVALO DE BUSCAS DE REDE PARA COMPATIBILIZAÇÃO DE DETECÇÃO COM COLETA A BORDO

Os dados coletados no experimento do Capítulo 5 representam informações com muitas horas de coleta de um mesmo dispositivo móvel, o que não representa o comportamento real de um único usuário quando o mesmo estiver dentro de um ônibus. Como exemplo, os tempos coletados de um único usuário no experimento realizado no Capítulo 4.3 não passa de 15 minutos pelo tempo de trajeto, e para o experimento realizado por Nunes (2018) não passa de 45 minutos pelo tempo da viagem.

Com o objetivo de ajustar os dados para simular que os dispositivos móveis dos voluntários estavam realizando viagens de ônibus, os dados foram agrupados em intervalos de 30 minutos.

Para realizar esse agrupamento, foram selecionados períodos de 30 minutos onde todos os voluntários estivessem com seus dispositivos móveis ligados e com a aplicação funcionando. Dessa forma garantiu-se uma situação similar ao caso em que todos estivessem “embarcados” ao mesmo tempo no mesmo ônibus. Ao todo foram considerados sete intervalos de 30 minutos, nesses intervalos há registros de buscas de redes realizadas por todos os dispositivos móveis dos voluntários.

Cada intervalo corresponde a uma viagem realizada pelos voluntários, o que acaba fornecendo um padrão de dados similar ao do experimento anterior (Capítulo 4.3 e Nunes (2018)), só que nesse caso não se considera paradas ao longo do trajeto, nem embarques e desembarques. Supõe-se que somente os seis voluntários estão embarcados e vão desembarcar juntos ao mesmo tempo.

As buscas de redes realizadas dentro desses intervalos agora são chamadas de detecções realizadas pelo MTR, já que anteriormente foi garantido que menos 1% das buscas de rede realizadas pelos dispositivos móveis são perdidas pelo MTR.

A Figura 17(a) representa o intervalo entre as detecções pelo número de dispositivos distintos detectados. Com ajuda da Figura 17(b), com a representação da taxa pelo número de dispositivos, é possível observar que aproximadamente 80% das detecções ocorrem em até 350 segundos. Esse comportamento indica que há um problema que será abordado no Capítulo 7 desta tese, em que em pontos de paradas de ônibus muito próximos essa porcentagem da taxa do número de buscas de rede tende a cair.

Com esses dados, é obtida a curva de distribuição da taxa do número de detecção

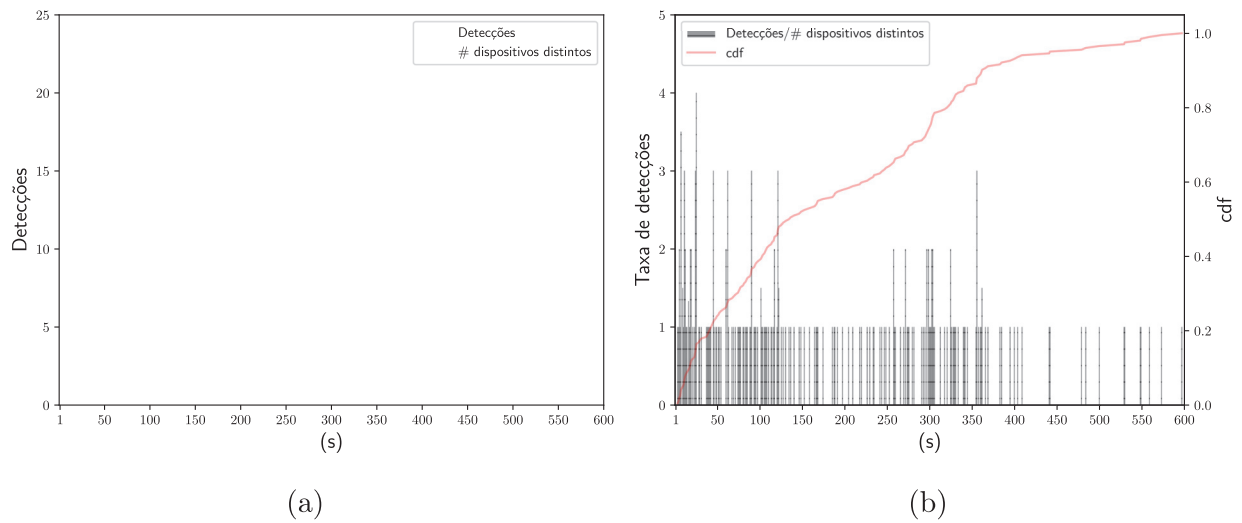


Figura 17: (a) Buscas de rede realizadas pelos seis voluntários com dados agrupados em 30 minutos para sete períodos. (b) Distribuição de probabilidade do intervalo de detecções.

em razão do número de dispositivos móveis detectados (conhecida como distribuição de probabilidade do intervalo de detecções) ao longo de sete viagens de ônibus de uma linha hipotética. A representação dessa distribuição é uma estimativa do comportamento médio das buscas de rede dos dispositivos móveis, sendo assim podendo afirmar que, para o conjunto de dados analisado, as buscas realizadas pelos dispositivos móveis tendem a se comportar de acordo com a Figura 17(b).

6.2 EVENTO BUSCA DE REDE X_i

Como já visto ao longo dessa tese, dispositivos móveis realizam buscas de rede quando o Wi-Fi está ativo. Nessa tese esse evento foi denominado como uma variável aleatória chamada de X_i . É importante observar que essa variável não é a busca de rede detectada pelo NTM, e sim a busca de rede realizada pelo dispositivo móvel. Com essa atribuição, podemos dizer que as distribuições obtidas através dos experimentos dos Capítulos 4.3 e 5, representam o comportamento dos dispositivos móveis, e assim representam o comportamento da variável aleatória X_i .

Mas é possível associar essa variável a detecção realizada pelo NTM. Isso porque só pode haver uma detecção se houver uma busca de rede. Com essa premissa em mente, podemos utilizar os dados de detecções de buscas de redes coletados pelo NTM dos experimentos realizados nessa tese e de outras fontes, e obter o comportamento dos dispositivos móveis.

6.3 DISTRIBUIÇÃO DE PROBABILIDADE DE BUSCAS DE REDE

Como não há possibilidade de instalar o aplicativo de monitoramento de buscas de redes em cada dispositivo móvel que irá embarcar em um ônibus, a única informação que

será disponibilizada é a busca de rede detectada pelo NTM. Mas como vimos anteriormente essa busca de rede detectada está associada há uma busca de rede realizada por um dispositivo móvel.

Com isso, a base de dados obtidas no Capítulo 4.3 pode ser convertida em buscas de redes realizadas por dispositivos móveis, representada pela variável aleatória X_i . A Figura 18(a) ilustra o número de detecções obtidas pelo NTM no experimento do Capítulo 4.3 e o número de dispositivos distintos que foram detectados. Por outro lado a Figura 18(b) ilustra o comportamento médio dessas buscas de redes detectadas. Podemos dizer que o comportamento médio da variável aleatória X_i pode ser representado pela Figura 18(b).

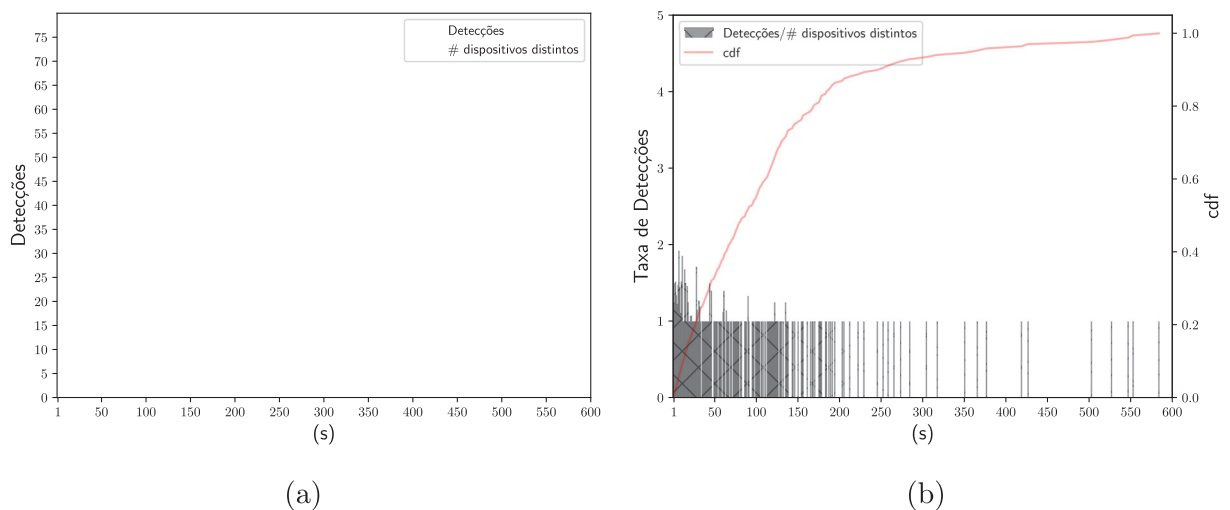


Figura 18: (a) Número de detecções de buscas de rede com dispositivos distintos registrados no experimento Embarcado (Capítulo 4.3); (b) média do intervalo de buscas de rede.

Obtendo fontes de dados que não foram coletadas para essa tese, mas foram coletadas de outras fontes, também é possível fazer as mesmas considerações. Para isso foram considerados os dados brutos disponibilizados por Nunes (2018), da sua pesquisa de dissertação de mestrado. A Figura 19(a) ilustra o número de detecções obtidas pelo detector desenvolvido pelo autor e o número de dispositivos distintos que foram detectados. Utilizando a mesma análise anterior a Figura 19(b) ilustra o comportamento médio dessas buscas de redes detectadas e o comportamento médio da variável aleatória X_i pode ser representado pela Figura 19(b)

Com conhecimento do comportamento das buscas de redes realizadas por dispositivos móveis, e a obtenção das distribuições de probabilidade, é possível aplicar o modelo de estimativas de embarques do capítulo a seguir nessas distribuições.

6.3.1 UTILIZAÇÃO DE DADOS DE COLETA WI-FI

Embora o comportamento de dispositivos móveis seja um tanto quanto irregular, há possibilidade de trabalhar com os dados coletados.

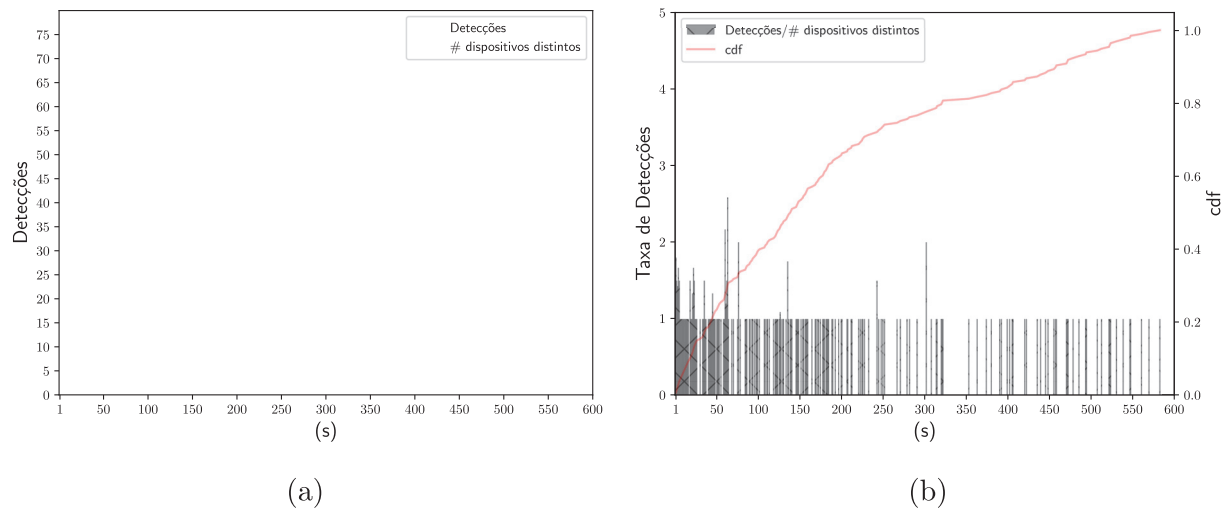


Figura 19: (a) Número de detecções de buscas de rede com dispositivos distintos registrados no experimento realizado por Nunes (2018); (b) média do intervalo de buscas de rede.

Mesmo com o comportamento irregular das buscas de redes realizadas por dispositivos móveis, é possível obter uma distribuição de probabilidade que possa indicar o comportamento dos dispositivos. Como é possível isolar cada dispositivo móvel da amostra, o ideal é que ao longo de um percurso de uma viagem de ônibus, seja obtido a distribuição de probabilidade individual de cada dispositivo móvel. Pensando em longo prazo, cada dispositivo poderia ter sua distribuição ao longo de um ano de viagens por exemplo. A grande vantagem apresentada aqui é a inserção de uma variável aleatória que representa esse comportamento.

Também a possibilidade de obtenção de dados de outras fontes, permite que seja ampliado o campo de estudo. Desde que o dado seja buscas de rede, é possível obter uma distribuição de probabilidade a partir dessa fonte. Como foi apresentado aqui, há fontes de dados advindas da própria busca de rede, mas também fontes de outros detectores que coletaram dados de buscas de redes em outros experimentos, experimentos esses não realizados por esse autor.

A utilização de outros dados permite verificar a robustez do método proposto, porque muitas vezes os dados coletados para a pesquisa podem estar "viciados" e podem levar aos resultados que não correspondem a uma realidade. Já com a utilização de dados de outras fontes, o método é realmente testado e demonstra sua eficácia.

7 ESTIMATIVA DE EMBARQUES DE PASSAGEIRO UTILIZANDO DADOS DE COLETA WI-FI E MÉTODOS ESTATÍSTICOS

Até o momento a etapa de realizar a análise do número de detecções e potência do sinal não é utilizada para a estimativa de embarques de passageiros. Ela é utilizada para validar se um dispositivo móvel se encaixa nos parâmetros definidos anteriormente (Capítulo 4.3) para um possível embarque ou para determinar se um dispositivo móvel realmente estava embarcado. Este capítulo apresenta uma abordagem para estimar onde ocorreu o embarque desse dispositivo móvel. Para isso foi apresentado um novo dado que estudos de outros autores não apresentaram. Esse dado é a distribuição da taxa do intervalo de buscas de rede de dispositivos móveis, isso é, o comportamento médio das buscas de redes realizadas por dispositivos móveis, e será utilizado para estimar onde ocorreu o embarque de um dispositivo móvel.

Sabendo-se que o dispositivo MTR é capaz de detectar as buscas de rede realizadas por dispositivos móveis com 99% de eficiência (Capítulo 5), foi desenvolvido um método estatístico Bayesiano para estimar os embarques de passageiros em um ponto de ônibus dada uma distribuição da taxa do intervalo de buscas de rede de cada dispositivo individualmente.

7.1 MODELO DE EMBARQUES DESAGREGADOS POR PONTO DE ÔNIBUS

Trabalhar com dados desagregados ao longo do trabalho mostrou-se o caminho mais adequado para buscar o comportamento mais realista do método proposto. Quando o passageiro embarca em um ônibus e está portando um dispositivo com o Wi-Fi ligado, há uma probabilidade que ocorra o evento de realizar uma busca de rede. Esse evento é chamado de S_i , sendo a primeira busca de rede realizada pelo dispositivo móvel i que foi detectado pelo MTR dentro do ônibus.

Essa primeira detecção só é considerada válida, caso ao longo do trajeto do ônibus ocorram mais buscas de rede detectadas do mesmo dispositivo, confirmando assim que essas buscas vieram de um passageiro embarcado. Isso garante que detecções de dispositivos externos não sejam consideradas.

O objetivo é saber se o passageiro portando um dispositivo móvel i embarcou no ponto k ou $k - 1$ até o primeiro ponto do itinerário. Como visto no Capítulo 5, o intervalo entre buscas de redes consecutivas de um dispositivo Wi-Fi pode ser descrito como a variável aleatória da função de densidade de probabilidade D_i (Dispositivo móvel i). Assim é natural associar o evento da busca de rede detectada pelo MTR S_i com o evento da busca de rede X_i . Mas a consideração mais importante dessa variável, é que ela nos fornece uma informação que não tínhamos, sobre a busca de rede ter ocorrido antes do embarque onde não havia formas de ser detectada. Essa informação é importante porque ela indica a probabilidade da próxima detecção ocorrer.

7.1.1 ESTIMADOR BAYESIANO PARA EMBARQUES EM PONTO

Seja B_{il} o evento em que o passageiro portando o dispositivo móvel i embarcou no ponto l . Supondo que i foi detectado e considerado a bordo do ônibus após o ponto de ônibus k e antes do ponto $k + 1$, como ilustrado na Figura 20, obtemos o a primeira busca de rede detectada do dispositivo móvel i após o ponto k , anteriormente denominado como evento S_i .

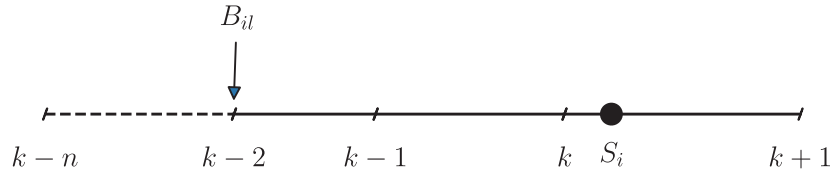


Figura 20: Passageiro embarcando no ponto l onde na representação sequencial é o ponto de ônibus $k - 2$ e sendo detectado após o ponto k .

Se só temos essa informação, então podemos assumir que a probabilidade de embarque em qualquer ponto anterior a k possui a mesma probabilidade $P(B_{il})$ da posição onde foi validado a primeira detecção pelo MTR dentro do ônibus, já que a probabilidade de embarque é igual em cada ponto. Então, a probabilidade *a priori* de que o passageiro portando o dispositivo móvel i embarcado no ponto de ônibus l , quando detectado pelo MTR a bordo entre os pontos k e $k + 1$ é dada por:

$$P(B_{il}) = \frac{1}{k}, \quad l = 1, 2, \dots, k, \quad (2)$$

A partir do momento da primeira detecção pelo MTR e com a informação prévia da distribuição da taxa do intervalo de buscas de rede X_i , que é conhecida, é possível de acordo a regra de Bayes determinar a probabilidade *a posteriori*, de que um passageiro portando um dispositivo móvel i embarcou em l e foi detectado pelo MTR a bordo após k e antes de $k + 1$. Esta probabilidade é dada por:

$$P(B_{il}|S_i) = \frac{P(B_{il})P(S_i|B_{il})}{P(S_i)}. \quad (3)$$

A probabilidade condicional $P(S_i|B_{il})$ define que a probabilidade da primeira busca de rede detectada pelo MTR de um dispositivo móvel que embarcado ocorreu após o embarque no ponto l e $P(S_i)$ a probabilidade de que a busca de rede a bordo do ônibus do dispositivo móvel i ocorreu após o ponto k . Com o objetivo de encontrar a probabilidade $P(S_i|B_{il})$, o horário relógio do ônibus partindo do ponto l , denotado por t_l , e a detecção do passageiro i em t_{ik} é dada por:

$$\Delta t_{il} = t_{ik} - t_l, \quad (4)$$

A Figura 21 ilustra um exemplo com os instantes de Δt_{il} para $k = 3$, e a Tabela 13 apresenta os dados obtidos. Nesta tabela, a probabilidade de embarque $P(B_{il})$ foi obtida através da equação 2, e é igual para todos os pontos de embarque anteriores ao momento da detecção.

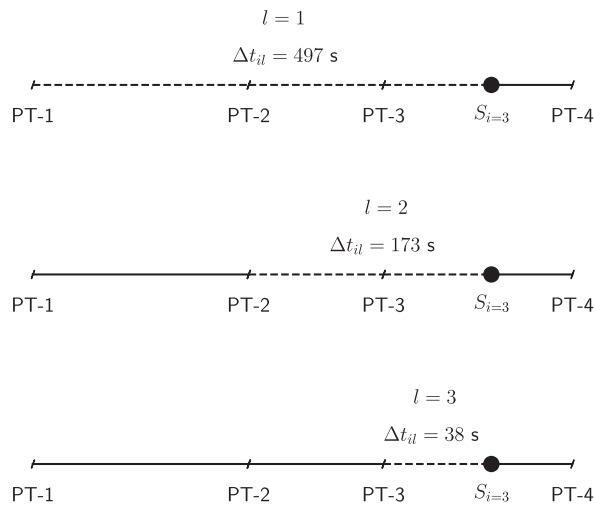


Figura 21: Exemplo de um dispositivo móvel $i = 3$, detectado após o ponto de embarque PT-3. As informações de distância entre os pontos são conhecidas. Sendo Δt_{il} , 497 s (ponto PT - 1 até a detecção), 173 s (ponto PT - 2), 38 s (ponto PT - 3).

Tabela 13: Tabela com resumo dos dados obtidos na Figura 21

Distribuição	i	k	l	Δt_{il}	$P(B_{il})$
Embarcado	3	3	1	497	0.333
			2	173	0.333
			3	38	0.333

Lembrando que S_i é a primeira busca de rede embarcado de i . Assumindo a condição de que o embarque ocorreu em l e que a primeira busca de rede detectada S_i ocorreu entre k e $k + 1$, a probabilidade desejada é equivalente à probabilidade de que a busca de rede anterior ocorreu antes do ponto l enquanto o passageiro i estava fora do veículo. Essa observação estabelece uma conexão direta entre Δt_{il} e X_i expressa por:

$$P(S_i|B_{il}) = P(X_i \geq \Delta t_{il}). \quad (5)$$

A Figura 22 ilustra as probabilidades desejadas, e a distribuição de probabilidade considerada é a de nome Embarcado. A área mais escura na pdf é $P(X_i \geq \Delta t_{il})$, isso é, a probabilidade de que intervalos consecutivos de buscas de redes ocorreram em um tempo maior ou igual ao tempo em que o passageiro portando o dispositivo móvel i embarcou no ponto l .

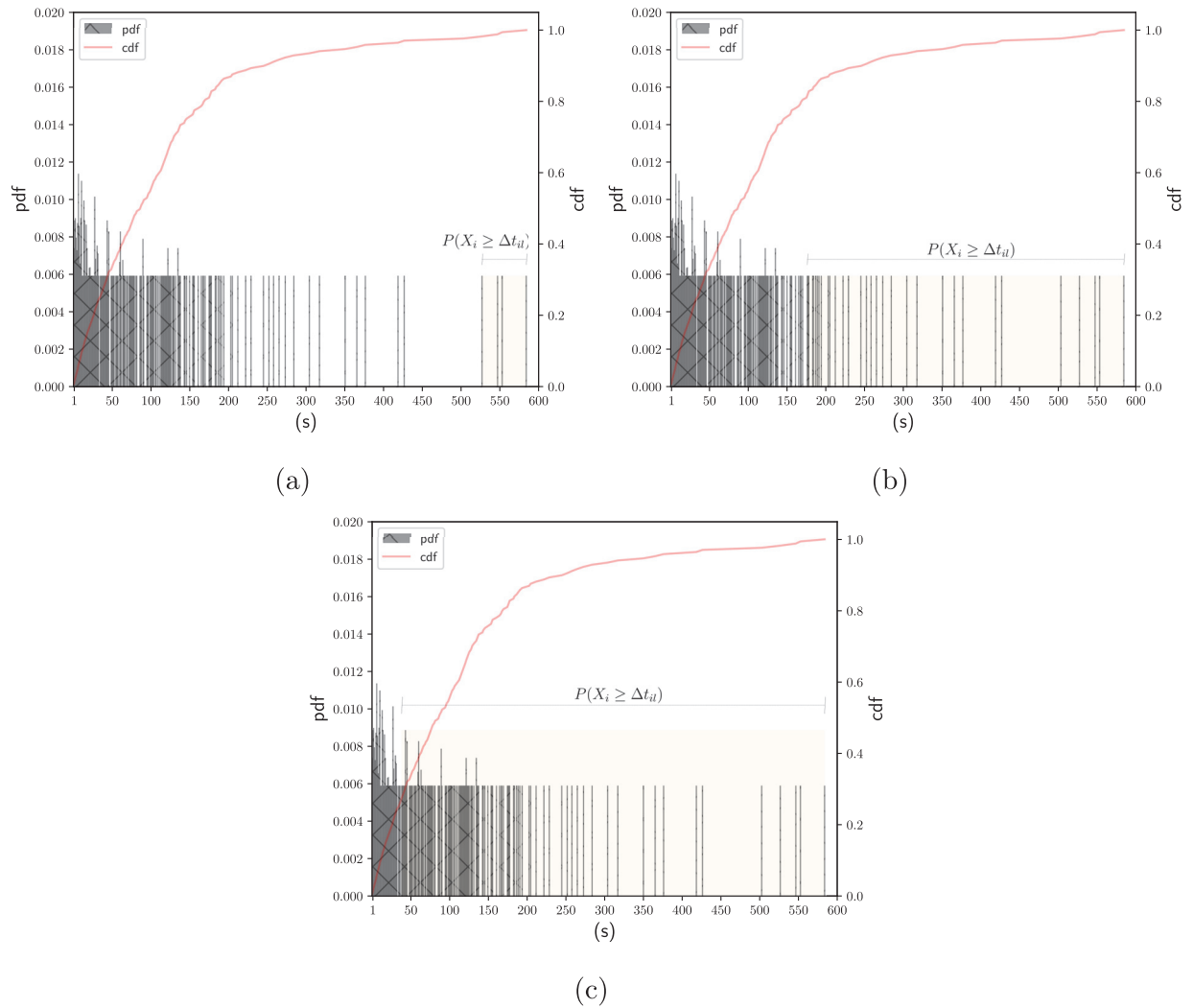


Figura 22: Probabilidade desejadas da distribuição Embarcado considerada como estudo de caso. (a) Probabilidade para o intervalo onde $l = 1$. (b) para $l = 2$. (c) para $l = 3$

Note que para $\Delta t_{il} \rightarrow 0$, $P(S_i|B_{il}) \rightarrow 1$. Em outras palavras, se a primeira busca de rede detectada pelo MTR ocorreu imediatamente após o embarque em l , então a busca de rede anterior quase certamente ocorreu fora do veículo. Tal conclusão resulta da premissa (condicional) de que o embarque de fato ocorreu em l . A Tabela 14 ilustra os resultados obtidos a partir da distribuição representada pela Figura 22.

Tabela 14: Tabela com a probabilidade de que a detecção S_i tenha ocorrido, dado que o evento B_{il} em que um passageiro portando um dispositivo móvel embarcou no exemplo da Figura 21.

Distribuição	i	k	l	Δt_{il}	$P(B_{il})$	$P(S_{ik} B_{il})$
Embarcado	3	3	1	497	0,333	0,029
			2	173	0,333	0,195
			3	38	0,333	0,736

A aplicação da Equação 6 implica em obter $P(S_i)$, ou a probabilidade de que a primeira busca de rede do dispositivo móvel embarcado tenha ocorrido após o ponto k .

Essa probabilidade pode ser calculada a partir da probabilidade total, dada por:

$$P(S_i) = \sum_{l=1}^k P(B_{il})P(S_i|B_{il}). \quad (6)$$

A Equação 3 resulta na distribuição de probabilidade de um passageiro ter embarcado em todos os pontos de ônibus anteriores possíveis a partir da primeira detecção, e deve obedecer a seguinte equação:

$$\sum_{l=1}^k P(B_{il}|S_i) = 1. \quad (7)$$

Aplicando a Equação 6 da probabilidade total na Equação de Bayes (3), obtém-se:

$$P(B_{il}|S_i) = \frac{P(B_{il})P(S_i|B_{il})}{\sum_{l=1}^k P(B_{il})P(S_i|B_{il})}, \quad l = 1, \dots, k. \quad (8)$$

Lembrando que as probabilidades *a priori* $P(B_{il})$ são obtidas através da Equação 2, levando a

$$P(B_{il}|S_i) = \frac{P(S_i|B_{il})}{\sum_{l=1}^k P(S_i|B_{il})}, \quad l = 1, \dots, k. \quad (9)$$

Aplicando o somatório para $l = 1, \dots, k$ nos dois lados da Equação 9 onde o resultado obtido é igual 1 tem-se

$$\sum_{l=1}^k P(B_{il}|S_i) = \frac{\sum_{l=1}^k P(S_i|B_{il})}{\sum_{l=1}^k P(S_i|B_{il})} = 1.$$

Aplicando a equação 6 no estudo de caso da Figura 21, onde o dispositivo móvel 3 foi detectado após o ponto de embarque 3, obtém-se os resultados da Tabela 15.

Tabela 15: Tabela com o exemplo do cálculo da probabilidade total para o estudo de caso.

Distribuição	i	k	l	Δt_{il}	$P(B_{il})$	$P(S_i B_{il})$	$P(B_{il})P(S_i B_{il})$	$P(S_i) = \sum_{l=1}^k P(B_{il})P(S_i B_{il})$
Embarcado	3	3	1	497	0,333	0,029	0,010	0,32
			2	173	0,333	0,195	0,065	
			3	38	0,333	0,736	0,245	

Com todos os dados obtidos na Tabela 15 é possível aplicar a Equação 9 para o exemplo. Os resultados obtidos estão representados na Tabela 16.

Se não fosse possível afirmar que o embarque ocorreu com 100% de certeza no ponto de embarque 3 como afirmado no exemplo do estudo de caso, então a partir do resultado de $P(B_{il}|S_i)$, podemos indicar que há probabilidade de 76,6% de que o passageiro portando o dispositivo móvel 3 tenha embarcado no ponto de embarque 3, dado que sua

Tabela 16: Tabela com o cálculo da Equação 9, onde $\sum_{l=1}^k P(B_{il}|S_i) = 1$.

Distribuição	i	k	l	$P(B_{il} S_i) = \frac{P(S_i B_{il})}{\sum_{l=1}^k P(S_i B_{il})}, l = 1, \dots, k.$
			1	0,031
Embarcado	3	3	2	0,203
			3	0,766

detecção ocorreu após o ponto 3. Da mesma forma há uma probabilidade de 20,3% de que o embarque tenha ocorrido no ponto de embarque 2 e 3,1% de que tenha ocorrido no ponto de embarque 1.

Essa é uma aproximação para somente um passageiro de uma única distribuição. Para realizar o cálculo da probabilidade de embarque de todos os passageiros, é necessário calcular a probabilidade de embarque dado que detecções ocorreram para todos os dispositivo móveis gerados.

7.1.2 ESTIMATIVA DO TOTAL DE EMBARQUES EM PONTOS DE PARADA DE ÔNIBUS

Quando está disponível a informação das probabilidades $P(B_{il}|S_i)$ onde um passageiro portando um dispositivo móvel i embarcou em um ponto l dado que sua detecção pelo MTR ocorreu após o ponto k , a estimativa de embarques em cada ponto pode ser determinado. Para simplificar a notação, considera-se

$$P(B_{ik}) \triangleq P(B_{il}|S_i).$$

Assim, deseja-se saber quantos passageiros embarcaram em cada ponto de parada de ônibus k no trajeto da Figura 27 para cada uma das três distribuições. Seja \mathcal{K} o conjunto de passageiros com $P(B_{ik}) \neq 0$. A tupla contendo as probabilidades de embarque de passageiros no ponto k é escrita como:

$$B_k = \langle P(B_{ik}), i \in \mathcal{K} \rangle.$$

Supondo o exemplo da Figura 23, onde é ilustrado os passageiros que possuem probabilidade de embarcar no ponto de parada de ônibus PT - 3. É possível observar que há 3 passageiros que foram detectados após o ponto PT - 4. Há possibilidade de que esses passageiros possam ter embarcado no ponto PT - 3, isso ocorre por que o tempo de deslocamento do ponto PT - 3 até o ponto PT - 4 é de 77 s. Sabendo que a detecção poderia ocorrer em até duas vezes a média da distribuição, que nesse caso é 244,26 s, o tempo de deslocamento faz com que as detecções possam ocorrer após o ponto PT - 4, mas não após o ponto PT - 5, por que o tempo de deslocamento do ponto PT - 3 até o ponto PT - 5 é 478 s.

Para o exemplo foi obtido:

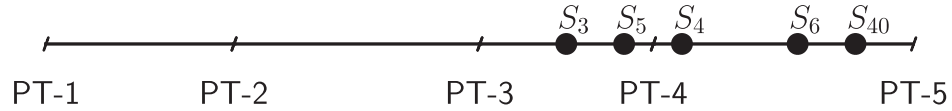


Figura 23: Passageiros que possivelmente embarcaram no ponto 3.

$$\mathcal{K} = [3, 4, 5, 6, 40]$$

$$B_3 = \langle P(B_{3,3}), P(B_{4,3}), P(B_{5,3}), P(B_{6,3}), P(B_{40,3}) \rangle$$

$$B_3 = \langle 0.766, 0.318, 0.796, 0.283, 0.302 \rangle$$

O total de embarques T_k em k é uma variável aleatória com distribuição de probabilidade dada por:

$$T_k = \langle P(j) \rangle, j = 0, \dots, N$$

em que N é a cardinalidade de \mathcal{K} e $P(j)$ é a probabilidade de j embarques em k . Note que $\sum_{j=0}^N P(j) = 1$.

Encontrar as probabilidades de T_k requer o cálculo de todas as combinações de embarques utilizando a probabilidade individual em B_k . Por exemplo, para o caso de $P(j=0)$ é a probabilidade de que todo passageiro $i \in \mathcal{K}$ não embarcou em k .

Como cada embarque é individual e um não depende do outro

$$P(0) = \prod_{i \in \mathcal{K}} P(\overline{B_{ik}}) = \prod_{i \in \mathcal{K}} [1 - P(B_{ik})].$$

Da mesma forma, para $j=1$, o cálculo é a soma de todas as combinações de que o passageiro i embarcou enquanto os outros não embarcaram, e é dada por:

$$\begin{aligned} P(1) &= \sum_{i \in \mathcal{K}} P(B_{ik}) \prod_{\substack{l \in \mathcal{K} \\ l \neq i}} P(\overline{B_{lk}}) \\ &= \sum_{i \in \mathcal{K}} P(B_{ik}) \prod_{\substack{l \in \mathcal{K} \\ l \neq i}} [1 - P(B_{lk})], \end{aligned}$$

No geral, para N passageiros e j embarques, o número de combinações é $\binom{N}{j}$. É inviável calcular essa combinação para um ônibus com carregamentos de, por exemplo, aproximadamente 50 passageiros e 25 embarques, que produz $\approx 1,3 \times 10^{14}$ combinações. Para

carregamentos menores, o cálculo é direto utilizando algoritmos para geração de todas as combinações possíveis, como em Ruskey e Williams (2009).

Para o exemplo, o cálculo é realizado para 5 possíveis embarques, dessa forma,

$$T_k = \langle P(0), P(1), P(2), P(3), P(4), P(5) \rangle$$

Calculando $P(0)$:

$$P(0) = (1 - P(B_{3,3})) \times (1 - P(B_{4,3})) \times (1 - P(B_{5,3})) \times (1 - P(B_{6,3})) \times (1 - P(B_{40,3}))$$

$$P(0) = (1 - 0.766) \times (1 - 0.318) \times (1 - 0.796) \times (1 - 0.283) \times (1 - 0.302)$$

$$P(0) = (1 - 0.766) \times (1 - 0.318) \times (1 - 0.796) \times (1 - 0.283) \times (1 - 0.302)$$

$$P(0) = 0.016$$

Calculando $P(1)$:

$$P(1) = (P(B_{3,3})) \times (1 - P(B_{4,3})) \times (1 - P(B_{5,3})) \times (1 - P(B_{6,3})) \times (1 - P(B_{40,3}))$$

$$+ (P(B_{4,3})) \times (1 - P(B_{3,3})) \times (1 - P(B_{5,3})) \times (1 - P(B_{6,3})) \times (1 - P(B_{40,3}))$$

$$+ (P(B_{5,3})) \times (1 - P(B_{3,3})) \times (1 - P(B_{4,3})) \times (1 - P(B_{6,3})) \times (1 - P(B_{40,3}))$$

$$+ (P(B_{6,3})) \times (1 - P(B_{3,3})) \times (1 - P(B_{4,3})) \times (1 - P(B_{5,3})) \times (1 - P(B_{40,3}))$$

$$+ (P(B_{40,3})) \times (1 - P(B_{3,3})) \times (1 - P(B_{4,3})) \times (1 - P(B_{5,3})) \times (1 - P(B_{6,3}))$$

Aplicando os valores das probabilidades em cada termo de $P(1)$ e calculando vamos obter:

$$P(1) = 0.137$$

Realizando os cálculos para $P(2)$, $P(3)$, $P(4)$ e $P(5)$ obtém-se:

$$P(0) = 0,016 \approx 1,6\%$$

$$P(1) = 0,138 \approx 13,8\%$$

$$P(2) = 0,368 \approx 36,8\%$$

$$P(3) = 0,336 \approx 33,6\%$$

$$P(4) = 0,125 \approx 12,5\%$$

$$P(5) = 0,017 \approx 1,7\%$$

Assim conclui-se que há 1,6% de probabilidade de que não tenha ocorrido algum embarque no ponto PT - 3 a partir dos dispositivos móveis detectados pelo MTR, e

que há em torno de 36,8% e 33,6% de que possam ter embarcado 2 ou 3 passageiros respectivamente no ponto PT - 3 a partir dos dispositivos móveis detectados pelo MTR.

Para efeito de exemplo, a informação de embarques nesse ponto é conhecida, e embarcaram 4 passageiros, se considerarmos a maior probabilidade que é 36.8% então estaríamos perdendo dois passageiros. Isso por que o cálculo tenta estimar com base nos eventos anteriores (olhando para o passado), e mesmo que a detecção tenha ocorrido somente após o ponto de embarque do passageiro, isso não quer dizer que ele não poderia ter embarcado em pontos anteriores devido as distribuições das taxas de buscas de rede serem irregulares.

Com base nos resultados, o ideal é obter distribuições individuais para cada dispositivo móvel de cada passageiro ao final do trajeto e executar o cálculo. Como para o exemplo foi utilizada uma distribuição resultante de experimentos com vários dispositivos móveis, a distribuição acaba sendo uma taxa que representa o comportamento médio de todos os dispositivos móveis, o que já foi discutido anteriormente que não é real, devido a inúmeros fatores. Tendo a distribuição individual dos dispositivos móveis, é possível ter uma estimativa melhor do embarque de cada passageiro, isso porque dependendo do comportamento do dispositivo móvel, é possível descartar pontos de embarques muito distantes.

Como exemplo, supondo que um único dispositivo móvel realiza buscas de redes a cada 10 segundos durante o trajeto, é possível descartar os pontos anteriores a sua detecção pelo MTR, por que se ele estivesse embarcado, e o tempo do trajeto entre os dois pontos for maior que 10 segundos, a detecção teria ocorrido 10 segundos antes do ponto da primeira detecção a bordo. Já dispositivos que apresentarem buscas de redes muito espaçadas, o efeito do exemplo calculado será repetido novamente, já que ele poderia ter embarcado em qualquer ponto antes de sua detecção pelo MTR.

Quando realizados os cálculos podem surgir casos em que há a probabilidade de que um número muito alto de passageiros portando dispositivos móveis possam ter embarcado a partir de um ponto de ônibus. Nesse caso é realizado uma aproximação de cálculo utilizando a heurística de classificação que será abordada na próxima seção.

7.2 CÁLCULO DE EMBARQUES APROXIMADO COM HEURÍSTICA DE CLASSIFICAÇÃO

Para lidar com a complexidade combinatória citada na Seção 7.1.2, considerando o caso particular em que as probabilidades de embarques no ponto de parada de ônibus k são os mesmos, por exemplo, $P(B_{ik}) = p$, $i \in \mathcal{K}$. Então, a distribuição T_k é binomial dada por:

$$T_k = \left\langle \binom{N}{j} p^j (1-p)^{N-j} \right\rangle, j = 0, \dots, N.$$

Com base nisso, o cálculo pode ser simplificado agrupando probabilidades de embarques similares em um único valor. Dessa forma, cria-se categorias de passageiros com (aproximadamente) igual probabilidade de embarque. A quantidade de embarques em k para cada categoria é dada pela fórmula binomial simplificando os cálculos.

O processo de classificação é realizado da seguinte forma. É definido um conjunto de categorias de probabilidades \mathcal{P}_k , os elementos são probabilidades únicas \bar{p}_c que caracterizam a categoria. Para este trabalho são consideradas dez categorias com $\mathcal{P}_k = \{0.05, 0.15, \dots, 0.95\}$. O passageiro portando o dispositivo móvel i é designado para uma categoria baseado na probabilidade $P(B_{ik})$ de acordo com:

$$\bar{p}_c - 0.05 < P(B_{ik}) \leq \bar{p}_c + 0.05, \quad (10)$$

onde c é o índice da categoria. Assim, todas as probabilidades de embarques são arredondadas para o mais próximo \bar{p}_c com um erro máximo de 0.05. Com essa variação de probabilidade, dez categorias são criadas ($c = 1, 2, \dots, 10$).

Utilizando como exemplo um ponto de parada de ônibus qualquer aqui denominado de PT - 9, onde após os cálculos $|B_9| = 75$, ou seja, há probabilidade de que 75 passageiros portando dispositivos móveis possam ter embarcado no ônibus naquele ponto. Utilizando o método direto para o cálculo de T_k , obtém-se

$$T_k = \langle P(0), P(1), \dots, P(75) \rangle.$$

Quando fosse realizado o cálculo para $P(38)$ o número de combinações possíveis seria $\approx 3.45 \times 10^{21}$, o que acarretaria em um esforço computacional muito grande. O que torna inviável realizar o cálculo por esse método. A Tabela 17 ilustra o cálculo realizado da divisão de classes das probabilidades do ponto PT - 9. É possível observar que alguns dispositivos móveis que foram considerados como possíveis embarques possuíam probabilidades muito baixas, e foram descartados, esses estão representados na tabela com um “x”. Os dispositivos móveis que tiveram suas probabilidades atribuídas para alguma classe, estão representados com “o” e com suas células mais escuras.

Dessa forma as classes são compostas pelos dispositivos móveis:

$$C_1 = \langle 70, 75, 92, 98, 99, 100, 101, 103, 108, 110 \rangle$$

$$C_2 = \langle 65, 67, 69, 72, 73, 79, 81, 85, 86, 89, 93, 94, 96, 97 \rangle$$

$$C_3 = \langle 61, 64, 66, 80, 82 \rangle$$

$$C_7 = \langle 49, 50, 52, 59, 60, 62, 68, 71, 74, 76, 77, 78, 83, 84 \rangle$$

$$C_8 = \langle 63 \rangle$$

Para cada classe criada, o total de embarques em k_c se torna:

$$T_k^c = \langle P_c(j') \rangle, \quad j' = 0, \dots, N^c, \quad (11)$$

Tabela 18: Tabela com o cálculo da distribuição binomial para as classes formadas por dispositivos móveis de passageiros com probabilidades de embarques no ponto PT - 9.

Distribuição	k	j'	$P_c(j')$				
			1	2	3	7	8
Embarcado	9	0	0,349	0,044	0,168	$4,78 \times 10^{-8}$	0,200
		1	0,387	0,154	0,360	$1,56 \times 10^{-6}$	0,800
		2	0,194	0,250	0,309	$2,37 \times 10^{-5}$	
		3	0,057	0,250	0,132	$2,21 \times 10^{-4}$	
		4	0,011	0,172	0,028	0,001	
		5	0,001	0,086	0,002	0,007	
		6	$1,38 \times 10^{-4}$	0,032		0,023	
		7	$8,75 \times 10^{-6}$	0,009		0,062	
		8	$3,65 \times 10^{-7}$	0,002		0,126	
		9	$9,00 \times 10^{-9}$	$3,36 \times 10^{-4}$		0,196	
		10	$1,00 \times 10^{-10}$	$4,20 \times 10^{-5}$		0,229	
		11		$3,82 \times 10^{-6}$		0,194	
		12		$2,39 \times 10^{-7}$		0,113	
		13		$9,18 \times 10^{-9}$		0,041	
		14		$1,64 \times 10^{-10}$		0,007	

Dessa forma T_k é a convolução discreta da variável binomial aleatória de cada categoria, o que pode ser calculado eficientemente. T_k pode ser escrito como:

$$T_k = \sum_{\substack{c_x \in \mathcal{C}_k \\ c_y \in \mathcal{C}_k}} P_{c_x}(j') \cdot P_{c_y}(j') \quad (12)$$

e pode ser simplificado por:

$$T_k = \sum_{c \in \mathcal{C}_k} T_k^c \quad (13)$$

Os elementos resultantes de T_k são valores aproximados de $P(j)$, cuja precisão pode ser controlada pelo intervalo de probabilidades no procedimento de classificação. Calculando T_k para o ponto PT - 9, a partir dos cálculos de T_k^c , obtemos:

Para $k = 9$, aplicando na Equação 13:

$$T_k = \sum_{c \in \mathcal{C}_k} T_9^1 \cdot T_9^2 \cdot T_9^3 \cdot T_9^7 \cdot T_9^8$$

Calculando em pares para $P_{c_1|c_2}(0) \rightarrow P_{c_A}(0)$:

$$P_{c_A}(0) = P_1(0) \times P_2(0)$$

Obtendo os valores da Tabela 18 e aplicando em $P_{c_A}(0)$ vamos obter:

$$P_{c_A}(0) = 0.015$$

Calculando para $P_{c_A}(1)$:

$$P_{c_A}(1) = P_1(0) \times P_2(1) + P_2(0) \times P_1(1)$$

Obtendo os valores da Tabela 18 e aplicando em $P_{c_A}(1)$ vamos obter:

$$P_{c_A}(1) = 0.071$$

Esse cálculo é realizado até $P_{c_A}(24)$, isso por que o total de elementos N é a soma da cardinalidade das duas classes, nesse caso $|T_9^A| = 24$. Assim:

$$|T_9^A| = \langle P_{c_A}(0), P_{c_A}(1), \dots, P_{c_A}(24) \rangle$$

Após o cálculo das classes 1 e 2, é possível calcular com a classe 3.

Dessa forma calculando em pares para $P_{c_A|c_3}(0)$:

$$P_{c_A|c_3}(0) = P_{c_A}(0) \times P_3(0)$$

$$P_{c_A|c_3}(0) = 0.015 \times 0.168 \approx 2.58 \times 10^{-3}$$

Calculando para $P_{c_A|c_3}(1)$:

$$P_{c_A|c_3}(1) = P_{c_A}(0) \times P_3(1) + P_3(0) \times P_{c_A}(1)$$

$$P_{c_A|c_3}(1) = 0.015 \times 0.360 + 0.168 \times 0.071 \approx 0.017$$

Esse cálculo é realizado até $P_{c_A|c_3}(29)$, onde o total de elementos N agora é a cardinalidade dos elementos de $|T_9^A|$ com os elementos de $|T_9^3|$.

O resultado do cálculo para todas as classes é ilustrado pela Tabela 19

Tabela 19: Cálculo de T_9 para as classes c_1, c_2, c_3, c_7 e c_8 , através do método da convolução.

Distribuição	k	j	$P(j)$	j	$P(j)$	j	$P(j)$
Embarcado	9	0	$2,465 \times 10^{-11}$	15	$1,409 \times 10^{-1}$	30	$4,170 \times 10^{-7}$
		1	$1,071 \times 10^{-9}$	16	$1,472 \times 10^{-1}$	31	$6,618 \times 10^{-8}$
		2	$2,208 \times 10^{-8}$	17	$1,342 \times 10^{-1}$	32	$9,216 \times 10^{-9}$
		3	$2,880 \times 10^{-7}$	18	$1,072 \times 10^{-1}$	33	$1,122 \times 10^{-9}$
		4	$2,667 \times 10^{-6}$	19	$7,508 \times 10^{-2}$	34	$1,187 \times 10^{-10}$
		5	$1,868 \times 10^{-5}$	20	$4,628 \times 10^{-2}$	35	$1,085 \times 10^{-11}$
		6	$1,028 \times 10^{-4}$	21	$2,514 \times 10^{-2}$	36	$8,487 \times 10^{-13}$
		7	$4,567 \times 10^{-4}$	22	$1,206 \times 10^{-2}$	37	$5,623 \times 10^{-14}$
		8	$1,668 \times 10^{-3}$	23	$5,115 \times 10^{-3}$	38	$3,109 \times 10^{-15}$
		9	$5,080 \times 10^{-3}$	24	$1,920 \times 10^{-3}$	39	$1,408 \times 10^{-16}$
		10	$1,303 \times 10^{-2}$	25	$6,384 \times 10^{-4}$	40	$5,078 \times 10^{-18}$
		11	$2,840 \times 10^{-2}$	26	$1,881 \times 10^{-4}$	41	$1,403 \times 10^{-19}$
		12	$5,294 \times 10^{-2}$	27	$4,906 \times 10^{-5}$	42	$2,787 \times 10^{-21}$
		13	$8,483 \times 10^{-2}$	28	$1,133 \times 10^{-5}$	43	$3,541 \times 10^{-23}$
		14	$1,174 \times 10^{-1}$	29	$2,314 \times 10^{-6}$	44	$2,160 \times 10^{-25}$

7.3 CÁLCULO POR CLASSIFICAÇÃO VS COMBINAÇÃO

O método de realizar o cálculo por classificação foi comparado com o método por combinação para verificar se a diferença entre os resultados seriam muito evidentes.

A ferramenta utilizada para realizar o cálculo conta com um computador com processador Core i7-7700HQ 2.80 GHz, 16 GB de memória RAM e um SSD WD Blue sn550 com leitura de 2400 MB/s e escrita de 1750 MB/s.

O primeiro resultado analisado será com relação ao tempo de execução. Para isso será considerado os dados de um ponto qualquer.

A tabela 20 ilustra a comparação entre os dois cálculos. Lembrando que quando é realizado o cálculo utilizando a classificação, algumas probabilidades podem ser descartadas. Para esse exemplo apresentado na tabela não houveram probabilidades descartadas.

Observando os tempos, foi identificado que o tempo computacional é muito relevante. O tempo total de cálculo do método por combinação foi de aproximadamente 7 minutos, enquanto o tempo de cálculo através da Heurística de Classificação levou menos de 1 segundo. Esse tempo varia de acordo com a quantidade de probabilidades de embarques em um ponto.

Se comparar os tempos de cada embarque, observa-se que o tempo é crescente quando o cálculo é feito por Combinações até $P(12)$ e após segue decrescente. Isso ocorre porque o maior número de combinações possíveis está em $P(12)$ com $2,7 \times 10^6$ combinações, e começa a diminuir a partir de $P(13)$. O mesmo não ocorre no método de classificação onde é crescente, mas os tempos são muito baixos.

Analisando o cálculo das probabilidades é possível verificar que há diferença quando os valores estão na cauda da distribuição binomial, mas no centro da distribuição a

diferença praticamente desaparece. Como pode ser observado entre P(4) e P(9). A Figura 24 ilustra o resultado do cálculo para comparação entre os dois métodos.

Levando em conta o tempo computacional, o método por Classificação se torna muito eficiente sem erros consideráveis se comparado com o método por combinação.

Tabela 20: Tempo de execução e cálculo da probabilidade utilizando modelo de combinação e Heurística de Classificação.

Embarques	Tempo de execução (s)		Cálculo de probabilidade		
	Combinação	Classificação	Combinação	Classificação	Diferença (%)
P(0)	0,034	0,002	$2,7 \times 10^{-5}$	$1,9 \times 10^{-5}$	-30
P(1)	0,023	0,004	$4,7 \times 10^{-4}$	$3,7 \times 10^{-4}$	-22
P(2)	0,032	0,007	$3,6 \times 10^{-3}$	$3,1 \times 10^{-3}$	-14
P(3)	0,057	0,009	$1,6 \times 10^{-2}$	$1,5 \times 10^{-2}$	-7
P(4)	0,276	0,012	$5,2 \times 10^{-2}$	$5,1 \times 10^{-2}$	-1
P(5)	0,971	0,015	$1,1 \times 10^{-1}$	$1,1 \times 10^{-1}$	4
P(6)	3,140	0,017	$1,7 \times 10^{-1}$	$1,8 \times 10^{-1}$	6
P(7)	8,043	0,019	$2,0 \times 10^{-1}$	$2,2 \times 10^{-1}$	5
P(8)	17,400	0,021	$1,8 \times 10^{-1}$	$1,8 \times 10^{-1}$	1
P(9)	36,695	0,024	$1,2 \times 10^{-1}$	$1,2 \times 10^{-1}$	-5
P(10)	50,579	0,026	$6,9 \times 10^{-2}$	$6,0 \times 10^{-2}$	-13
P(11)	63,264	0,028	$2,9 \times 10^{-2}$	$2,3 \times 10^{-2}$	-22
P(12)	70,593	0,030	$1,0 \times 10^{-2}$	$7,0 \times 10^{-3}$	-31
P(13)	67,510	0,032	$2,7 \times 10^{-3}$	$1,6 \times 10^{-3}$	-39
P(14)	50,712	0,034	$6,1 \times 10^{-4}$	$3,2 \times 10^{-4}$	-47
P(15)	31,620	0,036	$1,1 \times 10^{-4}$	$5,0 \times 10^{-5}$	-55
P(16)	18,287	0,038	$1,6 \times 10^{-5}$	$6,2 \times 10^{-6}$	-61
P(17)	8,447	0,040	$1,8 \times 10^{-6}$	$6,2 \times 10^{-7}$	-67
P(18)	3,305	0,043	$1,7 \times 10^{-7}$	$4,9 \times 10^{-8}$	-72
P(19)	1,013	0,046	$1,2 \times 10^{-8}$	$3,1 \times 10^{-9}$	-76
P(20)	0,240	0,048	$7,3 \times 10^{-10}$	$1,4 \times 10^{-10}$	-80
P(21)	0,047	0,050	$3,0 \times 10^{-11}$	$5,3 \times 10^{-12}$	-83
P(22)	0,008	0,052	$8,9 \times 10^{-13}$	$1,3 \times 10^{-13}$	-85
P(23)	0,001	0,054	$1,6 \times 10^{-14}$	$2,0 \times 10^{-15}$	-87
P(24)	0,001	0,056	$1,3 \times 10^{-16}$	$1,4 \times 10^{-17}$	-89
Total	432,298	0,742	1	1	0

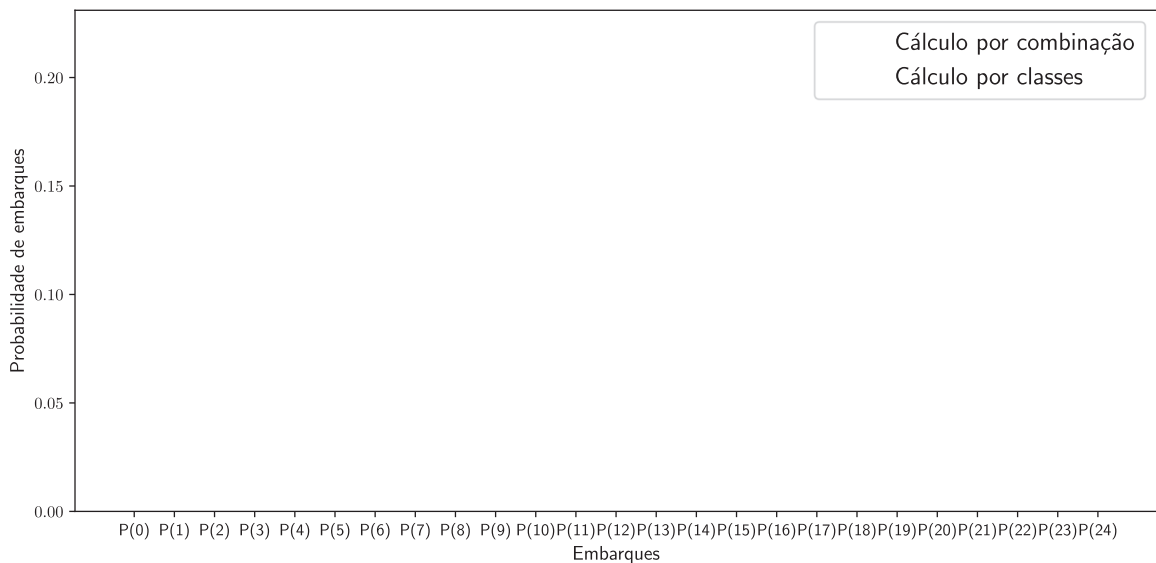


Figura 24: Comparação do resultado das probabilidades entre o método por combinação e por classes.

7.4 ESTIMATIVA DO TOTAL DE EMBARQUES EM PONTOS DE ÔNIBUS AGREGADOS POR ZONAS

Da mesma forma que os cálculos foram realizados para pontos de parada de ônibus com muitos embarques utilizando convolução, os mesmos cálculos podem ser aplicados para zonas de embarques. Isso pode ser útil em locais onde os pontos de ônibus são muito próximos. Considera-se uma zona de embarque, um agrupamento de pontos de parada de ônibus. Seja Z a zona formada por pontos de parada de ônibus, então

$$Z_n = \langle i \rangle, \quad i = 0, \dots, k, \quad (14)$$

onde n é o número identificador da zona e é atribuído de acordo com a necessidade, e i é o número do ponto de parada de ônibus que será agregado à zona.

Assim partindo da equação 13, obtém-se

$$T_{z_n} = \sum_{k \in Z_n} T_k \quad (15)$$

Calculando o total de embarques em zonas para o exemplo ilustrado pela Figura 27 com 17 pontos de paradas de ônibus. Para o agrupamento dos pontos de paradas de ônibus em zonas, pode ser usado qualquer critério desejado, para esse estudo de caso foram utilizadas duas regras. A primeira diz que se os pontos de paradas de ônibus estão na mesma zona de tráfego eles serão considerados na mesma zona de embarque; a segunda regra diz que caso os pontos de paradas de ônibus não estejam na mesma zona de tráfego e

o tempo de deslocamento entre eles for menor ou igual a 111 segundos eles são considerados na mesma zona. A Figura 25 ilustra o exemplo de agrupamento de pontos em zonas de embarques.

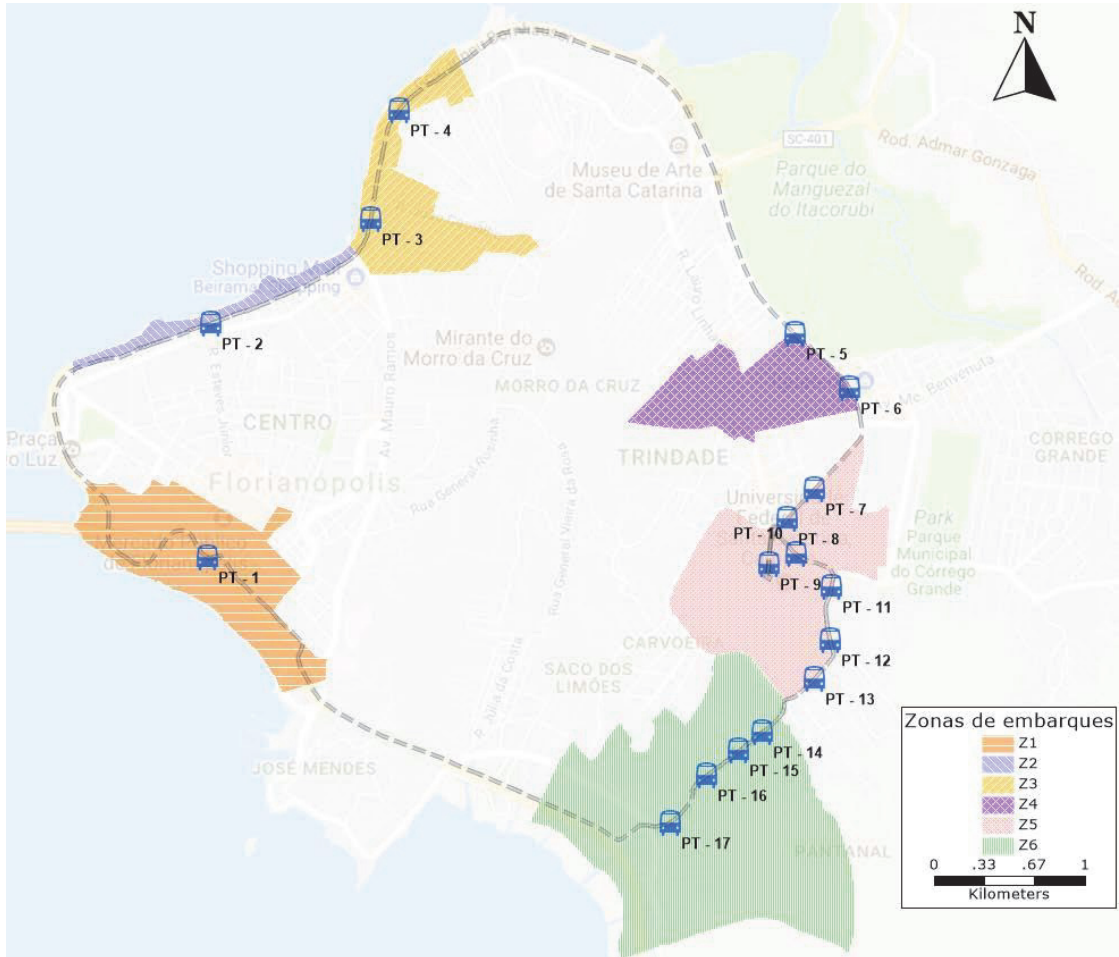


Figura 25: Ponto agrupados por Zonas de embarque.

Dessa forma as zonas compostas pelo agrupamento de pontos de embarques são:

$$Z_1 = \langle 1 \rangle$$

$$Z_2 = \langle 2 \rangle$$

$$Z_3 = \langle 3, 4 \rangle$$

$$Z_4 = \langle 4, 5 \rangle$$

$$Z_5 = \langle 7, 8, 9, 10, 11, 12, 13 \rangle$$

$$Z_6 = \langle 14, 15, 16, 17 \rangle$$

Com os pontos agrupados em zonas e com todos os valores de probabilidade de embarques T_k em cada ponto, é possível calcular de acordo com a Equação 15 os embarques realizados na zona. Assim calculando para o estudo de caso:

portando o dispositivo móvel i desembarcou no ponto de ônibus l , quando detectado pelo MTR a bordo entre os pontos k e $k + 1$ é dada por:

$$P(A_{il}) = \frac{1}{k+n}, \quad l = 1, 2, \dots, k+n, \quad (16)$$

A partir do momento da última detecção pelo MTR e com a informação prévia da distribuição da taxa do intervalo de buscas de rede X_i , que é conhecida, é possível de acordo a regra de Bayes determinar a probabilidade *a posteriori*, de que um passageiro portando um dispositivo móvel i desembarcou em l e foi detectado pelo MTR a bordo após k e antes de $k + 1$. Esta probabilidade é dada por:

$$P(A_{il}|R_i) = \frac{P(A_{il})P(R_i|A_{il})}{P(R_i)}. \quad (17)$$

A probabilidade condicional $P(R_i|B_{il})$ define que a probabilidade da última busca de rede detectada pelo MTR de um dispositivo móvel que estava embarcado ocorreu após o desembarque no ponto l e $P(R_i)$ a probabilidade de que a busca de rede a bordo do ônibus do dispositivo móvel i ocorreu após o ponto k . Para encontrar a probabilidade $P(R_i|B_{il})$, é considerado a Equação 4.

Lembrando que R_i é a última busca de rede embarcado de i . Assumindo a condição de que o desembarque ocorreu em l e que a última busca de rede detectada R_i ocorreu entre k e $k + 1$, a probabilidade desejada é equivalente à probabilidade de que a próxima busca de rede deveria ocorrer após l enquanto o passageiro i estava ainda dentro do veículo. Essa observação estabelece uma conexão direta entre Δt_{il} e X_i expressa por:

$$P(R_i|B_{il}) = P(X_i \geq \Delta t_{il}). \quad (18)$$

Se a última busca de rede detectada pelo MTR ocorreu imediatamente antes do desembarque em l , então a próxima busca de rede quase certamente ocorreu fora do veículo. Tal conclusão resulta da premissa (condicional) de que o desembarque de fato ocorreu em l .

A aplicação da Equação 19 implica em obter $P(R_i)$, ou a probabilidade de que a última busca de rede do dispositivo móvel embarcado tenha ocorrido após o ponto k . Essa probabilidade pode ser calculada a partir da probabilidade total, dada por:

$$P(R_i) = \sum_{l=1}^k P(B_{il})P(R_i|B_{il}). \quad (19)$$

A Equação 17 resulta na distribuição de probabilidade de um passageiro ter desembarcado em todos os pontos de ônibus posteriores possíveis a partir da última detecção, e

deve obedecer a seguinte equação:

$$\sum_{l=1}^k P(B_{il}|R_i) = 1. \quad (20)$$

A partir desse momento o processo é praticamente similar ao embarque, a diferença é que agora é necessário verificar os pontos a frente da última busca de rede detectada. Nessa Tese não serão apresentados os resultados do desembarque, somente o do embarque.

8 ESTUDO DE CASO DA ESTIMATIVA DE EMBARQUES

O modelo de estimativa de embarques será aplicado no trajeto da linha de transporte coletivo da cidade de Florianópolis, capital do estado de Santa Catarina. A linha considerada é a linha 185-UFSC SEMIDIRETO, que tem a maior demanda do sistema. A Tabela 21 apresenta as características da linha, que possui 17 pontos de paradas com trechos de velocidades variável onde não há corredor exclusivo para ônibus. O intervalo entre partidas dessa linha é de 7 minutos no pico. A Figura 27 ilustra o trajeto realizado pela linha. A numeração dos pontos de paradas de ônibus são sequenciais e indicam o sentido da linha, o número é indicado por PT - x, onde x é o número identificador do ponto de parada de ônibus. Além do número do ponto de parada de ônibus, a figura também apresenta a quantidade de passageiros que embarcou em cada ponto. A quantidade é indicada como $EMB = y$, onde y é o total de embarques que ocorreu naquele ponto. Os pontos de maiores carregamentos são o ponto 1 que corresponde ao Terminal de Integração Centro - TICEN, um terminal de pré embarque em que o pagamento ocorre fora do ônibus, e o ponto 9 que fica localizado na Universidade Federal de Santa Catarina. O ponto 9 não conta com uma estação de pré embarque, e em horários de pico possui um tempo de embarque com média de 5 minutos. A linha parte do Terminal de Integração Centro - TICEN e termina sua viagem no próprio terminal. Normalmente as viagens que terminam no terminal não recomeçam no mesmo veículo, e é o que está sendo considerado para esse estudo de caso. Os dados de tempo de viagem foram obtidos direto com o operador através do GPS embarcado nos veículos que fazem a linha. Esse tempo de viagem já inclui os tempos de paradas nos pontos de ônibus.

Tabela 21: Tabela com as características da Linha 185-UFSC SEMIDIRETO

Linha	Número de Pontos de ônibus	Ponto Origem	Ponto Destino	Velocidade entre pontos (km/h)	Distância entre pontos (m)	Tempo entre Pontos (s)
185 UFSC SEMIDIRETO	17	1	2	31	2783,20	323
		2	3	31	1168,44	135
		3	4	31	665,76	77
		4	5	31	3458,50	401
		5	6	25	447,92	64
		6	7	25	705,02	101
		7	8	18	232,64	46
		8	9	12	542,41	162
		9	10	12	362,93	108
		10	11	12	333,90	100
		11	12	12	316,81	95
		12	13	12	267,73	80
		13	14	12	461,18	138
		14	15	12	174,18	52
		15	16	12	235,23	70
		16	17	12	366,85	110
		17	1	31	3417,22	396

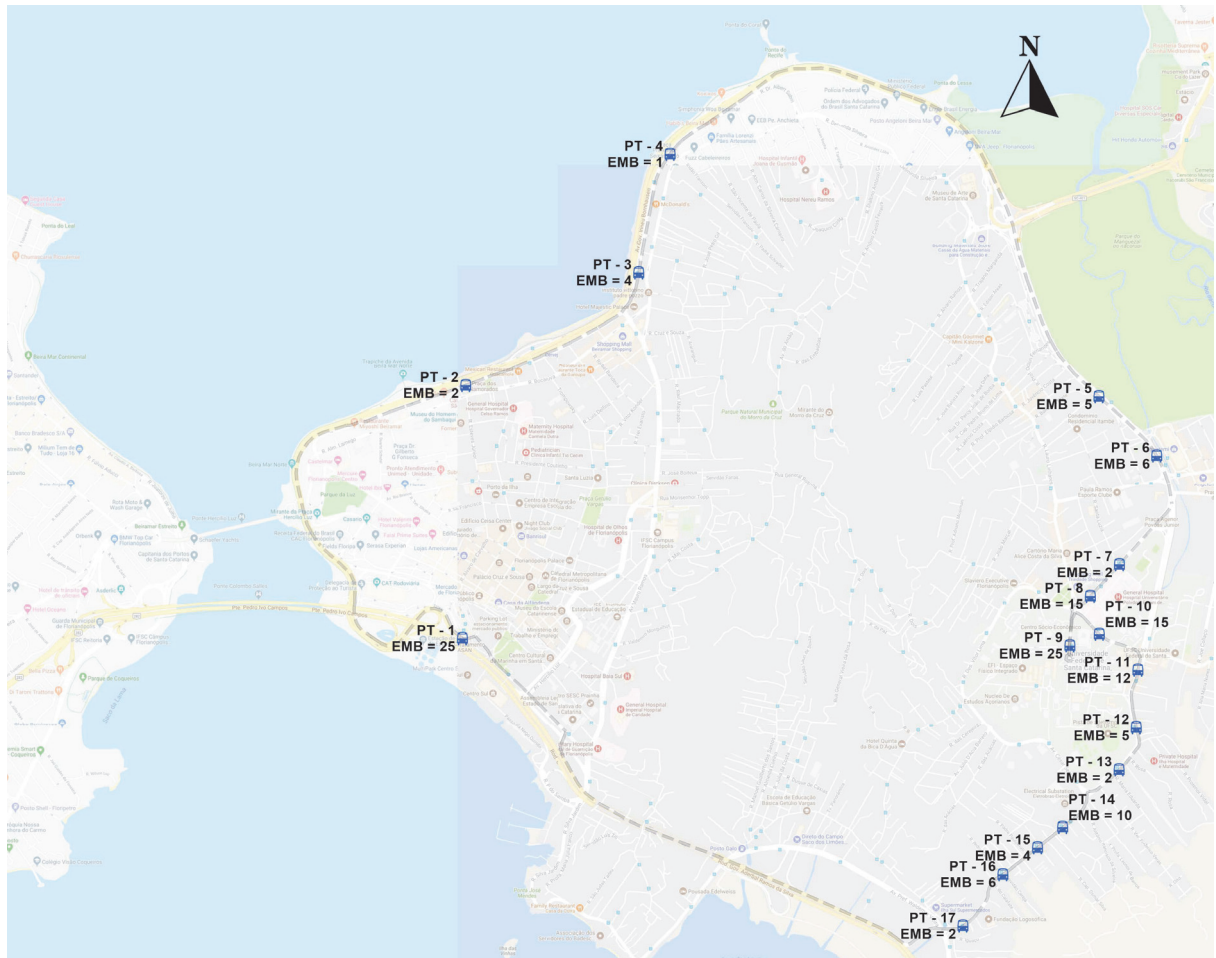


Figura 27: Trajeto da linha 185-UFSC SEMIDIRETO com 17 pontos de paradas de ônibus, onde EMB é o número de embarques conhecidos de cada ponto.

8.1 MODELO DE GERAÇÃO ARTIFICIAL DE DETECÇÕES

Devido ao cenário atual de uma pandemia, não foi possível se realizar experimentos controlados e obter um número razoável de amostras de dados de passageiros embarcados. Dessa forma foi desenvolvido um modelo de geração de detecções de forma artificial.

O modelo de geração de detecções gera a primeira detecção realizada pelo MTR através da primeira busca de rede realizada por dispositivos móveis. Esse modelo simula como se o MTR estivesse a bordo de um ônibus que estava realizando o trajeto do estudo de caso. Dentro desse ônibus em cada ponto embarcaram passageiros, todos portando dispositivos móveis com o Wi-Fi ativo. Ao longo da viagem os dispositivos móveis desses passageiros foram realizando buscas de rede assim que embarcaram, e essas buscas foram detectados pelo MTR a primeira vez. Como é um modelo gerado sinteticamente temos a certeza de que o passageiro estava embarcado por que ele irá preencher todos os requisitos para considerar que um passageiro está embarcado (Capítulo 4.3).

A definição do momento em que ocorreu essa detecção pelo MTR, é obtida através da distribuição da taxa de buscas de rede. A detecção é gerada através de uma distribuição

uniforme com mínimo 1 segundos e máximo duas vezes a média da distribuição, isso para garantir que no mínimo 80% das amostras de buscas de rede ocorreram dentro desse intervalo. Caso fosse escolhida somente a média da distribuição da taxa de buscas de rede, seria a mesma coisa que dizer que somente 50% dos dispositivos móveis se comportam dessa forma. A distribuição uniforme foi escolhido devido a aleatoriedade com que os dispositivos móveis realizam uma busca de rede, ela pode ocorrer em qualquer intervalo desde 1 segundo até um valor máximo. Três distribuições da taxa de busca de rede foram utilizadas nesta pesquisa.

A primeira distribuição da taxa de buscas foi obtida através dos resultados do experimento embarcado do Capítulo 4.3 e é ilustrado na Figura 28(a) que representa o número de buscas de rede e o número de dispositivos móveis distintos detectados. Ao todo foram detectados 122 MAC's de dispositivos móveis distintos com 856 detecções em um período de 45 minutos de experimento. A Figura 28(b) é a distribuição da taxa de buscas de rede da Figura 28(a), sendo essa a distribuição que representa o comportamento médio das buscas de rede de um dispositivo móvel.

A segunda distribuição da taxa de buscas foi obtida através dos resultados do experimento realizado com o aplicativo para dispositivos móveis do Capítulo 5, onde foram analisados os dados de 6 MAC's de dispositivos móveis distintos com 424 detecções. Foi escolhido um período de 3 horas de dados, já que um período maior apresentou um resultado similar na média. O processamento da média é ilustrado na Figura 28(d) e representa também o comportamento médio das buscas de rede de um dispositivo móvel.

Para a terceira distribuição da taxa de buscas, após uma filtragem dos dados, alguns *outliers*¹ foram encontrados. Esses outliers correspondem a dispositivos que sinalizam muitas vezes e causam um impacto significativo no número de detecções fazendo com que ocorram picos que distorcem a interpretação dos resultados. Essa distribuição foi obtida através dos dados de um detector da pesquisa de Nunes (2018), onde seu objetivo era realizar viagens de ônibus com um dispositivo detector de sinais Wi-Fi e obter a ocupação do veículo. A Figura 28(e) é a representação de 206 MAC's de dispositivos móveis distintos em 2 viagens de ônibus com um trajeto de aproximadamente 52 minutos. O processamento da média é ilustrado na Figura 28(f) e representa também o comportamento médio das buscas de rede de um dispositivo móvel.

A Tabela 22 ilustra de forma resumida os dados obtidos das distribuições.

Assim o modelo de geração de detecções assume algumas premissas, são elas:

- A primeira detecção gerada é a primeira detecção registrada pelo MTR;
- A detecção gerada do dispositivo móvel sempre é considerada válida, garantindo que

¹ Em estatística um *outlier* é um ponto nos dados que difere significativamente das outras observações. Um *outlier* pode ocorrer devido às variações nas medições ou até mesmo indicar um erro experimental. Na análise dos dados, algumas vezes um outlier é excluído do conjunto de dados (ZIMEK; FILZMOSER, 2018).

ocorreram outras detecções (mesmo que não geradas) do mesmo dispositivo ao longo do trajeto;

- Cada detecção gerada corresponde exclusivamente a um único dispositivo móvel, ou seja, não há duas detecções geradas atribuídas há um mesmo dispositivo móvel;
- Os pontos de ônibus dos locais de embarques dos dispositivos móveis gerados são conhecidos;
- A detecção ocorre de acordo com o intervalo $1 \leq \text{detecção} \leq 2 \times \text{média}$, onde a média é obtida através da distribuição média de buscas de rede escolhida.

Tabela 22: Tabela resumida das informações sobre buscas de rede obtidas de diferentes fontes de dados.

Distribuição	MAC's Distintos	Detecções	Média (s)	2x Média (s)	Desvio padrão
Embarcado	122	856	122,13	244,26	113,93
Aplicativo	6	424	221,69	443,38	176,63
Nunes (2018)	206	1145	196,54	393,08	162,53

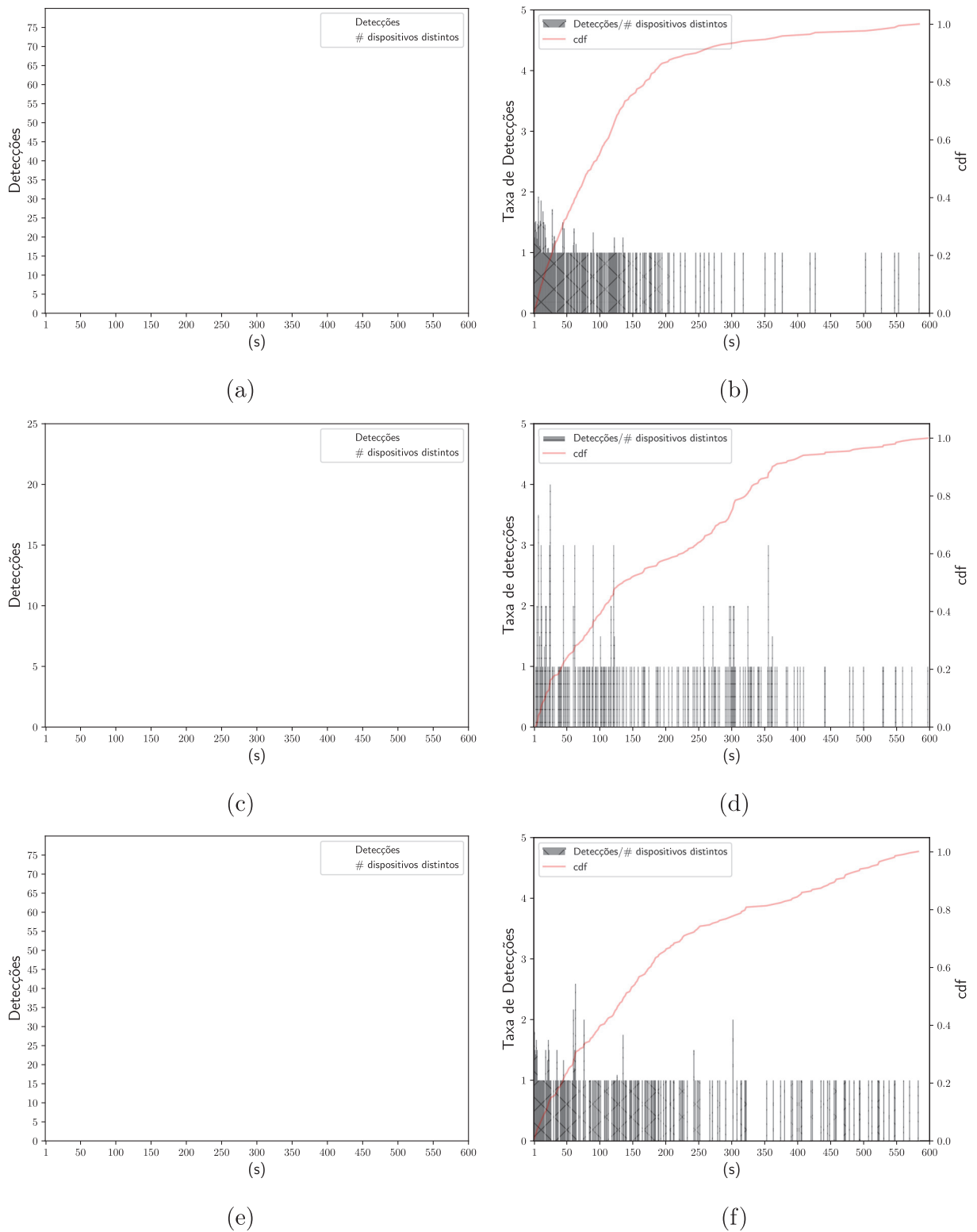


Figura 28: (a,c,e) Número de detecções de buscas de rede com dispositivos distintos registrados; (b,d,f) média do intervalo de buscas de rede. (a,b) Experimento embarcado (Capítulo 4.3); (c,d) Experimento aplicativo (Capítulo 5); (e,f) Experimento realizado por Nunes (2018)

Foi aplicado o modelo de geração de detecções no estudo de caso para as três distribuições de taxa de detecção. Foram geradas 141 detecções de dispositivos móveis para cada uma das três distribuições de taxa de buscas de rede. Os embarques são conhecidos

e estão representados na Figura 27. A representação de todas as detecções geradas para cada uma das distribuições em uma única imagem não é viável, assim a Figura 29 ilustra apenas três detecções geradas uma para cada distribuição individualmente. O dispositivo móvel que teve sua detecção gerada é representado pelo número 15.

Analisando a distribuição Embarcado:

- Representado por "DIST - Embarcado" na Figura 29;
- Nunca a detecção seria gerada após o ponto PT - 2, isso por que de acordo com a Tabela 21 o tempo de deslocamento entre os pontos PT - 1 e PT - 2 é de 323 s;
- Para gerar a detecção o tempo gerado precisa estar entre $1 \text{ s} \leq \text{detecção} \leq 244,26 \text{ s}$

Assim na distribuição Embarcado não haveria como essa detecção ser gerada após o ponto PT - 2. Como o ponto PT - 1 é o ponto de início da viagem, temos a certeza que o passageiro portando o dispositivo móvel 15 embarcou no ponto PT - 1.

Analisando a distribuição Nunes (2018):

- Representado por "DIST - Nunes (2018)" na Figura 29;
- Há possibilidade da detecção ser gerada após o ponto PT - 2, isso por que o intervalo de tempo gerado está entre $1 \text{ s} \leq \text{detecção} \leq 393,08 \text{ s}$ de acordo com a Tabela 21;
- O tempo máximo é maior do que o deslocamento entre os pontos PT - 1 e PT - 2 de 323 s.

Como o modelo de geração é aleatório entre 1 segundo e o máximo, e a detecção foi gerada em 214 s, ou seja, antes de completar o deslocamento até o ponto PT - 2. Da mesma forma que na Distribuição Embarcado, é possível afirmar que o passageiro portando o dispositivo móvel 15 embarcou no ponto PT - 1.

Analisando a distribuição Aplicativo:

- Representado como "DIST - Aplicativo" na Figura 29;
- O intervalo de tempo máximo que uma detecção poderia ser gerada é 443,38 s, de acordo com Tabela 21. Assim a detecção poderia ocorrer entre $1 \text{ s} \leq \text{detecção} \leq 443,38 \text{ s}$, e acabou sendo gerada aleatoriamente em 368 segundos.
- A detecção ocorreu após o ponto PT - 2, já que o trajeto entre o ponto 1 e ponto 2 é 323 segundos.

Para a distribuição aplicativo, o fato da detecção ocorrer após o ponto PT - 2, significa que um passageiro embarcou com seu dispositivo móvel no ponto PT - 1, e só foi detectado a primeira vez pelo MTR após passar o ponto PT - 2. E nesse momento não

temos mais a certeza de onde ele embarcou, ele pode ter embarcado no ponto PT - 1 ou ponto PT - 2.

Lembrando que o ponto PT - 1 possui 25 embarques, assim existem 25 detecções para cada distribuição nesse ponto, podendo a detecção ser gerada em qualquer ponto do trajeto desde que respeite o intervalo máximo de duas vezes a média da distribuição.

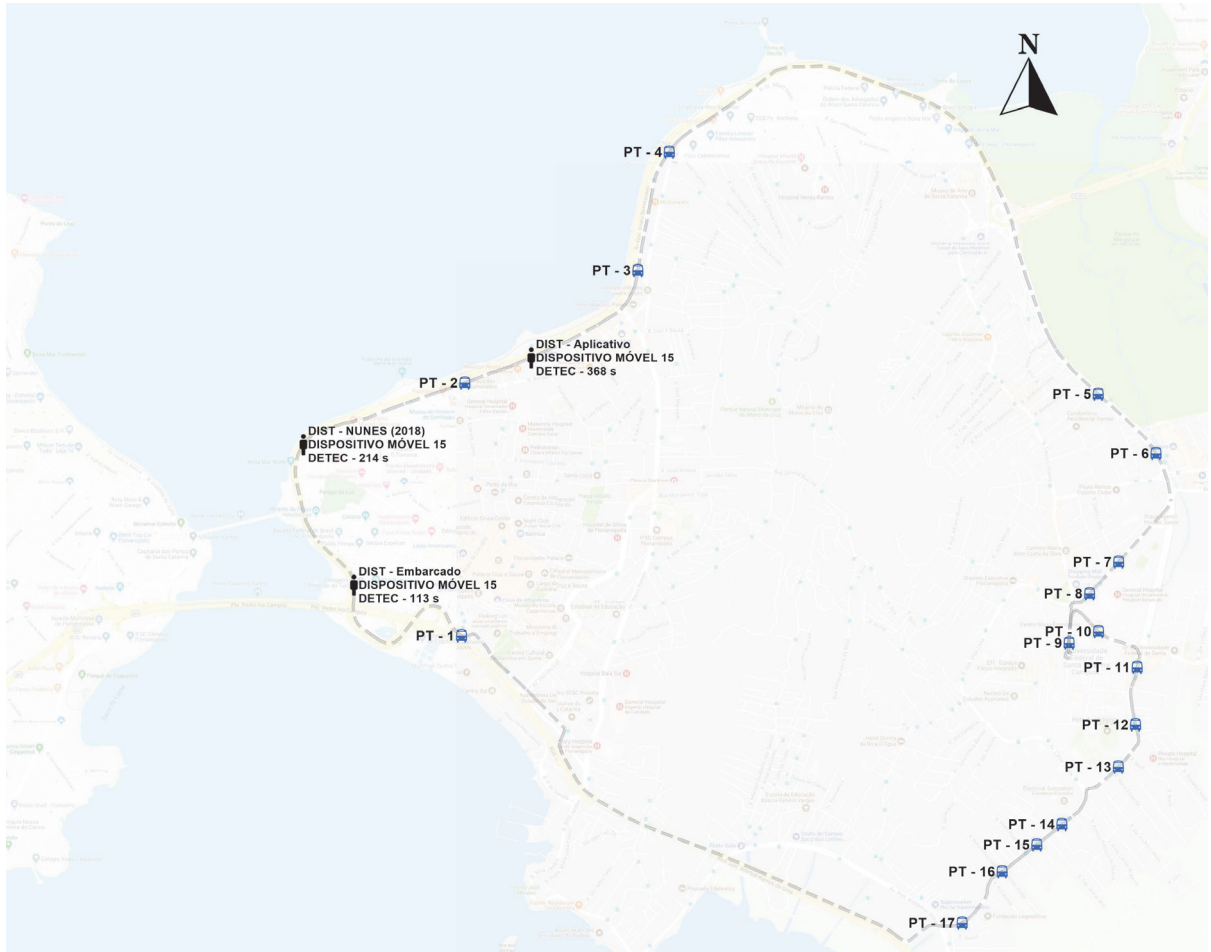


Figura 29: Exemplo de detecções gerados no software Transcad (CALIPER, 2020) através de uma distribuição uniforme, onde o mínimo é igual a 1 e o máximo é igual a duas vezes a média da distribuição escolhida através da Tabela 22.

Todo esse procedimento de geração é realizado ponto a ponto. Por exemplo, a partir do ponto PT - 2 são geradas duas detecções de mais 2 dispositivos móveis que embarcaram de acordo com a Figura 27, já para o ponto PT - 3, são geradas 4 detecções, da mesma forma para os demais o número de detecções geradas segue os embarques de cada ponto acordo com a Figura 27.

Um detalhe importante é sobre o ponto PT - 17 e a Distribuição Aplicativo. Como foi citado anteriormente, o ponto final da linha é o ponto PT - 1 e que a viagem não se repete com o mesmo veículo. Um novo veículo é necessário, sendo caracterizada uma nova viagem. Assim, um dispositivo móvel que foi detectado pelo MTR a partir de um embarque do ponto PT - 17, não pode ser detectado além do ponto PT - 1. O trajeto do

ponto P - 17 até o ponto PT - 1 é de 396 segundos, e o tempo máximo de detecção para essa distribuição é de 443,38 segundos seguindo a regra imposta do máximo da distribuição uniforme ser duas vezes a média da distribuição, poderia ser gerada alguma detecção além de 396 segundos. Caso isso ocorra, esse dispositivo móvel é descartado. A Figura 30 ilustra esse caso, em que o dispositivo móvel 140 embarcou no ponto PT - 17, e sua detecção foi gerada aleatoriamente em 430 segundos, sendo um tempo maior que o trajeto do ponto PT - 17 até o ponto 1. Como o desembarque no ponto PT - 1 é obrigatório, esse dispositivo móvel acaba sendo descartado, e não é considerado e seu embarque é descartado.

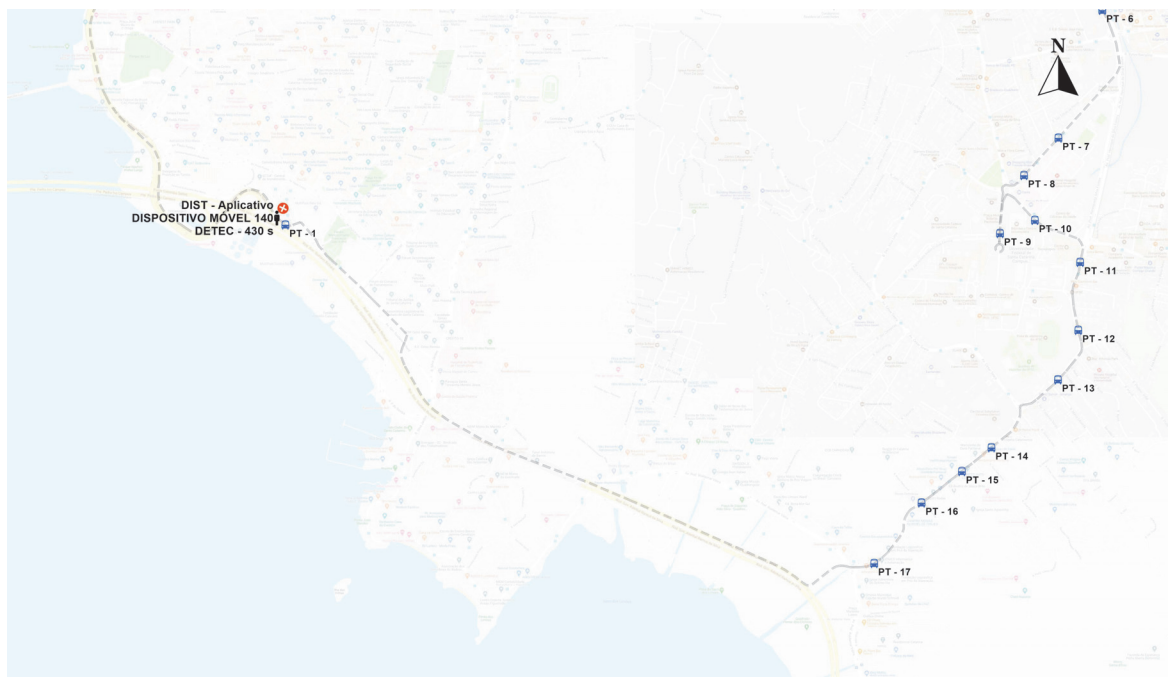


Figura 30: Descarte de detecção quando a geração ocorre em um intervalo de tempo maior que o trajeto até o ponto 1.

Com as detecções devidamente geradas, aplica-se o modelo de estimativa de embarques para determinar em qual ponto cada passageiro portando um dispositivo móvel embarcou.

A partir da geração de detecções foram realizados os cálculo das probabilidade de embarques aplicando o método proposto no Capítulo 7.

8.2 RESULTADO DO CÁLCULO DA ESTIMATIVA DO TOTAL DE EMBARQUES EM PONTOS DE PARADA DE ÔNIBUS

8.2.1 PONTO DE PARADA DE ÔNIBUS PT - 1

O ponto de parada de ônibus PT - 1 possui as características da Tabela 23. Com base nas informações da tabela é possível observar para a distribuição Embarcado que, mesmo com a geração de detecções sendo aleatória, não há possibilidade das detecções

serem geradas após o ponto PT - 2. Ou seja, todas as detecções foram geradas antes do ponto PT - 2 (Figura 31(a)). Isso garante uma probabilidade muito maior de que todos os embarques dessas detecções geradas sejam associadas a passageiros que embarcaram no ponto de parada de ônibus PT - 1 isso por que $P(B_{i1}) = 1$. Já para a distribuição Aplicativo foram geradas seis detecções após o ponto PT - 2 (Figura 31(b)), e para a distribuição Nunes (2018) foram geradas duas detecção após o ponto PT - 2 (Figura 31(c)). Essas detecções após o ponto PT - 2 faz com que a probabilidade de embarque $P(B_{i1}) = 1/2$ e $P(B_{i2}) = 1/2$. As detecções consideradas normalmente serão maiores do que o número de embarques conhecidos, isso porque como já comentado anteriormente há efeito das detecções geradas dos pontos posteriores aos que estão sendo analisados. Para esse caso na distribuição Embarcado, foram consideradas 5 detecções que estão entre os pontos PT - 2 e PT - 17 já que, essas detecções podem estar em qualquer lugar.

Tabela 23: Dados do ponto de parada de ônibus PT - 1

Ponto de parada de ônibus	Distribuição	$2 \times \text{Média}$	Nº Detecções Geradas	Tempo de Viagem PT - 1 \rightarrow PT - 2 (s)	Detecções Consideradas
PT - 1	Embarcado	244	25	323	30
	Aplicativo	443			28
	Nunes (2018)	393			27

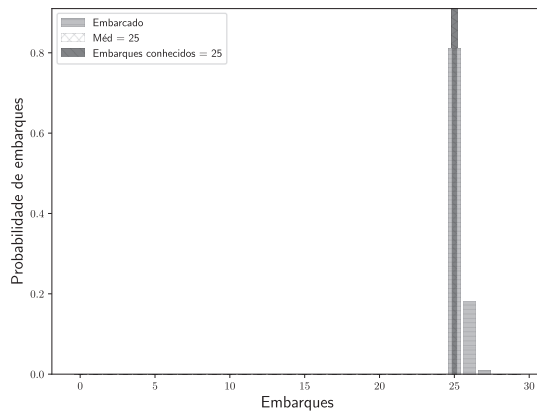
Para todas as distribuições foi aplicado o método de cálculo divisão por classes, isso porque a quantidade de detecções consideradas é muito alta para realizar o cálculo direto como já explicado no Capítulo 7. Dessa forma define-se como o número máximo detecções consideradas para cálculo direto o limite de 20 detecções por distribuição, qualquer distribuição com mais de 20 detecções consideradas é realizado o cálculo de divisão por classes. A Figura 32 ilustra o resultado da estimativa de embarques no ponto de parada de ônibus PT - 1 para cada uma das distribuições. A distribuição embarcado apresentou o melhor resultado dentre as três distribuições, estimando com 80% que ocorreram em média 25 embarques no ponto de parada de ônibus PT - 1. A distribuição Aplicativo apresentou o pior resultado, estimando com aproximadamente 40% que ocorreram em média 18 embarques. Isso ocorre devido ao fato de seis detecções terem ocorrido após o ponto 2, fazendo com que a probabilidade de embarque $P(B_{il})$ seja dividida entre o ponto de parada de ônibus PT - 1 e PT - 2. Já para o caso da distribuição de Nunes (2018) foi estimado com aproximadamente 58% que ocorreram em média 23 embarques. Para o caso da distribuição embarcado pode levar a impressão de que como não ocorreram detecções após o ponto de parada de ônibus PT - 2, a probabilidade de embarque em PT - 1 deveria ser 100%. Mas isso não ocorre porque supõe-se que não sabemos quantos embarques ocorreram em cada ponto, então os embarques que ocorreram no ponto PT - 2 também influenciam com alguma probabilidade de terem ocorrido em PT - 1. Isso ocorre para todos os pontos, principalmente pontos muito próximos.



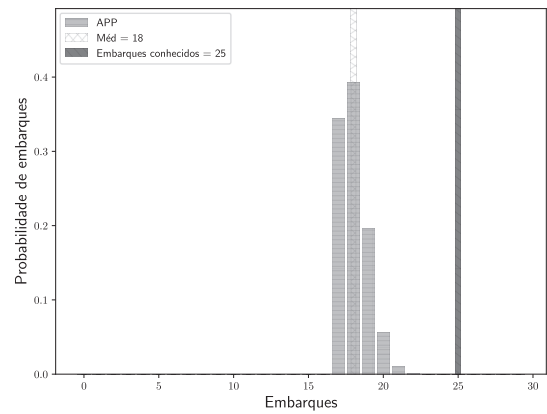
Figura 31: Detecções geradas para 25 embarques no ponto PT - 1. (a) Distribuição Embarcado, (b) Distribuição Aplicativo e (c) Distribuição Nunes (2018).

8.2.2 PONTO DE PARADA DE ÔNIBUS PT - 2

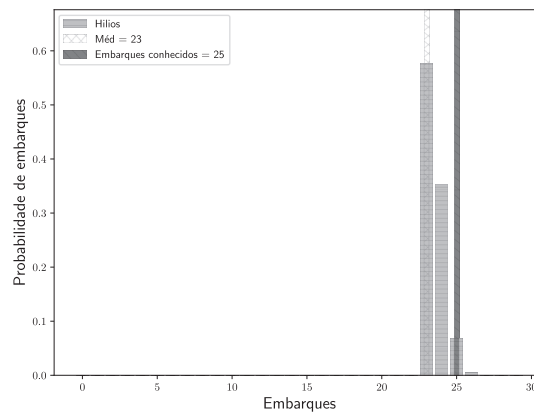
No ponto de parada de ônibus PT - 2, de acordo com a Tabela 24, foram geradas duas detecções. Como a geração de detecção para esse ponto é aleatória, há possibilidade de qualquer detecção de qualquer distribuição ser gerada após o ponto P3. Porém na distribuição Embarcado duas detecções foram geradas muito próximas (Figura 33(a)). Já na distribuição Aplicativo uma das detecções foi gerada após o ponto P3 (Figura 33(b)). Para a distribuição Nunes (2018) as duas detecções foram geradas após o ponto P4 (Figura 33(c)). A partir dos embarques em P2, as probabilidades de embarques começam a levar em consideração todos os pontos anteriores. Nesse caso, para detecções que ocorreram entre o trajeto P2 e P3 $P(B_{i2}) = 1/2$ e para detecções que ocorreram após P3 $P(B_{i3}) = 1/3$. A probabilidade passa a ser dividida por todos os pontos, porque há chances de que o



(a)



(b)



(c)

Figura 32: Estimativa de embarques realizados no ponto de parada de ônibus PT - 1 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

embarque possa ter ocorrido em qualquer um dos pontos anteriores à detecção. Da mesma forma que o ponto P1 o ponto P2 também sofre influência das detecções que ocorreram em pontos posteriores próximos como P3 e P4, dessa forma as detecções consideradas estão representadas na Tabela 24.

Tabela 24: Dados do ponto de parada de ônibus PT - 2

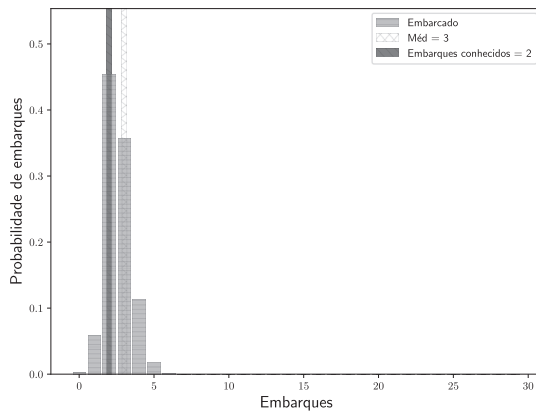
Ponto de parada de ônibus	Distribuição	$2 \times \text{Média}$	Nº Detecções Geradas	Tempo de Viagem PT - 2 \rightarrow PT - 3 (s)	Detecções Consideradas
PT - 2	Embarcado	244	2	135	7
	Aplicativo	443			15
	Nunes (2018)	393			9

Como o número de detecções consideradas para todas as distribuições é menor que 20, o método de cálculo aplicado foi o direto. A Figura 34 indica que, em média, as distribuições Embarcado e Nunes (2018) apresentaram o melhor resultado. A distribuição

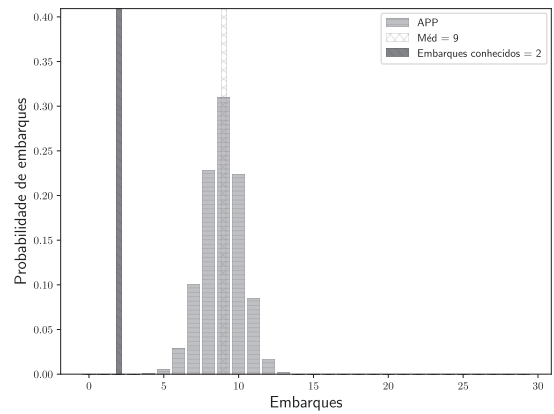


Figura 33: Detecções geradas para 2 embarques no ponto PT - 2. (a) Distribuição Embarcado, (b) Distribuição Aplicativo e (c) Distribuição Nunes (2018).

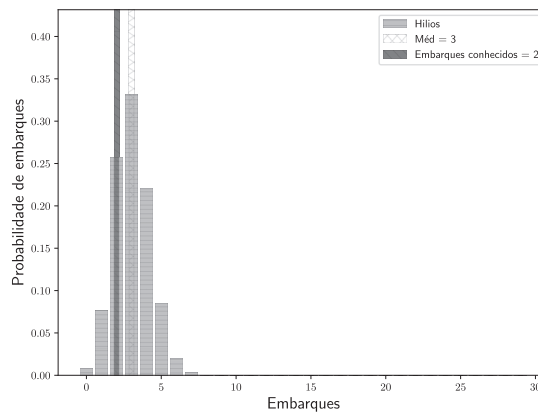
Embarcado indica que há aproximadamente 80% de probabilidade de terem 2 ou 3 detecções associadas a embarques realizados no ponto PT - 2, enquanto na distribuição Nunes (2018) há 80% de probabilidade de ocorrerem 2, 3 ou 4 detecções associadas a embarques realizados no ponto PT - 2. Já para a distribuição Aplicativo há aproximadamente 74% de probabilidade de terem 8, 9 ou 10 detecções associadas a embarques no ponto P2, ou seja, é uma indicação de que podem ter ocorrido no mínimo quatro vezes mais embarques do que realmente ocorreram. Isso é explicado devido à quantidade de detecções consideradas nessa distribuição sendo 7,5 vezes maior do que o que realmente foi gerado.



(a)



(b)



(c)

Figura 34: Estimativa de embarques realizados no ponto de parada de ônibus PT - 2 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

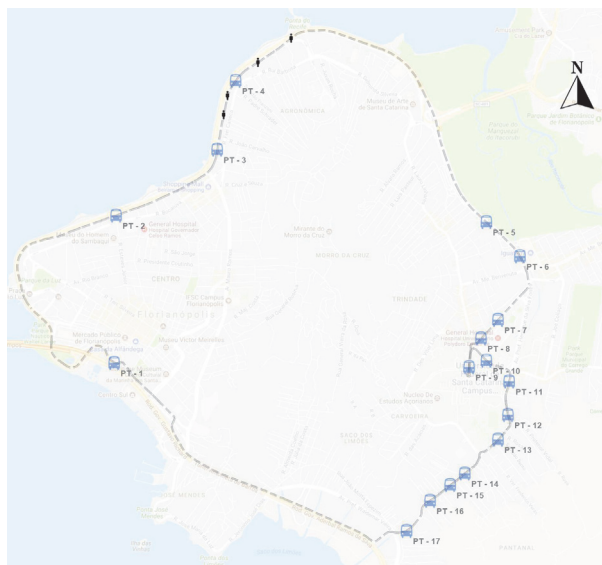
8.2.3 PONTO DE PARADA DE ÔNIBUS PT - 3

O trajeto entre os pontos PT - 3 e PT - 4 é muito curto, isso faz com que possa haver detecções geradas após o ponto 4, mas não após o ponto 5. Isso ocorre para a distribuição Embarcado onde duas detecções foram geradas após o ponto P4 (Figura 35(a)). Para a distribuição Aplicativo foram geradas 3 detecções após o ponto PT - 4 se aproximando muito do ponto P5 (Figura 35(b)). Já para a distribuição Nunes (2018) duas detecções foram geradas após o ponto PT - 4. O ponto PT - 3 da mesma forma que o ponto PT - 2, acaba sofrendo com a influência dos pontos posteriores, inclusive de um ponto mais distante como o PT - 5. Isso pode ser observado pelo número de detecções geradas em PT - 3 somado com o número de detecções geradas de P4. Se a influência fosse somente do ponto PT - 4, o número de detecções consideradas não poderia passar de cinco. Já para o caso das distribuições Aplicativo e Nunes (2018), o número de detecções consideradas é maior que cinco, indicando que há detecções geradas no ponto 5 influenciando no cálculo

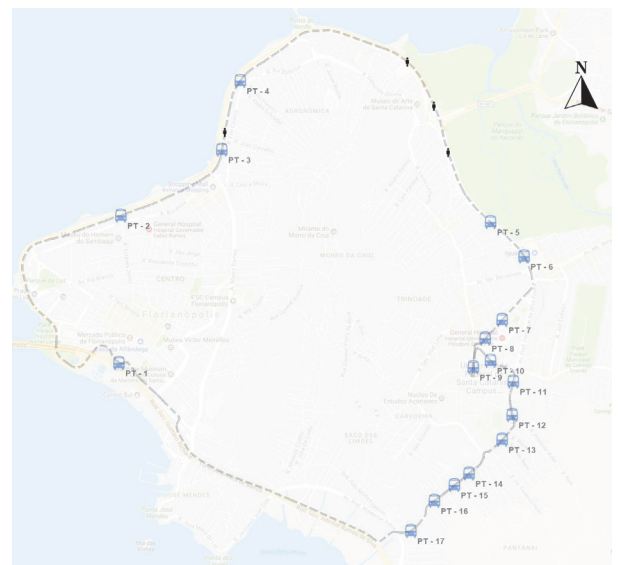
do ponto P3.

Tabela 25: Dados do ponto de parada de ônibus PT - 3

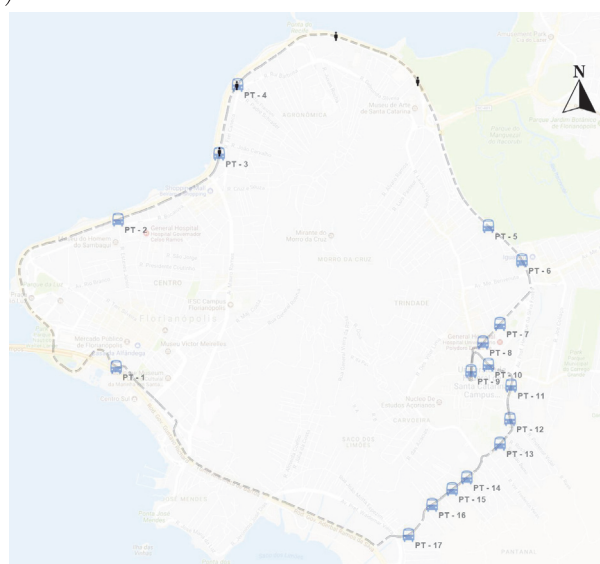
Ponto de parada de ônibus	Distribuição	$2 \times \text{Média}$	Nº Detecções Geradas	Tempo de Viagem PT - 3 \rightarrow PT - 4 (s)	Detecções Consideradas
PT - 3	Embarcado	244			5
	Aplicativo	443	4	77	6
	Nunes (2018)	393			7



(a)



(b)

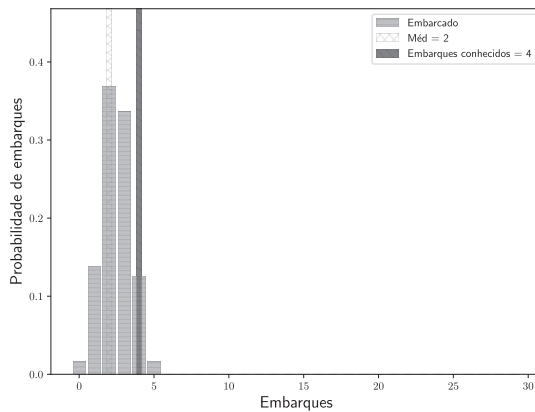


(c)

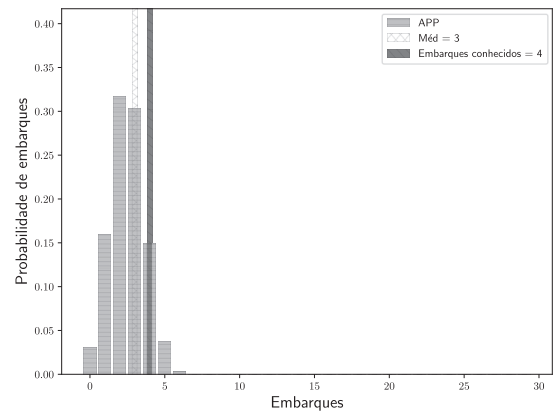
Figura 35: Detecções geradas para 4 embarques no ponto PT - 3. (a) Distribuição Embarcado, (b) Distribuição Aplicativo e (c) Distribuição Nunes (2018).

O método de cálculo aplicado ao ponto P3 é o direto, pelo número de detecções ser menor que 20. O resultado do cálculo é ilustrado na Figura 36. Com nenhuma distribuição foi possível estimar os quatro embarques do ponto P3. A média indica que é possível

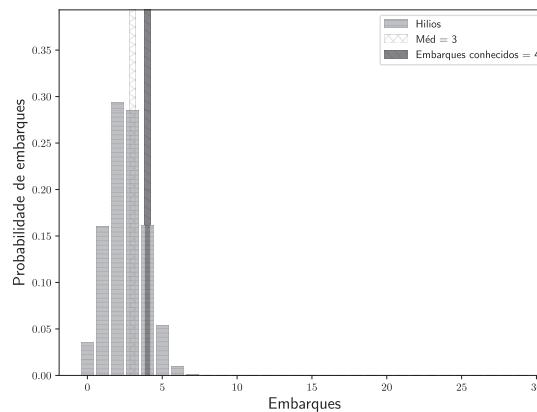
associar de 2 a 3 detecções geradas ao ponto de parada de ônibus PT - 3. O que pode ser observado é que a probabilidade maior em ambas as distribuições está entre 2 e 3 detecções com aproximadamente 70% para a distribuição Embarcado, 61% para a distribuição Aplicativo e 57% para a distribuição Nunes (2018) .



(a)



(b)



(c)

Figura 36: Estimativa de embarques realizados no ponto de parada de ônibus PT - 3 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

8.2.4 PONTO DE PARADA DE ÔNIBUS PT - 4

O ponto PT - 4 pela sua proximidade do ponto P3 tem um comportamento similar. Todos as detecções geradas estão entre o trajeto PT - 4 e PT - 5 para todas as distribuições. Embora houvesse uma chance de ter um embarque gerado após o ponto PT - 5 na distribuição Aplicativo, isso porque duas vezes a média da distribuição Aplicativo é maior que o tempo de trajeto entre PT - 4 e PT - 5, fato esse que não ocorreu. O ponto PT - 4 sofre influência dos pontos P5 e P6, então similar a PT - 3, as detecções geradas em PT - 6 serão consideradas para o cálculo de PT - 4. O cálculo aplicado em todas as distribuições

é o direto, pelo número de detecções ser menor que 20 de acordo com a Tabela 26. O resultado do cálculo é ilustrado na Figura 38..

Tabela 26: Dados do ponto de parada de ônibus PT - 4

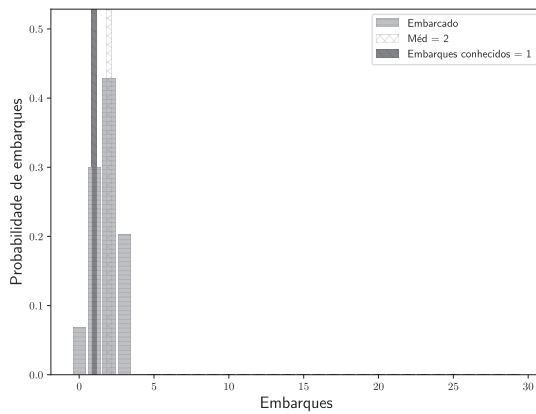
Ponto de parada de ônibus	Distribuição	2×Média	Nº Detecções Geradas	Tempo de Viagem PT - 4 → PT - 5 (s)	Detecções Consideradas
PT - 4	Embarcado	244	1	401	4
	Aplicativo	443			4
	Nunes (2018)	393			8



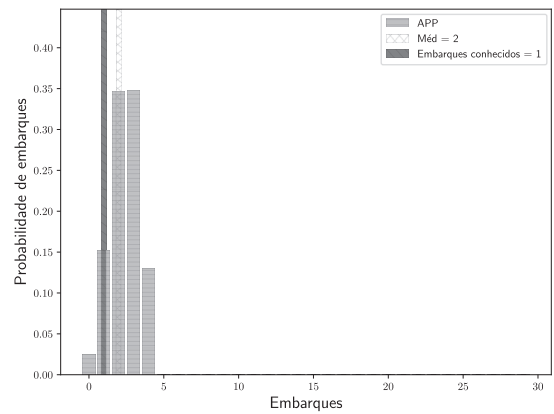
Figura 37: Detecções geradas para 1 embarque no ponto PT - 4. (a) Distribuição Embarcado, (b) Distribuição Aplicativo e (c) Distribuição Nunes (2018)

Os resultados para o ponto PT - 4 na Figura 38(a), que há probabilidade de 88% de que 1, 2 ou 3 detecções geradas para a distribuição Embarcado estão associadas a em-

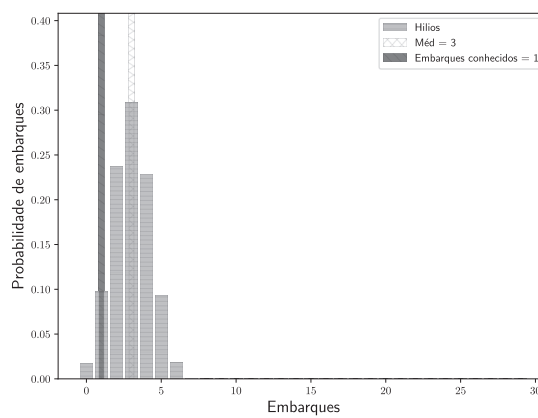
barques no ponto PT - 4. Para a distribuição Aplicativo, na Figura 38(b), a probabilidade é de 68% de que 2 ou 3 detecções geradas sejam associadas a embarques no ponto PT - 4. Já para a distribuição Nunes (2018) na Figura 38(c) a probabilidade é de 75% de que 2, 3 ou 4 detecções geradas sejam associadas a embarques no ponto PT - 4. Sendo que o número de embarques nesse ponto foi somente de um passageiro.



(a)



(b)



(c)

Figura 38: Estimativa de embarques realizados no ponto de parada de ônibus P4 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

8.2.5 DEMAIS PONTOS DE PARADAS DE ÔNIBUS

Como a dinâmica já foi explicada para os pontos PT - 1, PT - 2, PT - 3 e PT - 4, os próximos pontos serão apresentados em sequência com seus resultados. As próximas tabelas apresentam o tipo de cálculo realizado.

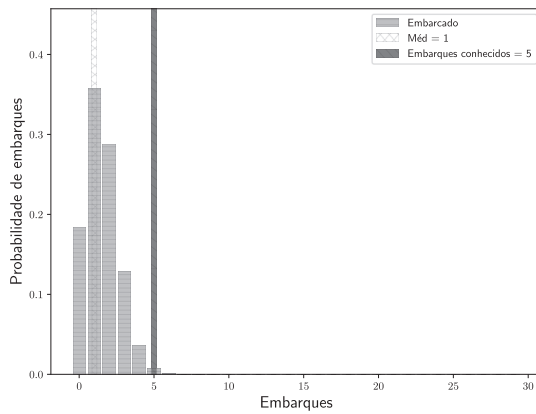
Nos cálculos realizados pelo método de classes é possível observar que a convolução das distribuições de probabilidades discretas fazem com que a curva seja mais suave e que uma forma Gaussiana começa a surgir quanto maior for o número de classes. O teorema matemático que explica isso é o Teorema do Limite Central, onde a soma de n variáveis aleatórias independentes é uma aproximação pela distribuição Gaussiana se n for grande. Isso nada mais é do que a distribuição Normal (OSGOOD, 2013).

Esse efeito será observado em pontos de parada com muitos embarques considerados, onde o número de classes é grande como no cálculo dos embarques por zonas na Seção 8.3.

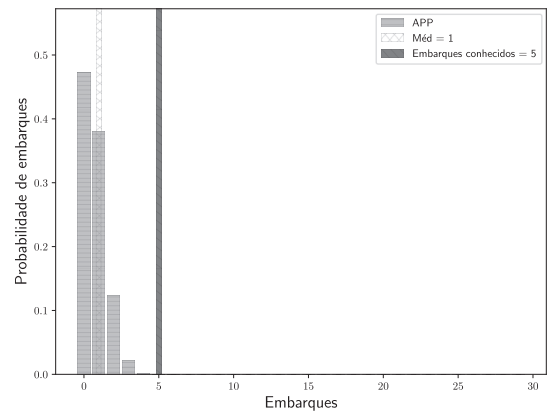
Para o ponto P5 há em média 1 detecção de dispositivos móveis na distribuição Embarcado, 1 na Aplicativo e 4 na Nunes (2018) que podem ser associadas a passageiros que tenham embarcado nesse ponto. O método de cálculo, assim como os resultados, estão apresentados na Tabela 27 e Figura 39 respectivamente.

Tabela 27: Dados do ponto de parada de ônibus PT - 5

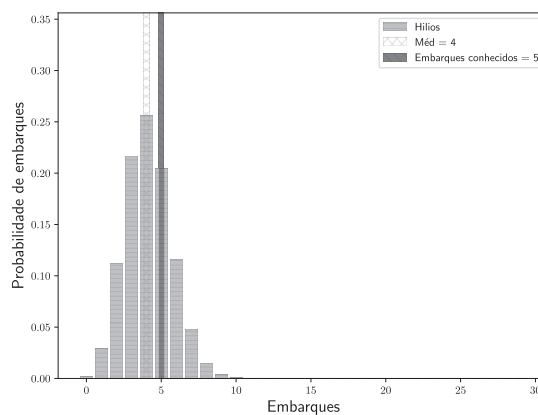
Ponto de parada de ônibus	Distribuição	Tipo de Cálculo	$2 \times$ Média	Nº Detecções Geradas	Tempo de Viagem PT - 5 \rightarrow PT - 6 (s)	Detecções Consideradas
PT - 5	Embarcado	Classes	244	5	64	27
	Aplicativo	Classes	443			21
	Nunes (2018)	Direto	393			20



(a)



(b)



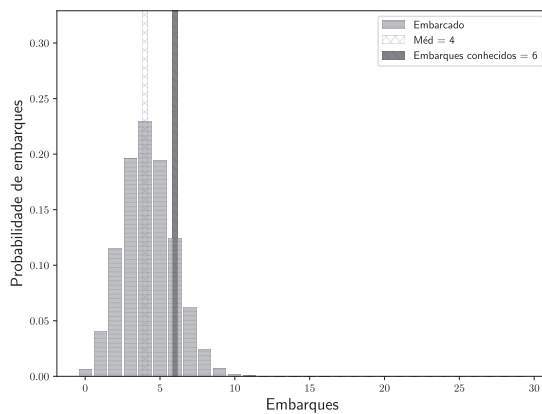
(c)

Figura 39: Estimativa de embarques realizados no ponto de parada de ônibus PT - 5 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

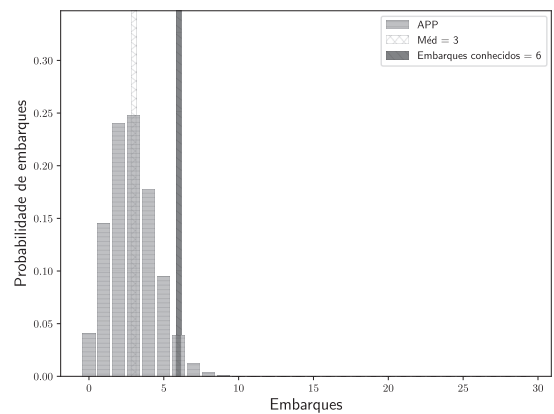
Para o ponto PT - 6 há em média 4 detecções de dispositivos móveis na distribuição Embarcado, 3 na Aplicativo e 5 na Nunes (2018) que podem ser associadas a passageiros que tenham embarcado nesse ponto. O método de cálculo, assim como os resultados, estão apresentados na Tabela 28 e Figura 40 respectivamente.

Tabela 28: Dados do ponto de parada de ônibus PT - 6

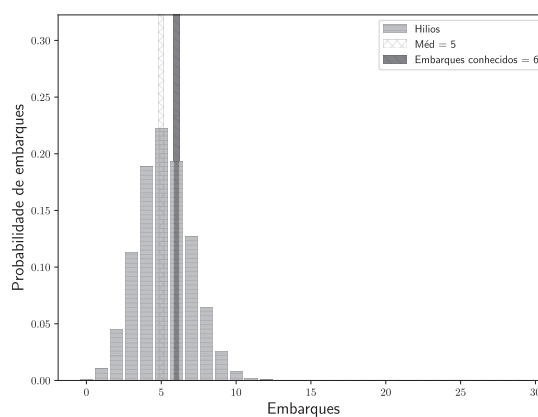
Ponto de parada de ônibus	Distribuição	Tipo de Cálculo	$2 \times$ Média	Nº Detecções Geradas	Tempo de Viagem PT - 6 \rightarrow PT - 7 (s)	Detecções Consideradas
PT - 6	Embarcado	Classes	244	6	101	48
	Aplicativo	Classes	443			35
	Nunes (2018)	Classes	393			33



(a)



(b)



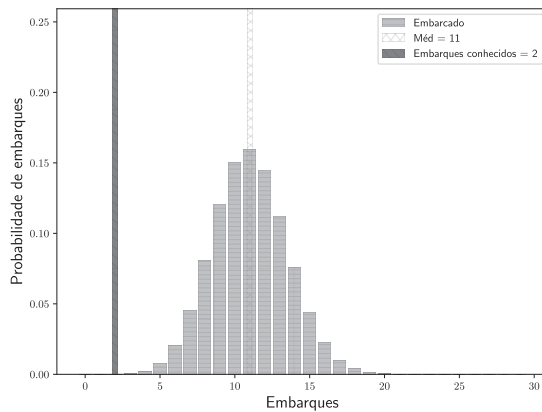
(c)

Figura 40: Estimativa de embarques realizados no ponto de embarque PT - 6 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

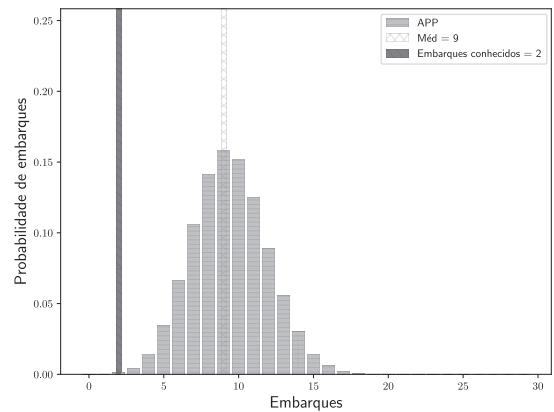
Para o ponto PT - 7 há em média 11 detecções de dispositivos móveis na distribuição Embarcado, 9 na Aplicativo e 9 na Nunes (2018) que podem ser associadas a passageiros que tenham embarcado nesse ponto. O método de cálculo, assim como os resultados, estão apresentados na Tabela 29 e Figura 41 respectivamente.

Tabela 29: Dados do ponto de parada de ônibus PT - 7

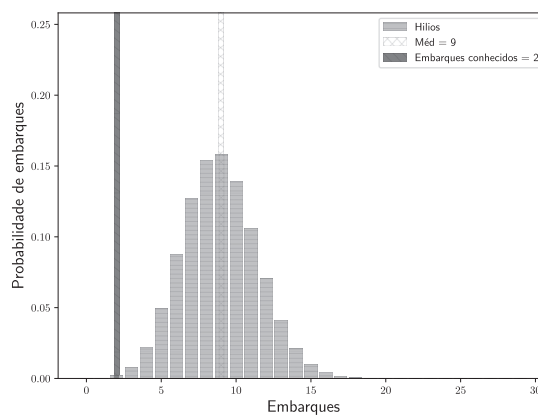
Ponto de parada de ônibus	Distribuição	Tipo de Cálculo	$2 \times$ Média	Nº Detecções Geradas	Tempo de Viagem PT - 7 \rightarrow PT - 8 (s)	Detecções Consideradas
PT - 7	Embarcado	Classes	244	2	48	78
	Aplicativo	Classes	443			68
	Nunes (2018)	Classes	393			66



(a)



(b)



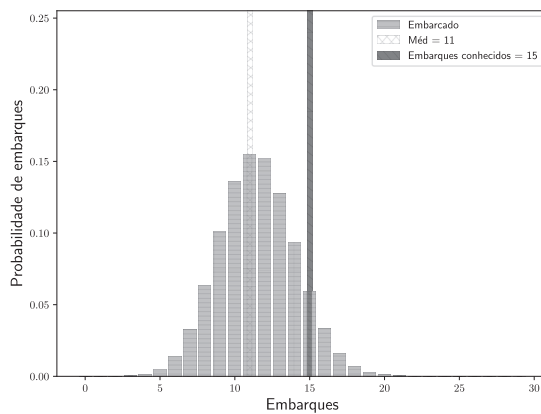
(c)

Figura 41: Estimativa de embarques realizados no ponto de parada de ônibus PT - 7 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

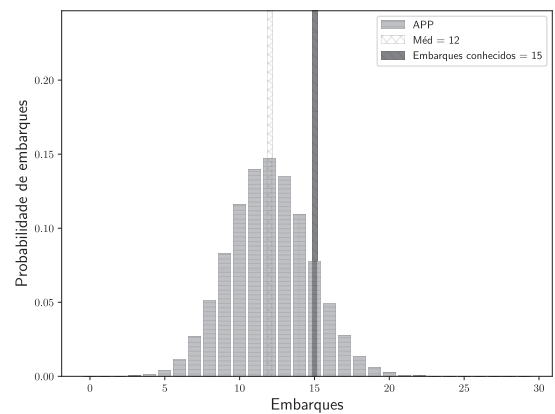
Para o ponto PT - 8 há em média 11 detecções de dispositivos móveis na distribuição Embarcado, 12 na Aplicativo e 9 na Nunes (2018) que podem ser associadas a passageiros que tenham embarcado nesse ponto. O método de cálculo, assim como os resultados, estão apresentados na Tabela 30 e Figura 42 respectivamente.

Tabela 30: Dados do ponto de parada de ônibus PT - 8

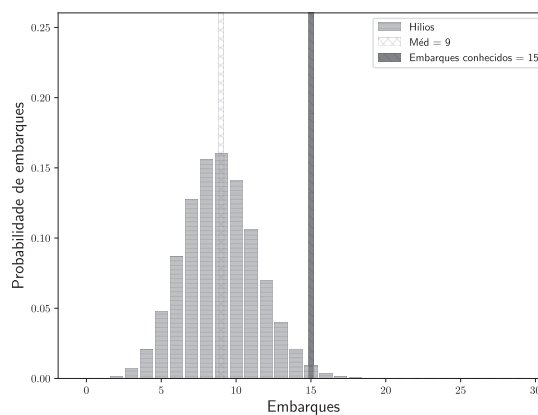
Ponto de parada de ônibus	Distribuição	Tipo de Cálculo	$2 \times$ Média	Nº Detecções Geradas	Tempo de Viagem PT - 8 \rightarrow PT - 9 (s)	Detecções Consideradas
PT - 8	Embarcado	Classes	244	15	162	76
	Aplicativo	Classes	443			80
	Nunes (2018)	Classes	393			72



(a)



(b)



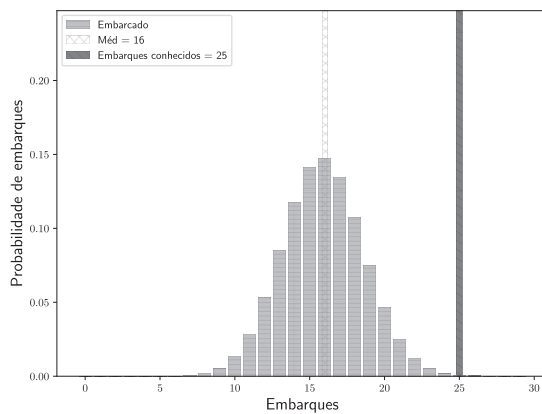
(c)

Figura 42: Estimativa de embarques realizados no ponto de parada de ônibus PT - 8 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

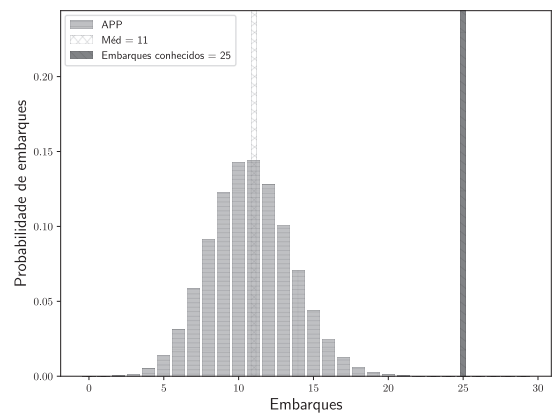
Para o ponto PT - 9 há em média 16 detecções de dispositivos móveis na distribuição Embarcado, 11 na Aplicativo e 13 na Nunes (2018) que podem ser associadas a passageiros que tenham embarcado nesse ponto. O método de cálculo, assim como os resultados, estão apresentados na Tabela 31 e Figura 43 respectivamente.

Tabela 31: Dados do ponto de parada de ônibus PT - 9

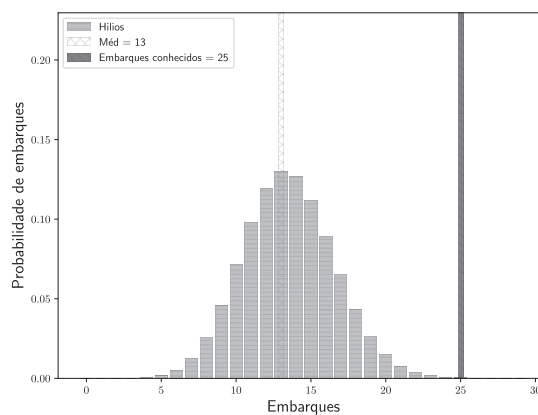
Ponto de parada de ônibus	Distribuição	Tipo de Cálculo	$2 \times \text{Média}$	Nº Detecções Geradas	Tempo de Viagem PT - 9 \rightarrow PT - 10 (s)	Detecções Consideradas
PT - 9	Embarcado	Classes	244	25	108	75
	Aplicativo	Classes	443			78
	Nunes (2018)	Classes	393			79



(a)



(b)



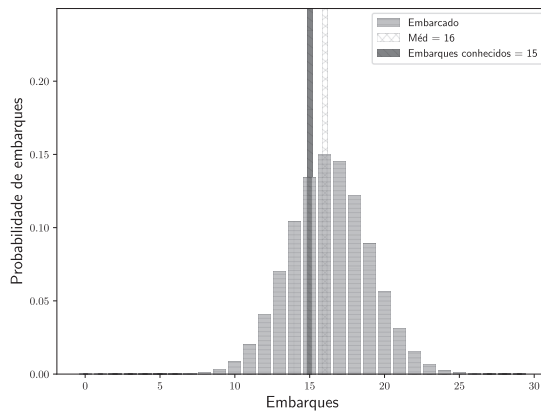
(c)

Figura 43: Estimativa de embarques realizados no ponto de parada de ônibus PT - 9 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

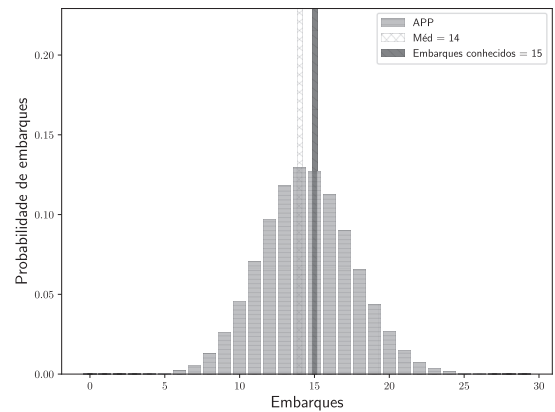
Para o ponto PT - 10 há em média 16 detecções de dispositivos móveis na distribuição Embarcado, 14 na Aplicativo e 17 na Nunes (2018) que podem ser associadas a passageiros que tenham embarcado nesse ponto. O método de cálculo, assim como os resultados, estão apresentados na Tabela 32 e Figura 44 respectivamente.

Tabela 32: Dados do ponto de parada de ônibus PT - 10

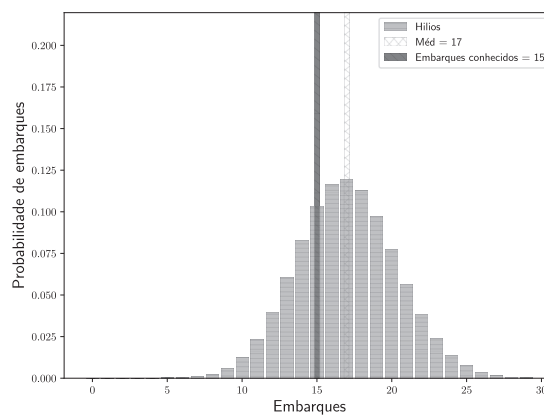
Ponto de parada de ônibus	Distribuição	Tipo de Cálculo	$2 \times$ Média	Nº Detecções Geradas	Tempo de Viagem PT - 10 \rightarrow PT - 11 (s)	Detecções Consideradas
PT - 10	Embarcado	Classes	244	15	100	64
	Aplicativo	Classes	443			80
	Nunes (2018)	Classes	393			75



(a)



(b)



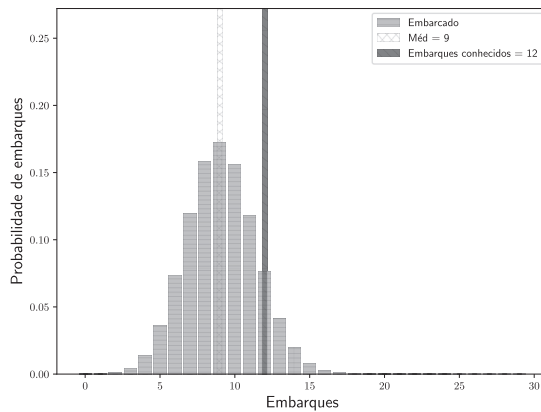
(c)

Figura 44: Estimativa de embarques realizados no ponto de parada de ônibus PT - 10 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

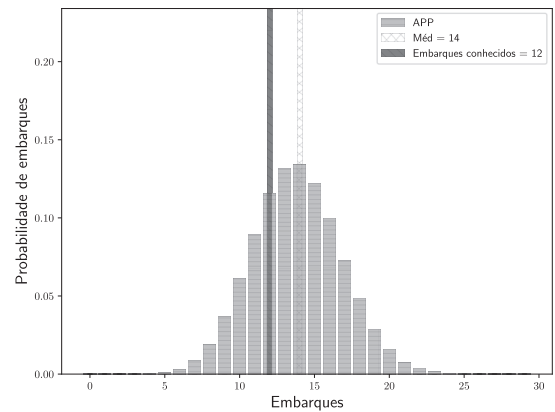
Para o ponto PT - 11 há em média 9 detecções de dispositivos móveis na distribuição Embarcado, 14 na Aplicativo e 14 na Nunes (2018) que podem ser associadas a passageiros que tenham embarcado nesse ponto. O método de cálculo, assim como os resultados, estão apresentados na Tabela 33 e Figura 45 respectivamente.

Tabela 33: Dados do ponto de parada de ônibus PT - 11

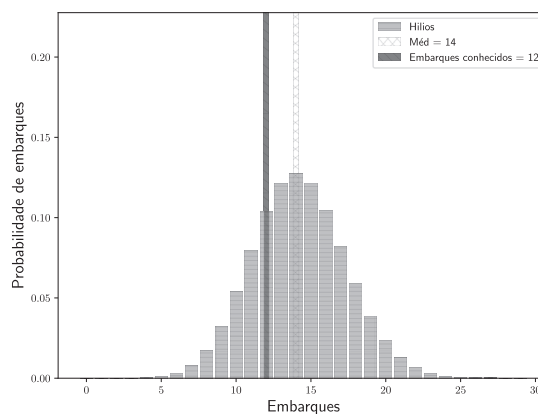
Ponto de parada de ônibus	Distribuição	Tipo de Cálculo	$2 \times$ Média	Nº Detecções Geradas	Tempo de Viagem PT - 11 \rightarrow PT - 12 (s)	Detecções Consideradas
PT - 11	Embarcado	Classes	244	12	95	52
	Aplicativo	Classes	443			70
	Nunes (2018)	Classes	393			65



(a)



(b)



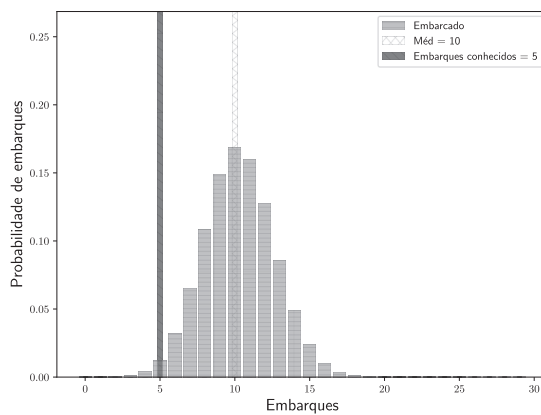
(c)

Figura 45: Estimativa de embarques realizados no ponto de parada de ônibus PT - 11 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

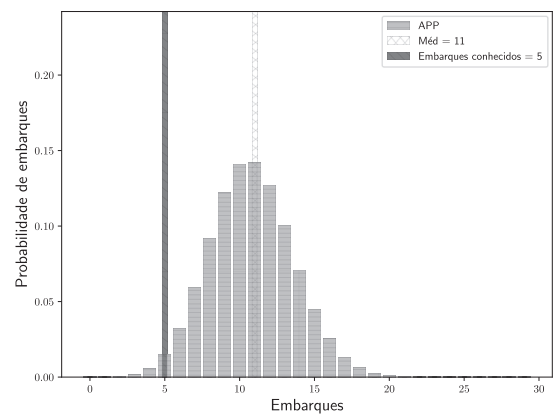
Para o ponto PT - 12 há em média 10 detecções de dispositivos móveis na distribuição Embarcado, 11 na Aplicativo e 12 na Nunes (2018) que podem ser associadas a passageiros que tenham embarcado nesse ponto. O método de cálculo, assim como os resultados, estão apresentados na Tabela 34 e Figura 46 respectivamente.

Tabela 34: Dados do ponto de parada de ônibus PT - 12

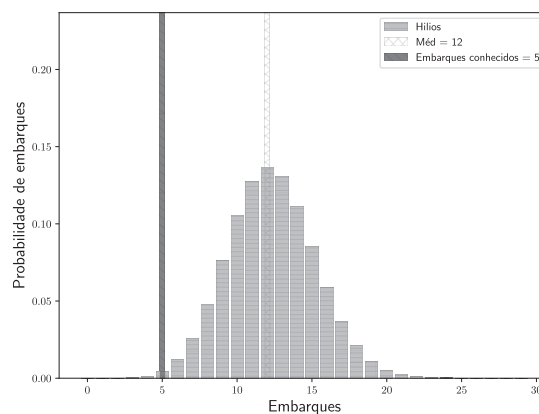
Ponto de parada de ônibus	Distribuição	Tipo de Cálculo	$2 \times$ Média	Nº Detecções Geradas	Tempo de Viagem PT - 12 \rightarrow PT - 13 (s)	Detecções Consideradas
PT - 12	Embarcado	Classes	244	5	80	44
	Aplicativo	Classes	443			56
	Nunes (2018)	Classes	393			54



(a)



(b)



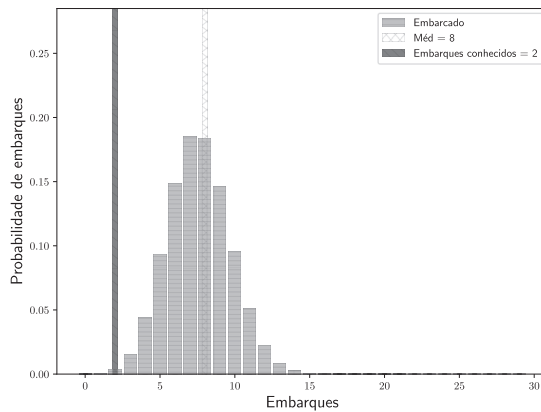
(c)

Figura 46: Estimativa de embarques realizados no ponto de parada de ônibus PT - 12 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

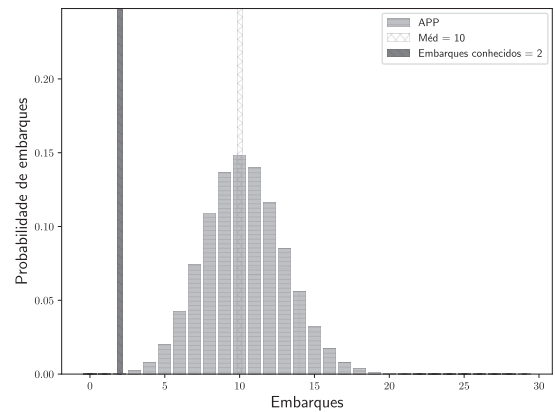
Para o ponto PT - 13 há em média 8 detecções de dispositivos móveis na distribuição Embarcado, 10 na Aplicativo e 6 na Nunes (2018) que podem ser associadas a passageiros que tenham embarcado nesse ponto. O método de cálculo, assim como os resultados, estão apresentados na Tabela 35 e Figura 47 respectivamente.

Tabela 35: Dados do ponto de parada de ônibus PT - 13

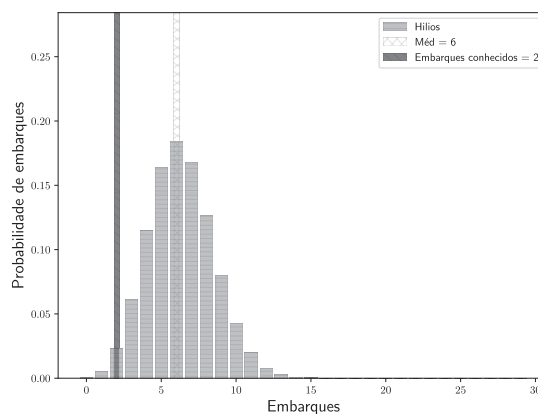
Ponto de parada de ônibus	Distribuição	Tipo de Cálculo	$2 \times$ Média	Nº Detecções Geradas	Tempo de Viagem PT - 13 \rightarrow PT - 14 (s)	Detecções Consideradas
PT - 13	Embarcado	Classes	244	2	138	33
	Aplicativo	Classes	443			47
	Nunes (2018)	Classes	393			36



(a)



(b)



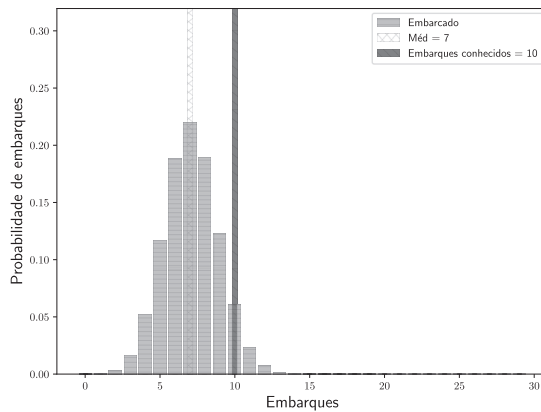
(c)

Figura 47: Estimativa de embarques realizados no ponto de parada de ônibus PT - 13 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

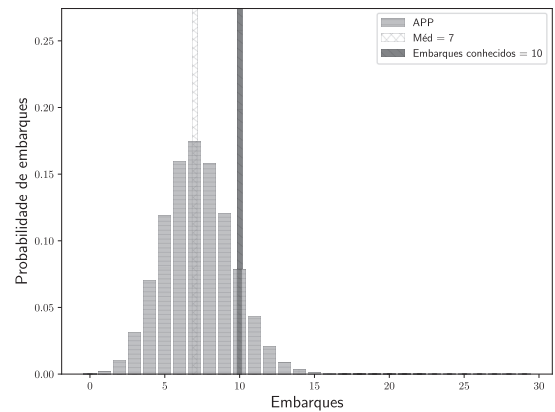
Para o ponto PT - 14 há em média 7 detecções de dispositivos móveis na distribuição Embarcado, 7 na Aplicativo e 6 na Nunes (2018) que podem ser associadas a passageiros que tenham embarcado nesse ponto. O método de cálculo, assim como os resultados, estão apresentados na Tabela 36 e Figura 48 respectivamente.

Tabela 36: Dados do ponto de parada de ônibus PT - 14

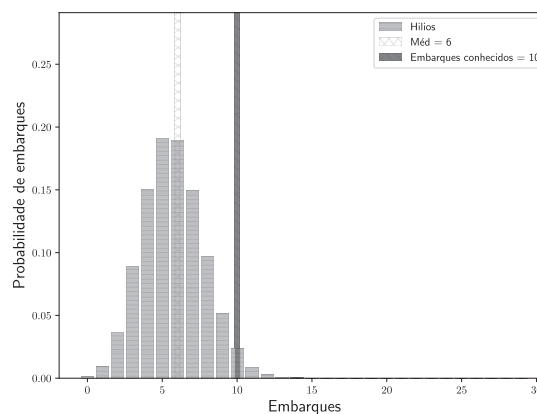
Ponto de parada de ônibus	Distribuição	Tipo de Cálculo	$2 \times$ Média	Nº Detecções Geradas	Tempo de Viagem PT - 14 \rightarrow PT - 15 (s)	Detecções Consideradas
PT - 14	Embarcado	Classes	244	10	52	24
	Aplicativo	Classes	443			32
	Nunes (2018)	Classes	393			28



(a)



(b)



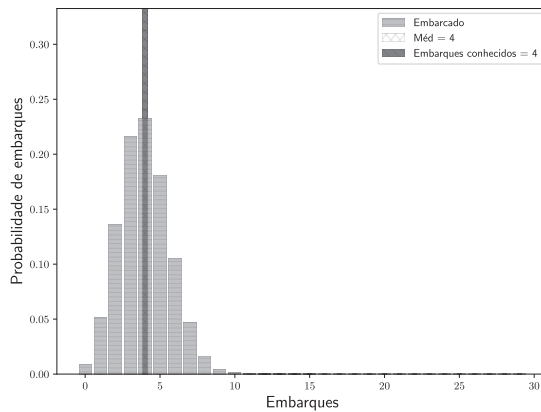
(c)

Figura 48: Estimativa de embarques realizados no ponto de parada de ônibus PT - 14 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

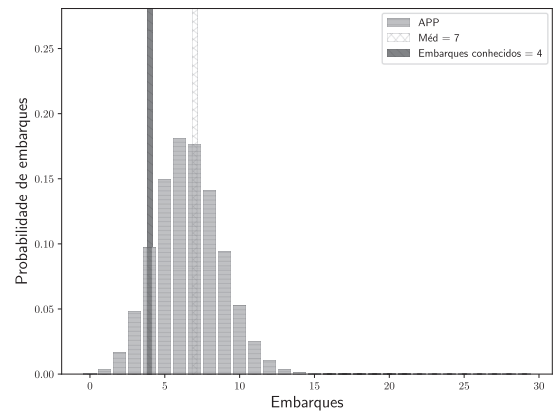
Para o ponto PT - 15 há em média 4 detecções de dispositivos móveis na distribuição Embarcado, 7 na Aplicativo e 5 na Nunes (2018) que podem ser associadas a passageiros que tenham embarcado nesse ponto. O método de cálculo, assim como os resultados, estão apresentados na Tabela 37 e Figura 49 respectivamente.

Tabela 37: Dados do ponto de parada de ônibus PT - 15

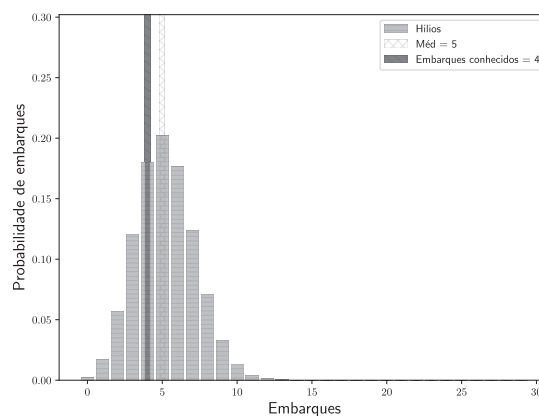
Ponto de parada de ônibus	Distribuição	Tipo de Cálculo	$2 \times$ Média	Nº Detecções Geradas	Tempo de Viagem PT - 15 \rightarrow PT - 16 (s)	Detecções Consideradas
PT - 15	Embarcado	Direto	244	4	70	17
	Aplicativo	Classes	443			27
	Nunes (2018)	Classes	393			25



(a)



(b)



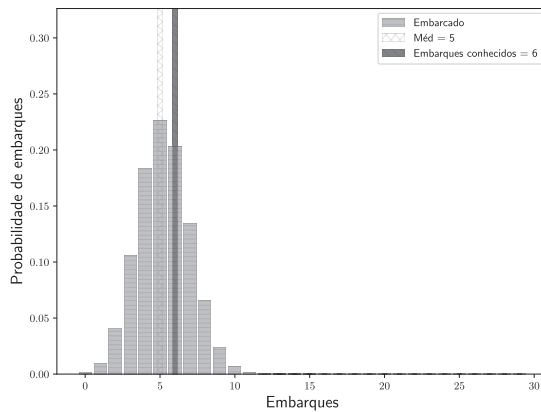
(c)

Figura 49: Estimativa de embarques realizados no ponto de parada de ônibus PT - 15 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

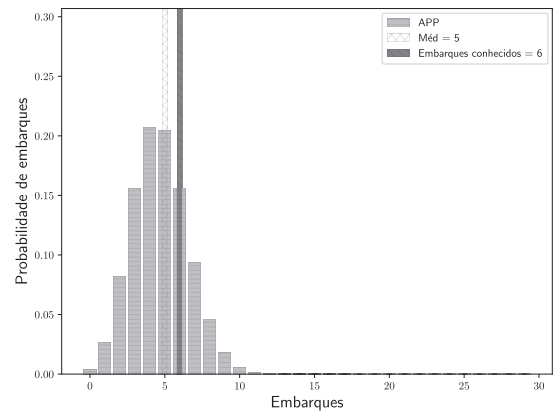
Para o ponto PT - 16 há em média 5 detecções de dispositivos móveis na distribuição Embarcado, 5 na Aplicativo e 5 na Nunes (2018) que podem ser associadas a passageiros que tenham embarcado nesse ponto. O método de cálculo, assim como os resultados, estão apresentados na Tabela 38 e Figura 50 respectivamente.

Tabela 38: Dados do ponto de parada de ônibus PT - 16

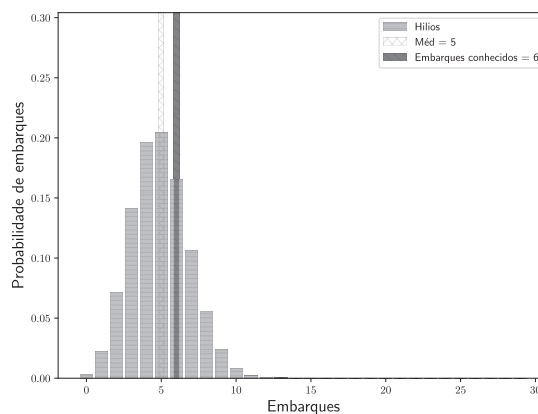
Ponto de parada de ônibus	Distribuição	Tipo de Cálculo	$2 \times$ Média	Nº Detecções Geradas	Tempo de Viagem PT - 16 \rightarrow PT - 17 (s)	Detecções Consideradas
PT - 16	Embarcado	Direto	244	6	110	15
	Aplicativo	Direto	443			20
	Nunes (2018)	Classes	393			21



(a)



(b)



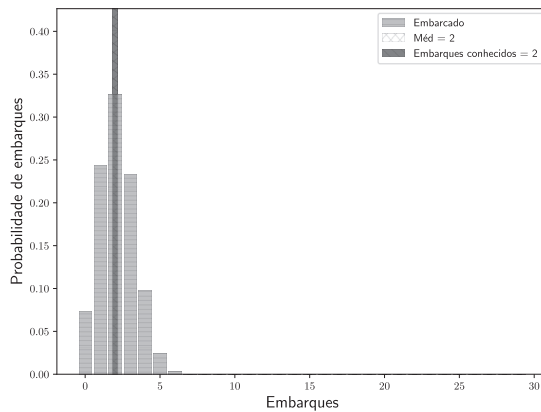
(c)

Figura 50: Estimativa de embarques realizados no ponto de parada de ônibus PT - 16 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

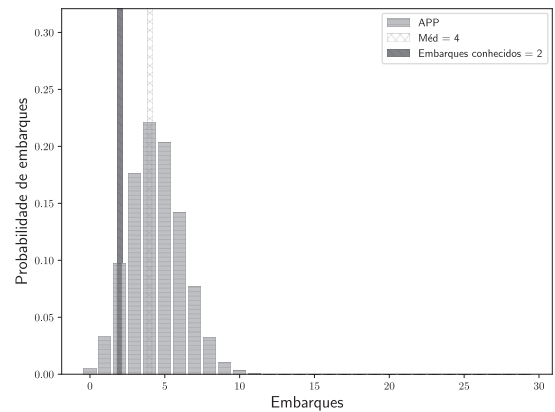
Para o ponto PT - 17 há em média 2 detecções de dispositivos móveis na distribuição Embarcado, 4 na Aplicativo e 3 na Nunes (2018) que podem ser associadas a passageiros que tenham embarcado nesse ponto. O método de cálculo, assim como os resultados, estão apresentados na Tabela 39 e Figura 51 respectivamente.

Tabela 39: Dados do ponto de parada de ônibus PT - 17

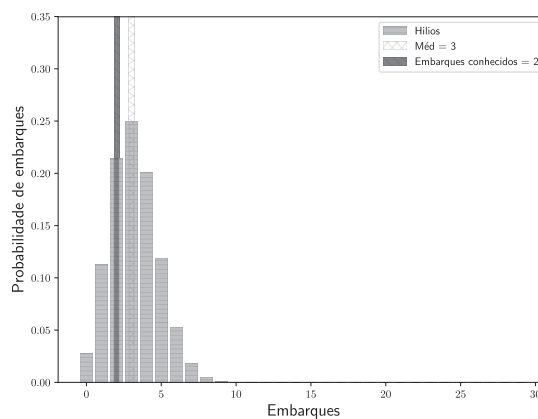
Ponto de parada de ônibus	Distribuição	Tipo de Cálculo	$2 \times \text{Média}$	Nº Detecções Geradas	Tempo de Viagem PT - 17 \rightarrow PT - 1 (s)	Detecções Consideradas
PT - 17	Embarcado	Direto	244	2	396	7
	Aplicativo	Direto	443			16
	Nunes (2018)	Direto	393			15



(a)



(b)



(c)

Figura 51: Estimativa de embarques realizados no ponto de parada de ônibus PT - 17 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

8.2.6 SÍNTESE DOS RESULTADOS

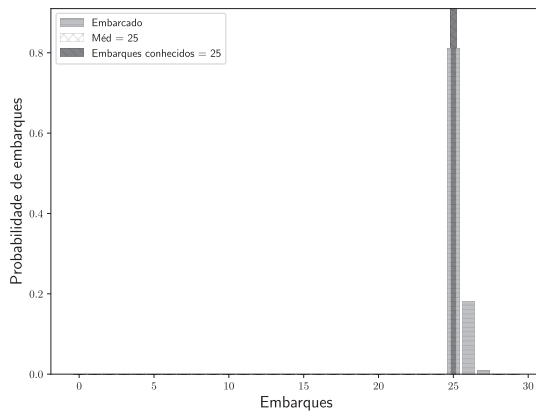
A Tabela 40 ilustra os resultados dos embarques estimados para todos os pontos de embarques. É possível observar através da tabela que para todas as distribuições, o total de embarques somando todos os pontos de embarques não é igual aos 141 embarques conhecidos. Para a distribuição Embarcado o total de embarques em média foram 136 embarques, enquanto para as distribuições Aplicativo e Nunes (2018) o total de embarques é 140. Olhando essa informação agregada passa uma impressão de que foi possível estimar os embarques com 99% de precisão para as distribuições Aplicativo e Nunes (2018) e com 96% de precisão para a distribuição Embarcado. Como já comentado, os dados agregados podem passar essa falsa impressão. Quando os dados desagregados são analisados ponto a ponto, as diferenças se tornam mais visíveis. Há casos em que a estimativa foi de mais passageiros do que realmente embarcaram no ponto. Por isso a análise individual é importante quando se quer uma estimativa do carregamento ponto a ponto.

Tabela 40: Tabela com as médias das estimativas de embarques de cada ponto, com erro relativo aos embarques conhecidos.

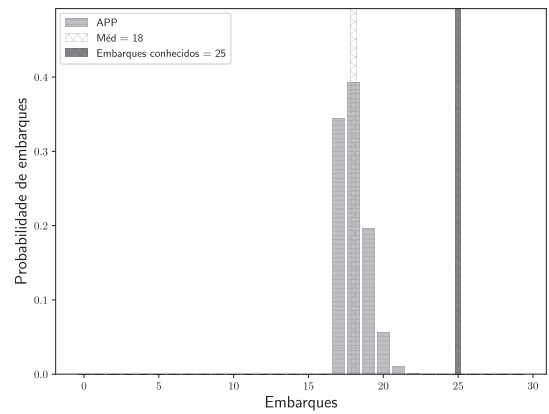
Ponto	Embarques Conhecidos	Distribuição					
		Embarcado		Aplicativo		Nunes (2018)	
		Média	Erro	Média	Erro	Média	Erro
PT - 1	25	25	0	18	7	23	2
PT - 2	2	3	-1	9	-7	3	-1
PT - 3	4	2	2	3	1	3	1
PT - 4	1	2	-1	2	-1	3	-2
PT - 5	5	1	4	1	4	4	1
PT - 6	6	4	2	3	3	5	1
PT - 7	2	11	-9	9	-7	9	-7
PT - 8	15	11	4	12	3	9	6
PT - 9	25	16	9	11	14	13	12
PT - 10	15	16	-1	14	1	17	-2
PT - 11	12	9	3	14	-2	14	-2
PT - 12	5	10	-5	11	-6	12	-7
PT - 13	2	8	-6	10	-8	6	-4
PT - 14	10	7	3	7	3	6	4
PT - 15	4	4	0	7	-3	5	-1
PT - 16	6	5	1	5	1	5	1
PT - 17	2	2	0	4	-2	3	-1

8.3 RESULTADOS DO CÁLCULO DA ESTIMATIVA DO TOTAL DE EMBARQUES EM ZONAS DE EMBARQUE

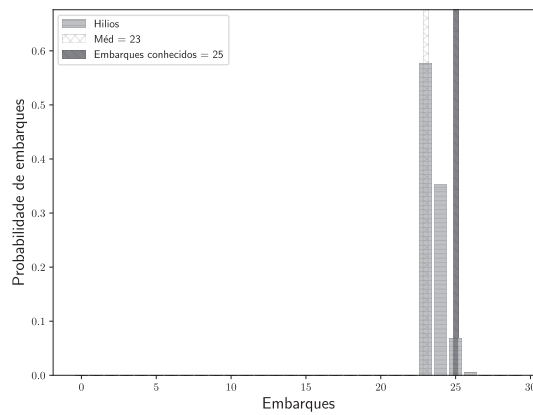
As zonas de embarques estão definidas na Figura 25. É possível observar que como o ponto de embarque PT - 1 e PT - 2 não foram agregados com outros pontos para formação da sua zona, os resultados das estimativas de embarques são os mesmos da Figura 32 e 34, e estão representados no formato das zonas na Figura 52.



(a)



(b)



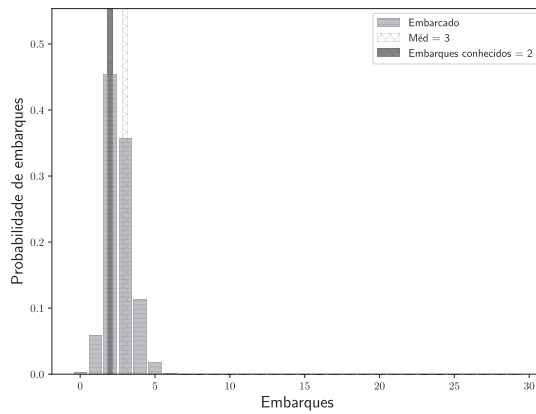
(c)

Figura 52: Estimativa de embarques realizados na zona de embarque Z1 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

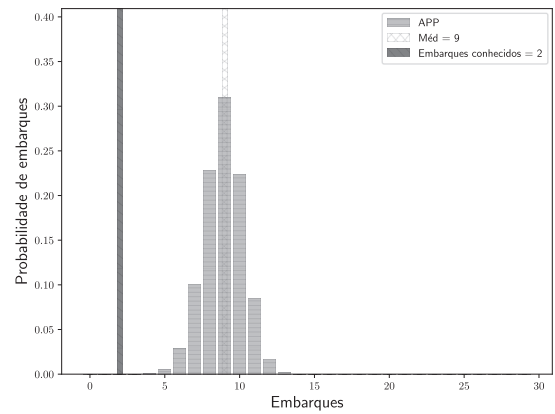
O método de cálculo aplicado para as demais zonas de embarques é o de classes, isso por que o agrupamento de pontos faz com que o número de probabilidades de embarques seja muito alto para realizar o cálculo direto. É possível observar nos cálculos seguintes através das Figuras 54, 55, 56 e 57, o teorema do limite central onde com muitas probabilidades, ou seja, n muito alto as curvas tendem a uma normal. A Tabela 41 apresenta o resumo das estimativas de embarques por zonas. Como citado anteriormente as zonas Z1 e Z2 não se alteram por não ter pontos agrupados.

Nos embarques agregados, os resultados indicam que as distribuições Aplicativo e Nunes (2018) passam a estimar todos os 141 embarques, enquanto quando analisado por ponto, faltava um embarque. Já para a distribuição Embarcado, mesmo com o agrupamento por zona não foi possível estimar os 141 embarques, contando com 138 embarques.

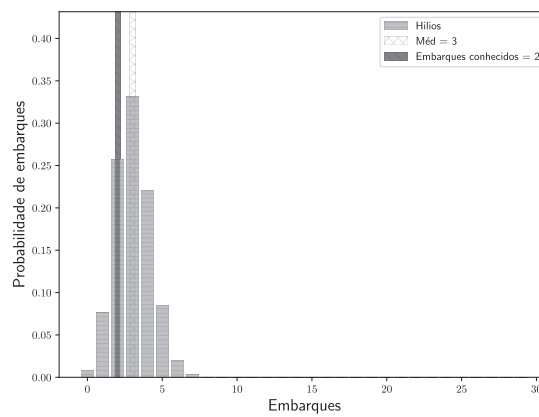
Analisando individualmente cada zona excluindo a zona Z1 e Z2 por já terem sido analisadas anteriormente, é possível observar que somente a zona Z4 apresentou um número menor de embarques para todas as distribuições, enquanto nas outras zonas ao



(a)



(b)



(c)

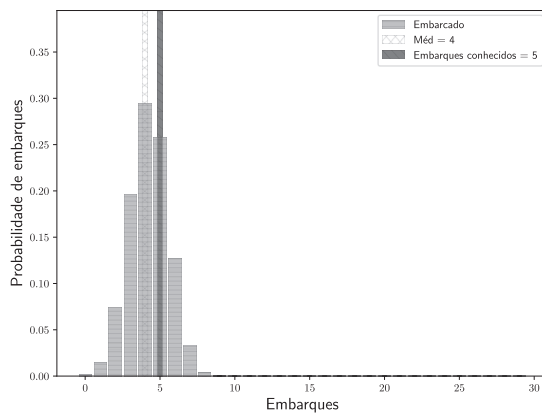
Figura 53: Estimativa de embarques realizados na zona de embarque Z2 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

menos uma distribuição apresentou um número maior de embarques, esses são chamados os falsos positivos, ou seja, embarques atribuídos a um ponto ao qual não pertence.

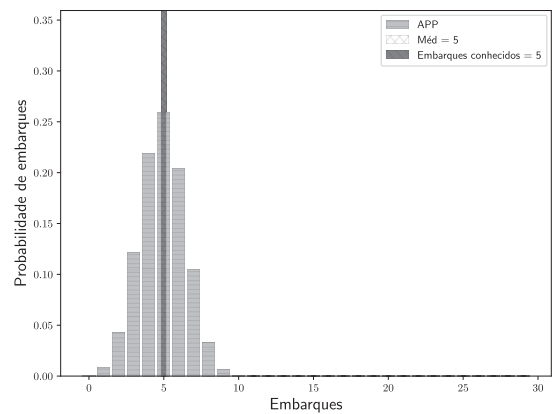
Mas podemos observar também que a forma de agregação da zona pode melhorar as estimativas de embarques. Por exemplo, ao agregar os embarques da Z1 com a Z2 agora teríamos 27 embarques conhecidos, que é a soma dos embarques da Z1 e Z2. Dessa forma podemos observar que para a distribuição Embarcado o número de embarques agregados agora, passa a ser 28, somente 1 embarque a mais do que o conhecido. Já para a distribuição Aplicativo, o erro que era muito presente agora é suprimido, com um total de 27 embarques, exatamente igual aos embarques conhecidos. Isso conclui que a forma de agregação das zonas também pode impactar nos resultados.

Tabela 41: Tabela com as médias das estimativas de embarques de cada zona.

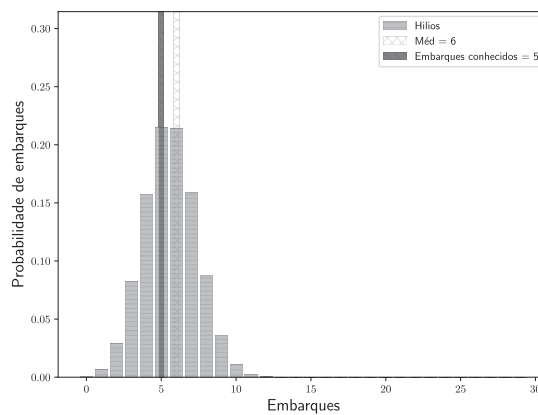
Ponto	Embarques Conhecidos	Média de Embarques					
		Distribuição				Nunes (2018)	
		Embarcado		Aplicativo		Estimativa	Erro(%)
Estimativa	Erro (%)	Estimativa	Erro(%)				
Z1	25	25	0	18	28	23	8
Z2	2	3	-50	9	-350	3	-50
Z3	5	4	20	5	0	6	-20
Z4	11	6	45	4	64	9	18
Z5	76	82	-8	82	8	81	-7
Z6	22	18	18	23	5	19	14
Total	141	138	2	141	0	141	0



(a)

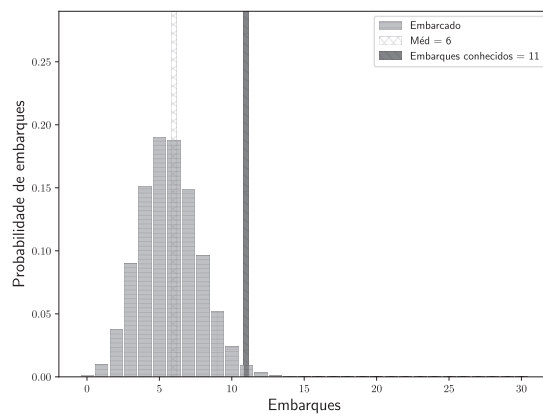


(b)

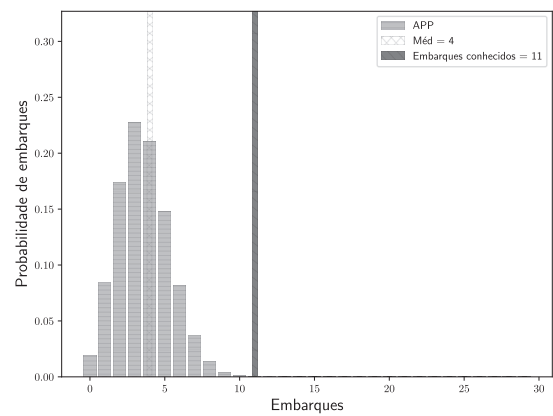


(c)

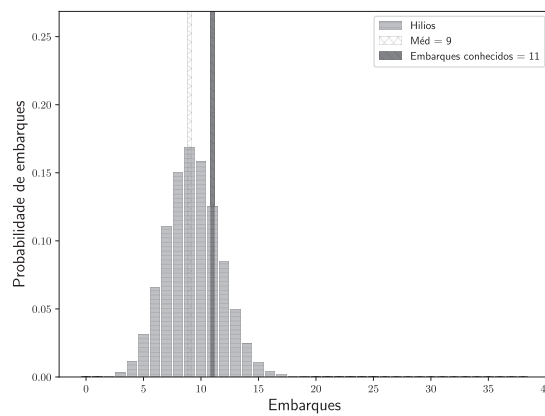
Figura 54: Estimativa de embarques realizados na zona de embarque Z3 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).



(a)

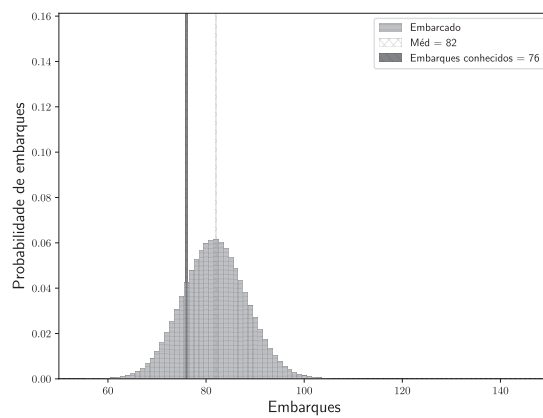


(b)

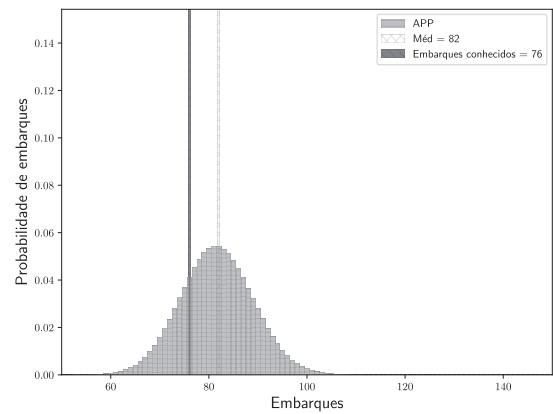


(c)

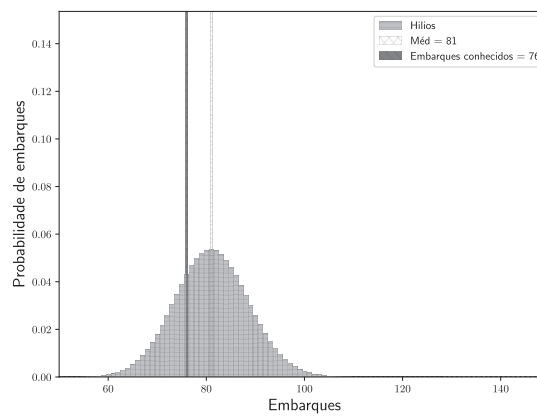
Figura 55: Estimativa de embarques realizados na zona de embarque Z4 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).



(a)

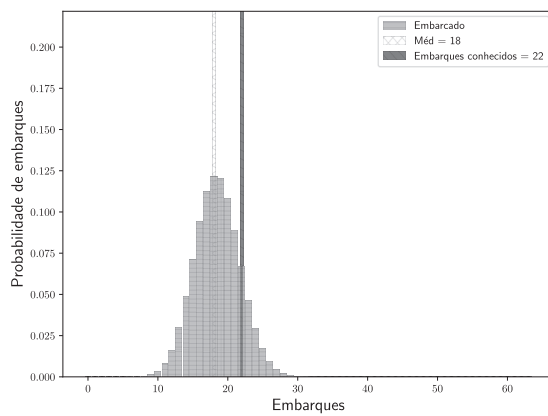


(b)

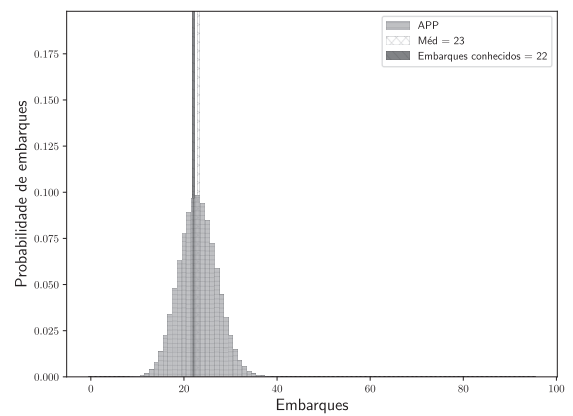


(c)

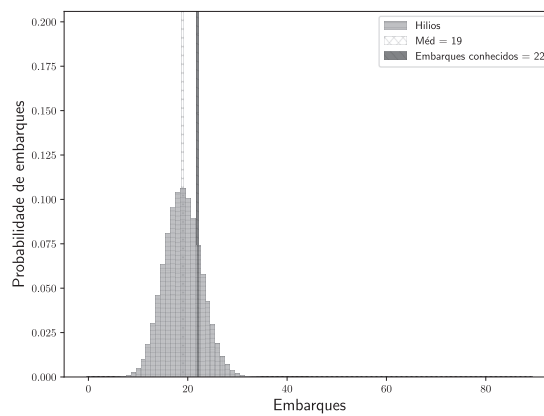
Figura 56: Estimativa de embarques realizados na zona de embarque Z5 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).



(a)



(b)



(c)

Figura 57: Estimativa de embarques realizados na zona de embarque Z6 a partir de detecções geradas. (a) Distribuição experimento Embarcado, (b) Distribuição experimento Aplicativo e (c) Distribuição experimento Nunes (2018).

9 CONCLUSÃO

Com o intuito de reduzir custos de pesquisas, este trabalho buscou trazer uma forma de buscar informações sobre os embarques e desembarques de passageiros com intuito de futuramente atualizar o modelo de transporte coletivo, com dados mais precisos. Com isso a pesquisa se concentrou em uma tecnologia que está disponível para grande parte da população do Brasil, que é o sistema Wi-Fi. Muitos estudos estão sendo realizados nessa área e buscou-se investigar a real viabilidade de implantação desse sistema como um método de contagem de embarques e desembarques de passageiros.

Através de dispositivos móveis como *smartphones* portados por passageiros que utilizam o sistema de transporte coletivo por ônibus, foi verificado que é possível utilizar um dispositivo de detecção aqui chamado de NTM para coletar as buscas de redes realizadas por dispositivos móveis embarcados, e que as perdas de buscas de redes que deveriam ser detectadas é de aproximadamente 1%.

A partir das informações coletadas e processadas, foi possível verificar que é possível associar um dispositivo móvel portado por uma pessoa há um passageiro embarcado em um ônibus.

Apesar de muitas pesquisas indicarem que há uma grande probabilidade de se estimar os embarques somente com base na informação do dado coletado, a pesquisa realizada neste trabalho mostrou o contrário, e indicou que é preciso cautela quando se caracteriza uma detecção Wi-Fi como um embarque de passageiro em um ponto de ônibus. Em alguns casos o erro pode passar de 100%.

As matrizes geradas por outras pesquisas podem ser superestimadas, e considerar detecções de dispositivos que poderiam não estar dentro do ônibus.

O grande problema encontrado não está na detecção do dispositivo portado pelo passageiro, o problema está em duas questões pontuais: a primeira é se o passageiro realmente está com o Wi-Fi do dispositivo móvel ligado; para isso não há solução, e a outra é taxa com que o dispositivo móvel desse passageiro realiza uma busca de rede. Cada vez mais com atualizações pelos desenvolvedores dos sistemas operacionais, esse aspecto se torna um ponto crítico. As taxas de buscas de redes realizadas por dispositivos móveis estão cada vez menos frequentes, isso porque a busca pela economia de bateria se torna o foco principal e o Wi-Fi é um dos principais fatores do aumento de consumo de bateria.

Através de uma investigação sobre o comportamento das buscas de redes realizadas pelos dispositivos móveis, foi identificado que a aleatoriedade com que essas buscas são realizadas causam grandes problemas de estimação. Esses problemas são desde um dispositivo móvel do passageiro realizar uma busca de rede a cada hora, e se considerarmos que em condições ideais de tráfego um passageiro não ficaria mais de uma hora em um ônibus, esse passageiro seria descartado, ou até mesmo se o dispositivo móvel desse passageiro realizar somente uma busca de rede ao longo de uma viagem inteira, como é preciso mais

detecções para verificar se realmente ele poderia estar no ônibus, nessa condição também seria descartado.

A aplicação do método estatístico proposto indicou que a estimativa dos embarques pode ser realizada desde que o comportamento das buscas de redes sejam mais frequentes. O método quando aplicado em dados reais, conseguiu estimar de certa forma os embarques, mesmo com as buscas de redes serem muito aleatórias. Quando foi aplicado um comportamento mais regular na frequência de buscas de redes (dados não apresentados aqui) o método se mostrou 100% eficaz, acertando todos os embarques dos passageiros.

Por outro lado, como não há um comportamento regular nas buscas de redes, outra solução encontrada foi o agrupamento de pontos de embarques em zonas de embarques. Com esse método foi possível observar que o método se torna mais eficiente, e que os erros individuais por cada ponto diminuem muito.

Uma solução para contornar o problema das buscas de redes, seria disponibilizar pontos de acesso Wi-Fi em todos os ônibus, e garantir de alguma forma que haja conexão entre os dispositivos móveis Wi-Fi dos passageiros com o ponto de acesso. Isso iria garantir que não fariam as buscas de redes ficarem frequentes, mas ao invés do quadro Requisição de Sondagem, seriam analisadas as trocas de pacotes TCP que são mais frequentes quando há conexão estabelecida.

Outro ponto importante é a individualização dos dados, o ideal é que cada passageiro tenha sua própria distribuição de probabilidade individual, fazendo com que ao final de cada viagem, a estimativa seja única daquele passageiro. Como os resultados mostraram o agrupamento de informações pode levar a interpretações equivocadas.

As pesquisas podem avançar considerando dois pontos, o primeiro é considerar melhorar a probabilidade a priori do método estatístico proposto. Inserindo dados reais como dados de bilhetagem podem melhorar a probabilidade que no método é igual para todos os pontos. Outro ponto é com relação a formação das zonas de embarque. A forma como as zonas podem ser criadas podem gerar impactos significativos nos resultados, então estudar as melhores formas de agrupar os pontos nas zonas podem ser um ponto importante a ser verificado.

Novas fontes de informação também podem ser consideradas. Ao invés da utilização do Wi-Fi de dispositivos móveis, há possibilidade de utilizar o próprio sinal de rede móvel do dispositivo móvel. Sabe-se hoje que cada dispositivo móvel é registrado em antenas de celulares através de um número único no mundo chamado de CellID. É possível desenvolver um dispositivo que seja capaz de capturar esse CellID e assim rastrear a localização do dispositivo móvel. As operadoras de telefonia celular só registram eventos, como uma ligação, recebimento de SMS ou conexão 3G/4G, eles não guardam a informação do registro na antena. Isso poderia ser uma solução porque bastaria o smartphone estar ligado que poderia ser rastreado.

REFERÊNCIAS

ABEDI, Naeim; BHASKAR, Ashish; CHUNG, Edward. Bluetooth and Wi-Fi MAC address based crowd data collection and ng: benefits, challenges and enhancement, 2013.

ACER, Utku Günay; WALTARI, Otto. WiPush: Opportunistic Notifications over WiFi Without Association. *In: PROCEEDINGS of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. [S.l.: s.n.], 2017. P. 353–362.

ALEXANDER, Lauren; JIANG, Shan; MURGA, Mikel; GONZÁLEZ, Marta C. Origin–destination trips by purpose and time of day inferred from mobile phone data. **Transportation research part c: emerging technologies**, Elsevier, v. 58, p. 240–250, 2015.

ALLEN, Lee; HERIYANTO, Tedi; ALI, Shakeel. **Kali Linux–Assuring security by penetration testing**. [S.l.]: Packt Publishing Ltd, 2014.

ANATEL, Agência Nacional de Telecomunicações. Resolução nº 680, de 27 de junho de 2017. **Diário Oficial da União**, 2017.

BARBOSA, Alexandre F. Pesquisa sobre o uso das tecnologias da informação e da comunicação no Brasil: TIC Domicílios. **São Paulo: Comitê Gestor da Internet no Brasil**, 2018.

BARBOSA, Alexandre F. Pesquisa sobre o uso das tecnologias da informação e da comunicação no Brasil: TIC Domicílios. **São Paulo: Comitê Gestor da Internet no Brasil**, 2019.

BAUER, Blaine. **Network traffic monitoring**. [S.l.]: Google Patents, 2004. US Patent App. 10/236,402.

BEGGS, Robert W. **Mastering Kali Linux for advanced penetration testing**. [S.l.]: Packt Publishing Ltd, 2014.

BEKHOR, Shlomo; COHEN, Yehoshua; SOLOMON, Charles. Evaluating long-distance travel patterns in Israel by tracking cellular phone positions. **Journal of Advanced Transportation**, Wiley Online Library, v. 47, n. 4, p. 435–446, 2013.

CALIPER, C. Transcad 8.0. **Caliper: Caliper Corporation**, p. 593, 2020.

CARVALHO, Ana Cristina Azevedo P. **Marco civil da internet no Brasil: análise da Lei 12,965/14 e do direito de informação**. [S.l.]: Alta Books Editora, 2014.

CASCIATI, F; CASCIATI, S; CHEN, ZC; FARAVELLI, L; VECE, M. Collecting data from a sensor network in a single-board computer. *In*: IOP PUBLISHING, 1. JOURNAL of Physics: Conference Series. [S.l.: s.n.], 2015. P. 012113.

CEDER, Avishai. **Public transit planning and operation: Modeling, practice and behavior**. [S.l.]: CRC press, 2016.

ÇOLAK, Serdar; ALEXANDER, Lauren P; ALVIM, Bernardo G; MEHNDIRATTA, Shomik R; GONZÁLEZ, Marta C. Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. **Transportation Research Record**, SAGE Publications Sage CA: Los Angeles, CA, v. 2526, n. 1, p. 126–135, 2015.

DUNLAP, Matthew; LI, Zhibin; HENRICKSON, Kristian; WANG, Yin Hai. Estimation of origin and destination information from Bluetooth and Wi-Fi sensing for transit. **Transportation Research Record**, SAGE Publications Sage CA: Los Angeles, CA, v. 2595, n. 1, p. 11–17, 2016.

FERRAZ, Antonio Clovis Pinto; TORRES, Isaac Guillermo Espinosa. **Transporte público urbano**. [S.l.]: RiMa Editora, 2001.

FREUDIGER, Julien. How talkative is your mobile device? An experimental study of Wi-Fi probe requests. *In*: PROCEEDINGS of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks. [S.l.: s.n.], 2015. P. 1–6.

GARFINKEL, Simson L; MCCARRIN, Michael. Can We Sniff Wi-Fi?: Implications of Joffe v. Google. **IEEE Security & Privacy**, IEEE, v. 12, n. 4, p. 22–28, 2014.

GOOGLE, Inc. **Wi-Fi scanning overview**. 2018. Disponível em: <https://developer.android.com/guide/topics/connectivity/wifi-scan>.

HÅKEGÅRD, Jan Erik; MYRVOLL, Tor Andre; SKOGLUND, Tor Rune. Statistical modelling for estimation of OD matrices for public transport using Wi-Fi and APC data.

In: IEEE. 2018 21st International Conference on Intelligent Transportation Systems (ITSC). [*S.l.: s.n.*], 2018. P. 1005–1010.

HERRERA-QUINTERO, Luis Felipe; BANSE, Klaus; VEGA-ALFONSO, Julián; VENEGAS-SANCHEZ, Andrés. Smart ITS sensor for the transportation planning using the IoT and Bigdata approaches to produce ITS cloud services. *In*: IEEE. 2016 8th Euro American Conference on Telematics and Information Systems (EATIS). [*S.l.: s.n.*], 2016. P. 1–7.

HUTCHENS, Justin. **Kali Linux network scanning cookbook**. [*S.l.*]: Packt Publishing Ltd, 2014.

IEEE. IEEE Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. **IEEE Std 802.11[^]**, 2007. DOI: 10.1109/IEEESTD.2007.373646.

JAMES, K; KEITH, R. **Redes de Computadores e a Internet: Uma abordagem top-down**. [*S.l.*]: São Paulo: Pearson Addison Wesley Gilligan R, 2005.

JI, Yuxiong; ZHAO, Jizhou; ZHANG, Zhiming; DU, Yuchuan. Estimating bus loads and OD flows using location-stamped farebox and Wi-Fi signal data. **Journal of Advanced Transportation**, Hindawi, v. 2017, 2017.

JIANG, Shan; FERREIRA, Joseph; GONZALEZ, Marta C. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. **IEEE Transactions on Big Data**, IEEE, v. 3, n. 2, p. 208–219, 2017.

JUNIOR, Marcos Paulino Roriz; MEDRANO, Ronny Aliagra. UAI-FI: UM MÉTODO BASEADO EM APRENDIZADO DE MÁQUINA PARA CONTAGEM AUTOMÁTICA DE PASSAGEIROS UTILIZANDO SINAIS WI-FI, 2018.

KORTEKAAS, Karst. Global mobile market report. Newzoo, 2018.

KOTLER, Philip; KARTAJAYA, Hermawan; SETIAWAN, Iwan. **Marketing 4.0: do tradicional ao digital**. [*S.l.*]: Sextante, 2017.

KUMKAR, Vishal; TIWARI, Akhil; TIWARI, Pawan; GUPTA, Ashish; SHRAWNE, Seema. Vulnerabilities of Wireless Security protocols (WEP and WPA2).

International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), v. 1, n. 2, p. 34–38, 2012.

KUROSE, James F; ROSS, Keith W. *Redes de Computadores e a Internet*. **São Paulo: Person**, p. 28, 2006.

LIND, Artjom; HADACHI, Amnir; BATRASHEV, Oleg. A new approach for mobile positioning using the CDR data of cellular networks. *In: IEEE. 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. [S.l.: s.n.], 2017. P. 315–320.

LIU, Jiayu; GAO, Jun; YU, Jiahui; HUANG, Yunjie; HUANG, Kun. The Design and Implementation of the Customer Traffic Statistics Analysis System Based on Wi-Fi Sniffing Technology. *In: IEEE. 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE)*. [S.l.: s.n.], 2019. P. 1274–1277.

LIU, Yang; LU, Dian-Jie; LI, Xing-Yue. The Remote Wifi Investigation of Network and Forensics System for Police Based on Raspberry Pi. *In: ATLANTIS PRESS. 3RD International Conference on Wireless Communication and Sensor Networks (WCSN 2016)*. [S.l.: s.n.], 2016.

LÓPEZ, Alberto Acosta; MONROY, Elver Yesid Melo; MURCIA, Pablo Andrés Linares. Evaluation of the WPA2-PSK wireless network security protocol using the Linset and Aircrack-ng tools. **Facultad de Ingeniería**, Universidad Pedagógica y Tecnológica de Colombia (UPTC), v. 27, n. 47, p. 71–78, 2018.

MANNING, Trevor. **Microwave radio transmission design guide**. [S.l.]: Artech house, 2009.

MARTIN, Jeremy; MAYBERRY, Travis; DONAHUE, Collin; FOPPE, Lucas; BROWN, Lamont; RIGGINS, Chadwick; RYE, Erik C; BROWN, Dane. A study of MAC address randomization in mobile devices and when it fails. **Proceedings on Privacy Enhancing Technologies**, Sciendo, v. 2017, n. 4, p. 365–383, 2017.

MATTE, Célestin; CUNCHE, Mathieu; ROUSSEAU, Franck; VANHOEF, Mathy. Defeating MAC address randomization through timing attacks. *In: PROCEEDINGS of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks*. [S.l.: s.n.], 2016. P. 15–20.

MEIRELLES, FS. 30^a Pesquisa Anual do Uso de TI nas Empresas. **FGVcia: Centro e Tecnologia de Informação Aplicada da EAESP**, 2019.

MUNIZ, Joseph; LAKHANI, Aamir. **Penetration testing with raspberry pi**. [S.l.]: Packt Publishing Ltd, 2015.

NAJERA-GUTIERREZ, Gilberto; ANSARI, Juned Ahmed. **Web Testing with Kali Linux: Explore the methods and tools of ethical hacking with Kali Linux**. [S.l.]: Packt Publishing Ltd, 2018.

NEWZOO. **Top Countries by Smartphone Users**. [S.l.: s.n.], 2018.
<https://newzoo.com/insights/rankings/top-countries-by-smartphone-penetration-and-users/>. [Acesso em Outubro-2018].

NITTI, Michele; PINNA, Francesca; PINTOR, Lucia; PILLONI, Virginia; BARABINO, Benedetto. iABACUS: A Wi-Fi-Based Automatic Bus Passenger Counting System. **Energies**, Multidisciplinary Digital Publishing Institute, v. 13, n. 6, p. 1446, 2020.

NUNES, Edson Hilios Marques. **Estimativa da ocupação utilizando sensoriamento wi-fi de telefones móveis: uma aplicação no transporte público urbano de passageiros por ônibus**. 2018. Tese (Doutorado) – Universidade de São Paulo.

O’FLAHERTY, Coleman A. **Transport planning and traffic engineering**. [S.l.]: CRC Press, 2018.

OSGOOD, Brad. Lecture notes for EE 261: the Fourier transform and its applications. **Electrical Engineering Department, Stanford University**, 2013.

PAOLO, Galluzzi; LONGO, Edoardo; REDONDI, Alessandro Enrico Cesare; CESANA, Matteo. Occupancy Estimation Using Low-Cost Wi-Fi Sniffers. *In*: 2ND Balkancomm international conference. [S.l.: s.n.], 2019. P. 0–0.

PARADEDADA, Diego Benites. **Arquivos dos experimentos e código dos aplicativos**. [S.l.: s.n.], 2018b. <https://github.com/paradedada/ScanWifi>.

PARADEDADA, Diego Benites; JUNIOR, Werner Kraus; CARLSON, Rodrigo Castelan. Bus passenger counts using Wi-Fi signals: some cautionary findings. **TRANSPORTES**, v. 27, n. 3, p. 115–130, 2019.

PARADEDADA, Diego Benites; JUNIOR, Werner Kraus; CARLSON, Rodrigo Castelan. Estimativa de embarque e desembarque de passageiros por detecção de sinal de Wi-Fi de dispositivos Móveis. *In: XXXII Congresso de Pesquisa e Ensino em Transportes*. [S.l.: s.n.], 2018. P. 1048–1059.

PIXEL RESEARCH LABS, Inc. **Tally Counter**. [S.l.: s.n.], 2018. <https://apps.apple.com/us/app/tally-counter/id288732372>. [Acesso em Outubro-2018].

RA, IEEE. **Guidelines for Use of Extended Unique Identifier (EUI), Organizationally Unique Identifier (OUI), and Company ID (CID)**. [S.l.]: pdf, Tech. Rep., August 2017.[Online]. Available: [https://standards.ieee ...](https://standards.ieee...), 2018.

RUSKEY, Frank; WILLIAMS, Aaron. The coolest way to generate combinations. en. **Discrete Mathematics**, v. 309, n. 17, p. 5305–5320, set. 2009. ISSN 0012365X. DOI: 10.1016/j.disc.2007.11.048. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0012365X07009570>. Acesso em: 18 jun. 2020.

RYU, Seunghan; PARK, Byungkyu Brian; EL-TAWAB, Samy. WiFi Sensing System for Monitoring Public Transportation Ridership: A Case Study. **KSCE Journal of Civil Engineering**, Springer, p. 1–13, 2020.

SEPEHRDAD, Pouyan; SUŠIL, Petr; VAUDENAY, Serge; VUAGNOUX, Martin. Smashing WEP in a passive attack. *In: SPRINGER. INTERNATIONAL Workshop on Fast Software Encryption*. [S.l.: s.n.], 2013. P. 155–178.

SMIRNOV, Maksim; SPIRICHEVA, Nataliya; SMIRNOVA, Valentina. Network Sniffer for Time Tracking. *In: IEEE. 2020 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT)*. [S.l.: s.n.], 2020. P. 0471–0474.

STATCOUNTER. **Mobile Operating System Market Share Worldwide**. [S.l.: s.n.], 2020. <https://gs.statcounter.com/os-market-share/androidwifi>. [Acesso em Outubro-2020].

SUTHERS, Elliott. **Wi-Fi device shipments to surpass 15 billion by end of 2016**. [S.l.: s.n.], 2018. <https://www.wi-fi.org/news-events/newsroom/wi-fi-device-shipments-to-surpass-15-billion-by-end-of-2016>. [Acesso em Outubro-2016].

EL-TAWAB, Samy; ORAM, Raymond; GARCIA, Michael; JOHNS, Chris; PARK, B Brian. Poster: Monitoring transit systems using low cost WiFi technology. *In: IEEE. 2016 IEEE Vehicular Networking Conference (VNC)*. [S.l.: s.n.], 2016. P. 1–2.

TSE, David; VISWANATH, Pramod. **Fundamentals of wireless communication**. [S.l.]: Cambridge university press, 2005.

TSUKANOV, Roman. **Counter**. [S.l.: s.n.], 2018.

<https://play.google.com/store/apps/details?id=me.tsukanov.counter>. [Acesso em Outubro-2018].

URAS, Marco; COSSU, Raimondo; FERRARA, Enrico; LIOTTA, Antonio; ATZORI, Luigi. PmA: A real-world system for people mobility monitoring and analysis based on Wi-Fi probes. **Journal of Cleaner Production**, Elsevier, p. 122084, 2020.

VANHOEF, Mathy; MATTE, Célestin; CUNCHE, Mathieu; CARDOSO, Leonardo S; PIESSENS, Frank. Why MAC address randomization is not enough: An analysis of Wi-Fi network discovery mechanisms. *In: PROCEEDINGS of the 11th ACM on Asia Conference on Computer and Communications Security*. [S.l.: s.n.], 2016. P. 413–424.

VASILEVSKI, Ivan; BLAZHEVSKI, Dobro; PACHOVSKI, Veno; STOJMENOVSKA, Irena. Five Years Later: How Effective Is the MAC Randomization in Practice? The No-at-All Attack. *In: SPRINGER. INTERNATIONAL Conference on ICT Innovations*. [S.l.: s.n.], 2019. P. 52–64.

VENDORS, MAC. **Find MAC Address Vendors. Now**. [S.l.: s.n.], 2020.

<https://macvendors.com/>. [Acesso em Outubro-2020].

VUCHIC, Vukan R. **Urban transit: operations, planning, and economics**. [S.l.]: John Wiley & Sons, 2017.

WALIULLAH, Md; MONIRUZZAMAN, ABM; RAHMAN, Md Sadekur *et al.* An experimental study analysis of security attacks at IEEE 802.11 wireless local area network. **International Journal of Future Generation Communication and Networking**, v. 8, n. 1, p. 9–18, 2015.

WIKIPEDIA. **Lockdown no Brasil em 2020**. [S.l.: s.n.], 2020.

https://pt.wikipedia.org/wiki/Lockdown_no_Brasil_em_2020. [Acesso em Outubro-2020].

WRIGHT, Lloyd; HOOK, Walter *et al.* Bus rapid transit planning guide. **Institute for Transportation and Development Policy, New York**, 2007.

XU, Yanyan; GONZÁLEZ, Marta C. Collective benefits in traffic during mega events via the use of information technologies. **Journal of The Royal Society Interface**, The Royal Society, v. 14, n. 129, p. 20161041, 2017.

ZHANG, Yi; QIN, Xiao; DONG, Shen; RAN, Bin. Daily OD matrix estimation using cellular probe data. *In: 89TH Annual Meeting Transportation Research Board.* [S.l.: s.n.], 2010.

ZIMEK, Arthur; FILZMOSE, Peter. There and back again: Outlier detection between statistical reasoning and data mining algorithms. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 8, n. 6, e1280, 2018.