



ÁREA DEPARTAMENTAL DE ENGENHARIA DE ELECTRÓNICA  
E TELECOMUNICAÇÕES E DE COMPUTADORES

# Inferência de Percursos dos Utentes de um Transporte Coletivo

Nuno José Bigodinho Baeta  
(Bacharel)

Trabalho Final de Mestrado para obtenção do Grau Mestre em  
Engenharia de Redes de Comunicações e Multimédia

JURI:

Presidente: Doutor Arnaldo Joaquim Castro Abrantes, Prof. Coordenador

Arguente: Doutor Paulo Jorge Freitas Oliveira Novais, Prof. Associado c/Agreg.

Orientador: Doutor João Carlos Amaro Ferreira, Prof. Adjunto

Dezembro 2015



# Resumo

Nos dias de hoje, os dispositivos móveis inteligentes com o *Smartphone* fazem parte do dia-a-dia de cerca 46% dos utilizadores de telemóvel, representando por isso quase metade da população portuguesa.

Estes dispositivos são dotados de muitas funcionalidades que permitem a deteção do movimento dos seus utilizadores, quer através dos seus sensores GPS e acelerómetro, quer por via dos módulos de comunicação sem fios Bluetooth e WiFi, e por isso podem ser usados para passivamente colherem dados da mobilidade destes.

Este facto, embora controverso devido a questões de privacidade de dados e de possível identificação do seu utilizador, está a revolucionar a forma como se avalia os hábitos, tendências, contagens e mobilidade dos fluxos de população, quer dentro de grandes superfícies ou de espaços públicos.

A proposta deste trabalho é a elaboração de uma prova de conceito que explora as tecnologias de deteção e contagem baseadas em dispositivos WiFi, aplicadas à problemática da contabilização de passageiros num transporte público específico – os autocarros.

A informação sobre o número de passageiros bem como a mobilidade entre origem e destinos é informação extremamente difícil e onerosa de obter de forma recorrente. Esta informação é bastante valiosa para a correta gestão de um sistema de transportes.

O contacto com a empresa Tecmic ajudou a perceber este problema, que hoje ainda não tem uma solução ideal.

**Palavras Chave:** Transportes Coletivos, Autocarro, Tecmic, Carris, XTraN, Contagens, Percursos, Mobilidade, WiFi, *Smartphones*



## ABSTRACT

Nowadays smart mobile devices like the Smartphone are part of the everyday life of about 46% from the mobile phone user's base, representing almost half of total Portuguese population.

These devices are equipped with many features that enable motion detection of its users, either through the GPS and accelerometer sensors, or by the WiFi or Bluetooth wireless communication modules, making possible passive collection of mobility data of its users.

This fact, although controversial due to issues with privacy of personal data and possible user identification, is changing the way on how to evaluate habits and trends, mobility and population flows, both within large service infrastructures, or in public open spaces.

The purpose of this work is to develop a proof of concept that exploits sensing and counting technologies based on WiFi devices, applied to the problem of passenger counting for a specific public transport: the BUS.

Information about passenger counting as well as mobility between origin and destination are extremely difficult and costly to obtain on a recurring basis. This information is very valuable for the proper management of a transport system.

The Contact with Tecmic Company helped to realize this problem, which today does not have yet an ideal solution.

**Key words:** Public Transport systems, BUS, Tecmic, Carris, XTraN, Counting, Routing Path, Mobility, WiFi, *Smartphones*



# AGRADECIMENTOS

Agradeço ao meu orientador, Professor Doutor João Carlos Amaro Ferreira o apoio, incentivo, a orientação e incansável disponibilidade prestados ao longo do projeto.

Agradeço á minha família a paciência, o apoio incondicional e as muitas horas de ausência, para que fosse possível concretizar este objetivo pessoal.

Agradeço também aos meus colegas de mestrado pelo espírito de interajuda e motivação que me proporcionaram ao longo destes anos académicos.





# Conteúdo

1	Introdução .....	1
1.1	Envolvência do Trabalho.....	2
1.2	Contexto Empresarial da Tecmic .....	4
1.3	Objetivo .....	5
1.4	Organização deste documento .....	6
2	Estado da Arte .....	7
2.1	A Carris e a Eco-Condução.....	7
2.2	Sistemas de Localização e Contagens .....	8
2.2.1	Trabalhos Académicos.....	8
2.2.2	Cisco Location Base Services.....	11
2.3	Sistemas de informação de tráfego e Visualização.....	12
2.3.1	GTFS – Static Transit .....	12
2.3.2	GTFS – Realtime.....	13
2.3.3	Ferramentas GIS <i>open source</i> .....	13
2.3.4	Caso de estudo TRAVIC.....	14
3	Proposta de solução .....	15
3.1	Requisitos gerais.....	15
3.2	Modelo e componentes do sistema.....	16
3.2.1	Access Point WiFi .....	17
3.2.2	Unidade XTraN.....	17
3.2.3	Sistema Central de Dados.....	18
3.2.4	Processamento e correlação dos dados .....	18
3.2.5	Extração de informação .....	21
3.3	Dados do Sistema.....	22
3.3.1	Dados WiFi.....	22
3.3.2	XTraN Passenger.....	23
3.3.3	Sensores.....	26
3.3.4	Dados GPS .....	27

3.4	Potenciais problemas.....	28
3.4.1	Frequência das mensagens de <i>Probing</i> .....	28
3.4.2	Ofuscação do endereço MAC .....	29
3.4.3	Terminal WiFi invisível.....	29
3.4.4	Ligação entre o MAC e o indivíduo .....	29
3.4.5	Discriminação dos terminais WiFi .....	29
3.4.6	Precisão GPS .....	30
4	Implementação.....	31
4.1	Configuração do AP WiFi.....	31
4.1.1	Sistema com 1 AP em modo autónomo ( <i>standalone</i> ).....	31
4.2	Módulo de Processamento de dados.....	33
4.2.1	Pré-processamento.....	33
4.2.2	Correlação de Dados.....	36
4.2.3	Filtragem e Agregação de dados .....	38
4.3	Extração de Informação .....	39
5	Prova de Conceito.....	41
5.1	Ensaio de Campo.....	41
5.2	Inserção de dados na BD.....	43
5.3	Correlação de dados.....	47
5.4	Resultados do Ensaio com Viatura Própria.....	50
5.5	Resultados do Ensaio com Autocarro .....	56
5.6	Análise dos resultados .....	59
6	Conclusões .....	67
6.1	Resultados obtidos:.....	68
6.2	Trabalho Futuro.....	69
7	Anexos .....	71
7.1	Anexo 1 - XTraN SQL Query View .....	71
7.2	Anexo 2 – Modificação AP de <i>lightweight</i> para <i>autónomo</i> .....	71
7.3	Anexo 3 - Configuração do AP1252 em modo autónomo .....	71
7.4	Anexo 4 – Rotas Carreira nº400 Carris .....	71
7.5	Anexo 5 – Código para manipulação e ingestão de dados .....	71
7.6	Anexo 6 – Código para Correlação de Dados.....	71

7.7	Anexo 7 – Código para Filtragem de Dados .....	71
7.8	Anexo 8 – Queries para Extração de Informação .....	72
8	Bibliografia .....	73



# Índice de Tabelas

Tabela 1: Comportamento dos terminais face a “Probe Response” [17].....	9
Tabela 2: Principais ficheiros GTFS.....	13
Tabela 3: Excerto de dados GPS .....	27
Tabela 4: Dados Operacionais Sensores – Eventos de Porta .....	34
Tabela 5: Dados Operacionais WiFi .....	35
Tabela 6: Dados Estáticos Carreiras .....	36
Tabela 7: Passageiros por Segmento .....	38
Tabela 8: Validação de Passageiros .....	39
Tabela 9: Tabela de Origem e Destinos por passageiro .....	39
Tabela 10: Informação sobre Origem – Destinos para um percurso.....	40
Tabela 11: Tabela Paragens (Carreira 400 ascendente) .....	44
Tabela 12: Tabela WiFi após função HASH .....	44
Tabela 13: Eventos de Porta por Paragem (Carreira 400 ascendente).....	45
Tabela 14: Passageiros por segmento .....	48
Tabela 15: Matriz Origem-Destino por passageiro .....	49
Tabela 16: Passageiros por segmento, viagem 1 .....	50
Tabela 17: Comparação entre filtros utilizados (viagem1) .....	53
Tabela 18: Distribuição de passageiros por segmento, viagem 2 .....	54
Tabela 19: Comparação entre filtros utilizados (viagem2) .....	55
Tabela 20: Distribuição de passageiros por segmento, viagem 3 .....	56
Tabela 21: Distribuição de passageiros por segmento, viagem 4 .....	57



# Índice de Figuras

Figura 1: Rácio de Smartphones sobre Telemóveis (utilizadores) em 2014 [5].....	2
Figura 2: Metodologia de trabalho no processamento dos dados XTraN [1].....	8
Figura 3: Comportamento dos terminais face a “Probe Response” [17].....	9
Figura 4: Relação entre RSSI e distância com e sem filtro EMD [17].....	10
Figura 5: Contagem passageiros Bluetooth [32].....	11
Figura 6: Cisco Location-Based Services Solution Architecture [6].....	12
Figura 7: TRAVIC – worldwide transport movement tracker [26].....	14
Figura 8: Diagrama geral do modelo proposto.....	16
Figura 9: Casos de discriminação de terminas.....	19
Figura 10: Função de Discriminação.....	21
Figura 11: Base de dados XTraN Passenger (parcial).....	25
Figura 12: Diagrama do modelo de processamento de dados.....	33
Figura 13: Mapa Carreira 400 Ascendente [ <a href="http://carris.transporteslisboa.pt/">http://carris.transporteslisboa.pt/</a> ].....	43
Figura 14: Alteração da posição das paragens (Carreira 400 ascendente).....	47
Figura 15: Distribuição de passageiros por segmento, viagem 1.....	51
Figura 16: Paragens na zona da Gare Oriente.....	52
Figura 17: Distribuição de passageiros por segmento, viagem 2.....	54
Figura 18: Distribuição de passageiros por segmento, viagem 3.....	57
Figura 19: Distribuição de passageiros por segmento, viagem 4.....	58
Figura 20: Comparação de ensaios – percurso ascendente.....	59
Figura 21: Comparação de ensaios – percurso ascendente.....	60
Figura 22: Comparação com Contagem Real – percurso ascendente.....	61
Figura 23: Comparação com Contagem Real – percurso descendente.....	61
Figura 24: Matriz de confusão para percurso Ascendente no Autocarro.....	63
Figura 25: Matriz de confusão percurso Descendente no Autocarro.....	63
Figura 26: Gráfico do valor RSSI para um terminal fixo a 6 metros do AP.....	65
Figura 27: Gráfico do valor RSSI para um terminal no limite do alcance do AP.....	65





# 1 INTRODUÇÃO

---

Um sistema de transporte coletivo eficaz é um serviço essencial nas sociedades modernas, e desempenha um importante papel na mobilidade das pessoas num meio urbano, ao qual estão associadas emissões de CO<sub>2</sub>. Para maximizar a eficiência é necessário a adaptação da oferta às reais necessidades da população, que têm variações de procura de forma horária, diária e sazonal. É portanto essencial possuir informação precisa e em tempo útil para atuar corretamente e para tal, é necessário ter o conhecimento real dos fluxos de mobilidade da população alvo a ser servida.

Outra área recente que tem vindo a ganhar importância é a redução dos consumos pela introdução de boas práticas de condução: o *Eco-Driving* [1, 2].

Estes conceitos têm vindo a ser implementados na Carris por meio de uma empresa que presta serviços na área dos transportes, a Tecmic ([www.tecmic.pt](http://www.tecmic.pt)). A empresa Tecmic disponibiliza atualmente o sistema XTraN [3], desenvolvido a pensar nos operadores de transportes públicos, possibilitando aos seus operadores a consulta de tempos de chegada das carreiras da frota, em tempo-real. Ao mesmo tempo esta aplicação permite controlar os consumos dos autocarros e faz de interface para uma aplicação que permite identificar consumos mensais dos condutores.

Dado que o incentivo é condição essencial para a aplicação destas boas práticas de condução, a Carris introduziu um prémio mensal função de um ranking de consumos dos diversos condutores. Este ranking usa um conjunto de heurísticas baseadas nos parâmetros recolhidos pelo sistema XTraN.

A diversidade de dados é bastante grande e no entanto falta uma informação relevante para o cálculo do ranking: o número de passageiros no autocarro. Este parâmetro influencia bastante o consumo e por isso foi identificado com prioridade pela Tecmic. Atualmente é utilizada uma solução com recurso a câmaras de vídeo, baseada em processamento de imagem para deteção dos passageiros. Além de dispendiosa, verificou-se ser bastante difícil de implementar por necessitar de grandes ajustes iniciais, por exemplo, para mitigar as interferências causadas pela reflexão da luz solar. Por estas razões mantém-se a procura de soluções alternativas.

A Tecmic mostrou interesse na aferição do número de passageiros com recurso a redes e dispositivos sem fio, principalmente o WiFi. A secção seguinte 1.1 (Envolvência do Trabalho), explica as razões desta escolha. A estimativa do número de passageiros permite melhorar o ranking dos condutores no que respeita ao consumo (litros por 100 km) uma vez que é um parâmetro importante que impacta nos consumos do autocarro.

Por outro lado, a possibilidade de perceber de forma passiva onde o passageiro entrou e saiu (matriz de origens e destinos) é informação valiosa para os operadores de transportes. Outro aspeto importante é que este tipo de aproximações pode ser um dos primeiros passos para a cobrança de bilhetes em função do percurso feito, da hora ou do dia.

De igual forma pretende também comparar resultados e o impacto desta solução com outra a ser testada baseada em Bluetooth.

## 1.1 ENVOLVÊNCIA DO TRABALHO

Dado a grande proliferação de dispositivos móveis, no final de 2014 existiam em Portugal cerca de 16,4 milhões de dispositivos ativos ligados às redes móveis em Portugal dos quais 79% se encontram ativos - dados ANACOM [4].

Em Agosto de 2014, 46.4% dos possuidores de telemóvel têm modelos *Smartphones* - dados Marktest [5] (ver gráfico da Figura 1), representando cerca de 4 Milhões de indivíduos na posse de telemóveis com funcionalidades extra como por exemplo o acelerómetro, recetor GPS, capacidade de ligação a redes WiFi, *Bluetooth*, etc.

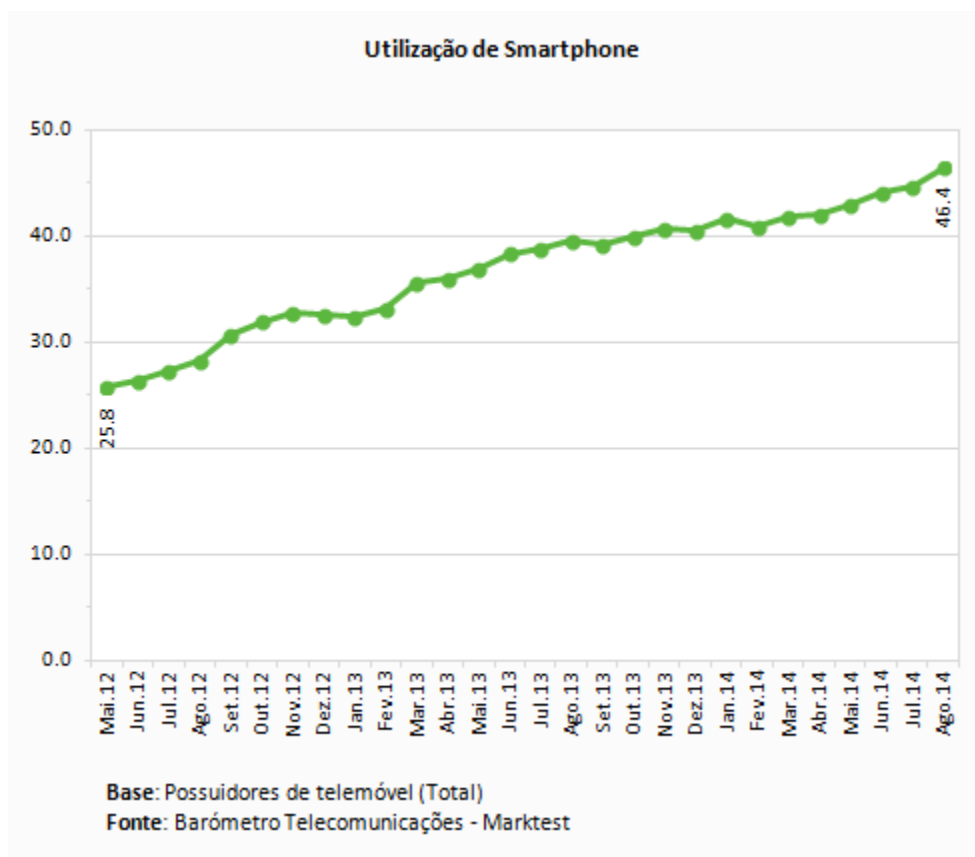


Figura 1: Rácio de Smartphones sobre Telemóveis (utilizadores) em 2014 [5]

Estes dispositivos fazem parte do dia-a-dia dos utilizadores e por isso podem ser usados para passivamente colherem dados da mobilidade destes, usando os sensores GPS e acelerómetro. O mesmo se passa com o módulo WiFi, normalmente ativo, estes dispositivos deixam o seu registo nos equipamentos de infraestrutura rede WiFi APs - *Access Points*, bem como a intensidade do sinal recebido. Com estes dois parâmetros é também possível aferir a localização do utilizador como é descrito pela solução Cisco Location Based Services [6]. Estes dados têm muita relevância para os fornecedores de serviços e de publicidade dirigida, pois permitem extrair conhecimento acerca de determinados percursos, hábitos e preferências dos utilizadores, normalmente dentro de grandes espaços de comércio e serviços, onde existem grandes fluxos de pessoas: Centros Comerciais, Aeroportos, Centros Intermodais de Transportes Coletivos, Campus Universitários, etc.

Um dos grandes fabricantes e impulsionadores da tecnologia WiFi, a Cisco, tem criado novos *standards* nesta tecnologia. Desde 2002, o programa CCX iniciado pela Cisco – *Cisco Compatible Extension* [7] – tem como um dos principais objetivos “*enhance mobility services for reliable business applications*”. Hoje em dia cerca de 90% dos fabricantes de *chipsets* WiFi são certificados pela CCX.

A própria Cisco disponibiliza uma plataforma para recolha e processamento dos parâmetros RSSI “*Receiver Signal Strength Indicator*”, através da sua infraestrutura de rede WiFi, de forma a obter a localização do utilizador com uma precisão que pode chegar aos 5 metros: “Accuracy of less than or equal to 10 meters, with 90 percent precision” [6].

Nos dias de hoje, esta tecnologia já é explorada em algumas grandes superfícies comerciais e de serviços, quer para inferência de localização e percursos, quer para publicidade dirigida.

Por outro lado, várias empresas de transportes coletivos têm introduzido serviços adicionais a bordo, tais como serviços de vídeo ou televisão, música, e acesso à internet gratuito para os utentes, que se baseia em ter instalado um *access point* WiFi no autocarro, que por sua vez está conectado a um *router* com interface de comunicações móveis 3G/4G.

No caso particular da CARRIS, esta implementou o serviço NET BUS a 22 de setembro de 2010, disponibilizando aos Clientes internet gratuita para uma frota inicial de 30 autocarros.

Podemos assumir que o acesso à internet é um fator motivador para passageiros e diferenciador para as empresas de transportes. Estes últimos podem explorar o potencial interesse comercial dos prestadores de serviços via internet, por exemplo para publicidade dirigida ou para oferta de outros serviços aos utentes.

## 1.2 CONTEXTO EMPRESARIAL DA TECMIC

A Tecmic é uma empresa Portuguesa fundada em 1988 que tem como principal objetivo o desenvolvimento de soluções para a gestão inteligente de frotas. O XTraN Passenger [3] é uma solução completa e poderosa que permite acompanhar em tempo real toda a atividade da frota de transporte públicos, comunicar por voz e dados com o tripulante, determinar e comunicar a previsão de chegada dos próximos autocarros às paragens assim como informar o utilizador dos transportes públicos sobre tempos de espera através de diversos canais.

Atualmente a empresa comercializa a terceira geração dos sistemas de Gestão de Frotas desenvolvido pela TECMIC e é uma solução para gestão de frotas, transporte de passageiros, transporte de resíduos, forças de segurança e serviços de emergência, logística e transporte de mercadorias.

Esta solução encontra-se subdividida no seguinte conjunto de módulos:

### ***XTraN Dispatch Manager***

O módulo Dispatch Manager do XTraN completa a solução de gestão de frotas, com uma ferramenta para o planeamento e gestão dos serviços, a qual permite realizar a otimização do planeamento, em função dos recursos disponíveis e das suas competências.

### ***XTraN Enforcement***

O XTraN Enforcement permite às Forças de Ordem Pública e Emergência terem o conhecimento total e completo do terreno, dos meios e dos mais variados dados do cenário.

### ***XTraN Passenger***

É um sistema de Gestão de Frotas de Transporte de Passageiros que visa melhorar o aproveitamento dos recursos, assegurar a regularidade dos horários, informar os passageiros, aumentar a segurança a bordo e garantir uma importante redução dos custos de exploração.

### ***XTraN Web***

O XTraN Web é um interface que disponibiliza, de uma forma restrita ou livre, a informação relevante aos clientes e colaboradores de uma empresa.

A TECMIC disponibiliza também o seu sistema XTraN em ASP ("Application Service Provider"), para facilitar a consulta aos utilizadores autorizados de diversas informações relevantes acerca das atividades desenvolvidas pelas frotas.

### 1.3 OBJETIVO

O objetivo deste projeto é inferir o número de passageiros num autocarro de forma passiva e respetivos percursos dos utilizadores. Esta informação é determinada com base na introdução de um Access Point (AP) WIFI em modo *monitoring* nos autocarros e usando o pressuposto que a grande maioria dos passageiros tem dispositivos móveis com WIFI ativo. É um projeto ligado a uma necessidade real na área dos transportes o qual pretende ser uma prova de conceito para a Tecmic aferir se vale a pena investir nessa aproximação para o problema da contagem do número de passageiros. A informação recolhida é trabalhada por forma a extrair informação relacionada com o número de passageiros e respetivos percursos. Esta informação é importante para corrigir o ranking dos consumos dos autocarros [1, 2] através de uma estimativa do número de passageiros num determinado percurso.

De igual forma este trabalho propõe a determinação de percursos feitos pelos utentes através da análise e correlação de dados de variadas fontes, tal como a localização das paragens, dos recetores GPS e sensores de portas do autocarro, e dados de mensagens de *Probe Request WiFi* dos terminais móveis dos utentes. Esta informação será muito importante para o planeamento de rotas, frequência e horários dos transportes.

O valor desta solução será assim muito inferior a outras soluções existentes no mercado baseado em câmaras de vídeo, as quais têm um processo de instalação e de parametrização inicial bastante complexo, para não falar dos elevados custos associados.

Assim o presente trabalho pretende criar uma prova do conceito, pela introdução do AP WiFi em modo monitor, a manipulação dos ficheiros de *logging* associados para extrair informação relevante. A informação recolhida deveria ser enviada para um servidor central usando a infraestrutura existente do XTraN nos autocarros o qual tem capacidade de armazenar eventos e comunicações remotas 3G. No entanto para a presente prova de conceito, estes dados WiFi serão recolhidos e armazenados por um sistema independente da unidade XTraN, através de um PC portátil dedicado configurado como servidor Syslog, pretendendo simular a unidade embarcada XTraN nesta componente de processamento. Posteriormente estes dados serão correlacionados com os restantes dados dos sensores e com os dados das rotas dos autocarros obtidos através da base de dados central XTraN Passenger.

## **1.4 ORGANIZAÇÃO DESTE DOCUMENTO**

Este documento está organizado em 8 capítulos: Introdução, Estado da Arte, Solução Proposta, Implementação, Prova de conceito, Conclusões, Anexos e Bibliografia.

Nos capítulos Introdução e Estado da Arte descrevem-se a motivação, contexto e objetivos deste trabalho, em que neste último se aborda o estado da arte relativamente às principais áreas de abrangência que este trabalho evoca.

No capítulo 3, Proposta de Solução, elabora-se respetivamente no modo de como se efetuou a abordagem ao problema, caracterizando-o em termos de requisitos e de dados disponíveis, bem como a descrição do modelo da solução.

No capítulo 4, Implementação, descreve-se a implementação em detalhe de cada componente da solução.

Por forma a validar o conceito que este trabalho pretende implementar, o capítulo 5, Prova de Conceito consiste num trabalho de campo com recolha de medidas reais, recorrendo ao uso de um protótipo, em que se efetua a validação do modelo proposto.

No capítulo 6, Conclusões, efetua-se um resumo sobre os aspetos principais dos problemas a que a solução proposta pretende resolver, comparando com outras soluções. Este capítulo inclui também uma perspetiva de evolução da solução.

No final deste documento estão os Anexos com todas as configurações e scripts utilizados e a Bibliografia com as referências aos documentos de suporte a todo este trabalho.

## 2 ESTADO DA ARTE

---

Neste capítulo apresenta-se o enquadramento do trabalho prévio efetuado no contexto do presente trabalho, discute-se a familiarização com a realidade da Carris e o conceito da Eco-Condução, sistemas de localização, contagem e de informação de tráfego, com recurso a sistemas informação geográfica, bem como trabalhos efetuados nessa área.

### 2.1 A CARRIS E A ECO-CONDUÇÃO

A Carris possui uma frota de 745 autocarros, conduzidos por 1410 condutores num total de 152 rotas, com uma extensão superior a 1000 km.

A recente introdução de medidas de *Eco-Driving* na Carris permitiu o aumento da qualidade geral do serviço, da segurança e economia de combustível.

Este conceito foi implementado pela empresa TECMIC, baseado no seu sistema XTraN, que oferece um sistema de gestão integrado de frotas de transportes cujas principais funcionalidades são a localização GPS, controlo consumos, tempos de percursos e ranking dos condutores. Durante um ano o XTraN produz cerca de 2,5 milhões de eventos. Os dados operacionais são enviados para uma base de dados num servidor SQL, existente no sistema XTraNPassenger [3] da TECMIC.

Nas conclusões do trabalho de análise dos dados exportados pelo sistema XTraN “The Impact of Driving Styles on Fuel Consumption” de João C. Ferreira, José Almeida e Alberto Rodrigues da Silva [1], é referido o seguinte:

*“Overall, our findings show that adopting appropriate driving styles can reduce fuel consumption on an average between 3 to 5 liters per 100 km. This can save 15 to 30 liters per bus in just a working day. Taking into account fuel prices, these savings represent 20€ to 40€ a day, per bus. Considering the working days in a year and with around 1,500 involved drivers, this may impact significant savings that can go up to 1.5M€ per year.”*

É ainda de assinalar que a identificação das ações de condução que influenciam o consumo é ainda mais importante nos casos de operadoras de frotas de autocarros, devido ao elevado número de horas que os condutores fazem numa base diária.

O investimento na formação da eco-condução e na promoção de bons estilos de condução pode resultar em economias relevantes para as empresas de transporte e

logística. No entanto, um fator importante que ainda não é mensurado e contribui para o consumo de combustível, é o número de passageiros. Este é um dado importante em falta no sistema XTran e é o principal objetivo deste trabalho.

A metodologia de análise conduzida ao longo desse processo está representado na Figura 2.

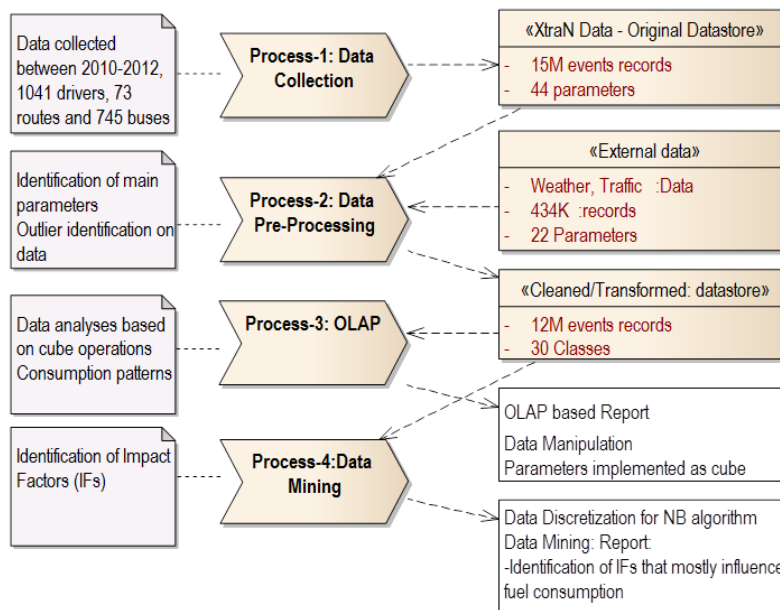


Figura 2: Metodologia de trabalho no processamento dos dados XTran [1]

## 2.2 SISTEMAS DE LOCALIZAÇÃO E CONTAGENS

### 2.2.1 Trabalhos Acadêmicos

Os dados WiFi extraídos das mensagens de gestão 802.11 têm sido ultimamente utilizados para os mais diversos fins, nomeadamente para a determinação de trajetórias de movimentação [21], relações sociais entre indivíduos [19], tempos de espera em filas [22] e estimação de densidades populacionais em determinados espaços [18, 20].

Para o sucesso nos resultados da estimação, recorre-se normalmente a medidas extra para aumentar o número de mensagens recebidas, uma vez que a frequência destas mensagens não é normalizada e cada fabricante opera com valores diferentes. Estas medidas farão parte da uma monitorização ativa em vez de passiva, e requer *software* específico (não *standard*) para o *Access Point*.



No trabalho de Zhuliang Xu et All, em “Pedestrian Monitoring System using Wi-Fi Technology And RSSI Based Localization” [17], um dos métodos utilizados foi a injeção de mensagens *Probe Response* por via de *Access Points* com *software* adaptado. Esta medida não é *standard* e nem todos os terminais respondem de igual forma, de modo a ser recolhido o seu *MAC Address*, sendo este o objetivo primordial (ver Figura 3). Os testes efetuados demonstraram um aumento global de 10% de mensagens dos terminais.

Brands	Operation System	Sniffing	Packet Injection
HTC Corporation	Android	YES	YES
Apple Inc.	IOS	YES	NO
Sony Mobile	Android	YES	YES
Samsung Corporation	Android	YES	YES
Blackberry	Blackberry OS	YES	NO

Tabela 1: Comportamento dos terminais face a “Probe Response” [17]

Para estimação da localização com necessidades de maior precisão ou de contagens de terminais, A. B. M. Musa e Jakob Eriksson no seu trabalho “Tracking Unmodified Smartphones Using WiFi Monitors” [21] propõem a adoção de medidas adicionais: difundindo SSIDs muito populares de forma a aumentar a deteção do número de terminais, provocando pedidos de *Attach* ou ainda difundir SSIDs que estão a ser inquiridos pelos terminais, aumentando o número de mensagens desses terminais.

De modo a obter a localização dos terminais e determinar as suas trajetórias recorre-se às medias RSSI “*Receiver Signal Strength Indicator*”. Estes valores são essenciais para se obter a posição do terminal através da trilateração com vários APs, usando a relação logarítmica com o inverso da distância.

No entanto estas medidas são efetuadas pelos terminais, convertendo a intensidade do sinal recebido para um valor indicativo. Este valor tem variações que dependem das condições e do meio de propagação, pelo que apresentam flutuações ao longo do tempo para uma posição estática. Num meio de elevada concentração de terminais e de obstáculos em movimento, é necessário um algoritmo de filtragem para retirar a componente de ruído do sinal.

A Figura 4 mostra a relação do valor RSSI com a distância, antes e depois de ser aplicado o processamento EDM – *Empirical Mode Decomposition* [17], que retira as componentes de alta frequência.

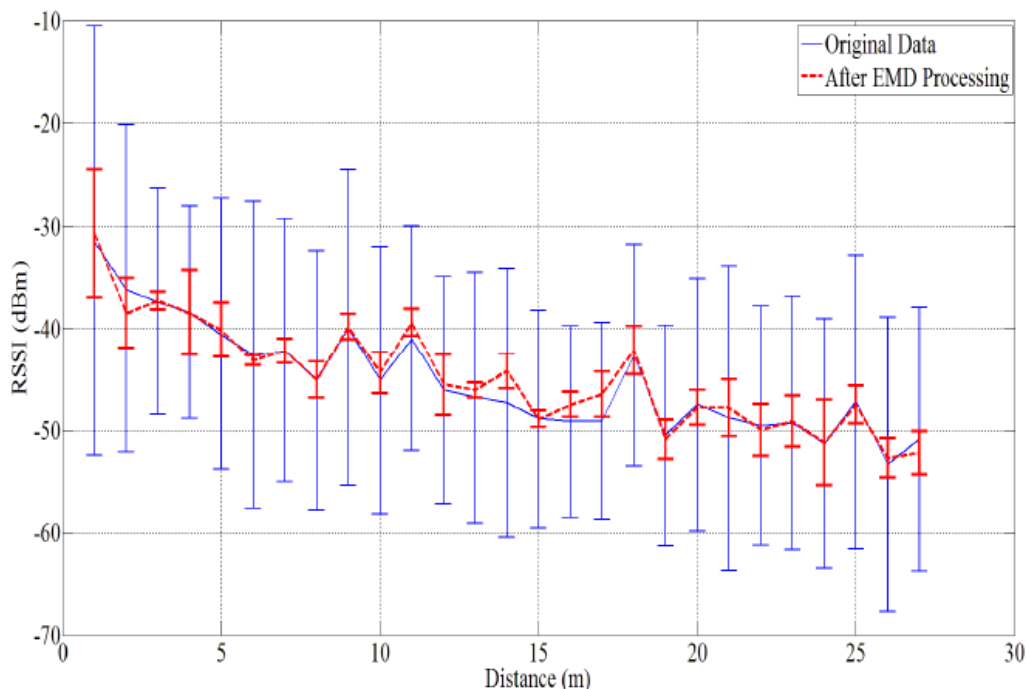


Figura 4: Relação entre RSSI e distância com e sem filtro EMD [17]

No trabalho de Lorenz Schauer, Martin Werner, Philipp Marcus em “Estimating Crowd Densities and Pedestrian Flows Using Wi-Fi and Bluetooth” [18], foi efetuado uma comparação entre as contagens de pessoas que passavam pela zona de segurança de embarque no aeroporto de Munique, usando WiFi ou Bluetooth. As correlações entre a “ground truth” do sistema de bilhética contagem disponibilizada pelas autoridades e as inferências efetuadas através dos sinais WiFi e Bluetooth revelaram que o número de contagens únicas entre estas duas tecnologias foi substancialmente diferente: o rácio das contagens Bluetooth / Wifi foi menor que 4%. Relativamente à correlação para com o sistema de bilhética, os índices foram de 0,53 para BT e de 0,75 para WiFi.

Em complemento existe um trabalho efetuado por Vassilis Kostakos et all [32] em parceria com o operador Horários do Funchal, que explorou a tecnologia Bluetooth para a contagem de passageiros. Este estudo revelou uma boa correlação de 0,73 entre os passageiros detetados para com a contagem de bilhetes cobrados, embora o nº de deteções corresponda a cerca de 10% dos passageiros reais (ver Figura 5). A desvantagem deste sistema é que para um terminal ser descoberto, o terminal tem de estar em nodo “discoverable”. A vantagem é que o protocolo permite que a estação de monitoria envie ativamente mensagens de pesquisa de terminais Bluetooth.

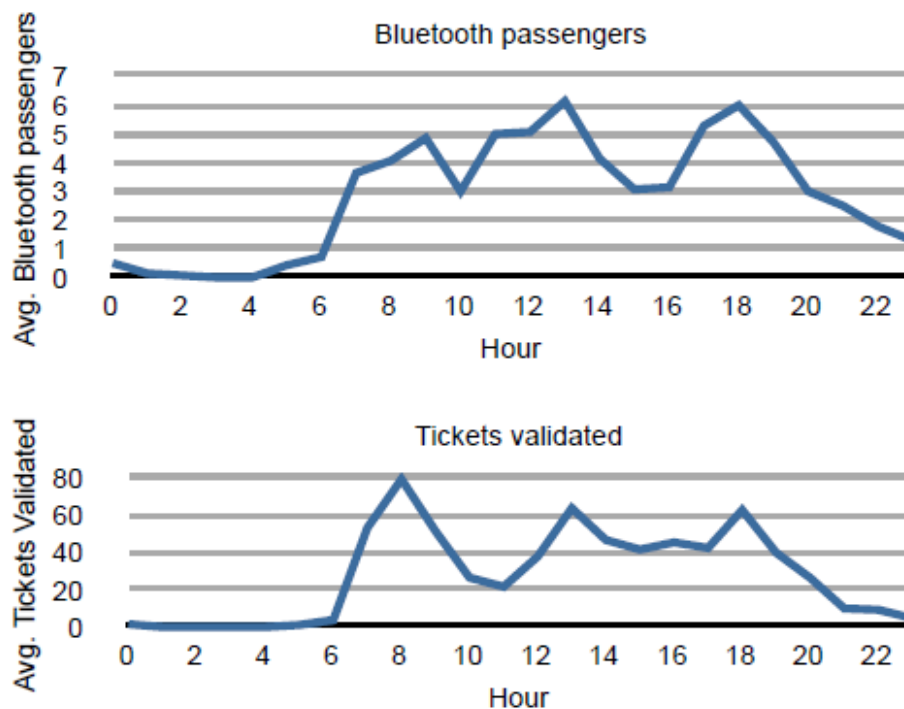


Figura 5: Contagem passageiros Bluetooth [32]

### 2.2.2 Cisco Location Base Services

A Cisco implementou o seu próprio sistema e algoritmos de localização com base nas mensagens CCX [7], *standard* patrocinado pela própria Cisco. Através destas mensagens de gestão, o sistema consegue de forma sistematizada e global, solicitar aos terminais métricas o valor de sinal recebido dos APs, permitindo assim obter as informações sobre o ambiente de rádio a partir da perspetiva do cliente. O sistema LBS da Cisco utiliza estas medições para que, de forma rápida e precisa, calcular a localização do cliente.

Os APs emitem mensagens de solicitação de medição rádio para os clientes que suportem as versões CCXv2 ou posteriores. Estas mensagens são enviadas para cada SSID e para cada interface de rádio, num intervalo de tempo pré-configurado. Neste processo de medição rádio, os clientes enviam os probe request para todos os canais especificados no pedido.

Para aumentar a precisão da localização, os APs emitem pedidos de medição diretamente a determinados clientes que suporte CCXv2 ou superior. Os clientes enviam

então em resposta, vários relatórios de medição para o AP e para o controlador WLC. Estes relatórios incluem informações sobre o ambiente de rádio e dados que permitem interpretar a localização dos clientes, tais como o relatório de path-loss measurement.

Desta forma, a Cisco evita o problema da dependência das mensagens de *scanning* por parte dos terminais, controlando totalmente a periodicidade e tipo de medidas a obter. A Cisco anuncia que o seu sistema (dependendo do número de APs instalados) consegue precisões de localização de cerca de 5 metros [6].

A arquitetura da solução está apresentada na Figura 6.

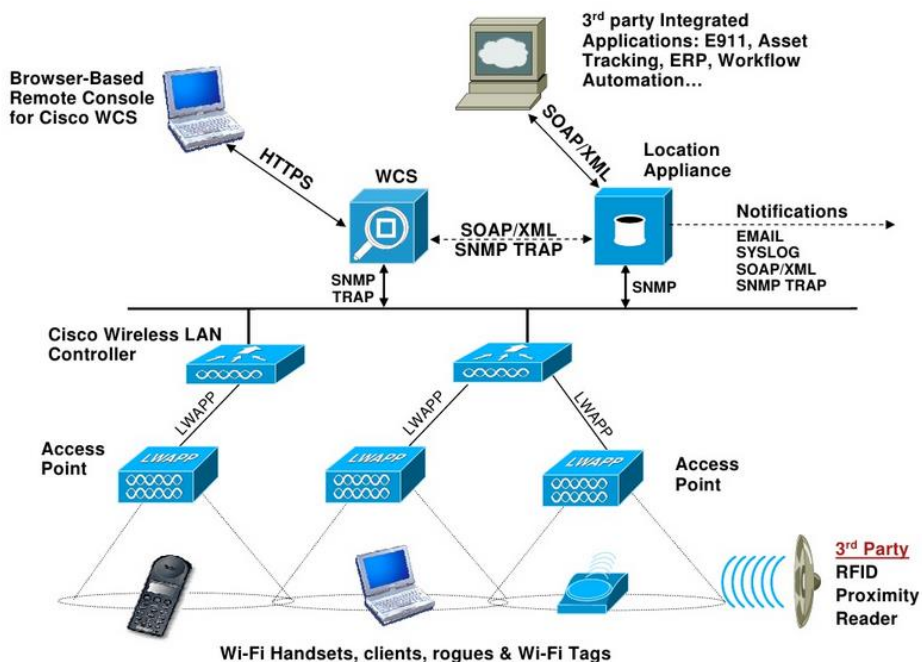


Figura 6: Cisco Location-Based Services Solution Architecture [6]

## 2.3 SISTEMAS DE INFORMAÇÃO DE TRÁFEGO E VISUALIZAÇÃO

### 2.3.1 GTFS – Static Transit

Nos últimos anos, o General Transit Feed Specification [14] tornou-se no formato mais popular para descrever as rotas e horários das carreiras de redes de transporte de passageiros. Grande parte dos operadores de transportes de todo o mundo já disponibiliza as suas fontes de dados neste formato. Esta especificação de dados estáticos permite modelar horários diários ou semanais, informação das paragens, informação geográfica de percurso em multi-linhas “*shapes*”, taxas e alguns dados mais específicos tais como admissão de bicicletas e cadeiras de rodas, etc.

Uma fonte de dados GTFS consiste em 13 ficheiros CSV (alguns são opcionais), sendo os mais importantes listados na Tabela 2:

agency.txt	calendar.txt
stops.txt	calendar_dates.txt
stop_times.txt	shapes.txt
trips.txt	routes.txt

Tabela 2: Principais ficheiros GTFS

Exemplos de dados GTFS bem como o diagrama UML detalhado que descreve as dependências entre as tabelas podem ser obtidos nas seguintes ligações [16].

<https://developers.google.com/transit/gtfs/examples/gtfs-feed>

[http://www.google.com/help/hc/images/transitpartners\\_1106431\\_objecttablelarge\\_en.gif](http://www.google.com/help/hc/images/transitpartners_1106431_objecttablelarge_en.gif)

### 2.3.2 GTFS – Realtime

Desde 2011 a especificação GTFS foi estendida para suportar informação de tempo real, na forma de atualizações de tempos (atrasos) de determinadas carreiras face ao horário oficial.

Esta extensão dá suporte a três novos tipos de dados [15]:

**Trip updates** - onde são publicados os atrasos, cancelamentos ou alterações de percurso relativamente a um identificador de rota GTFS.

**Service alerts** – campo genérico para publicação de notícias

**Vehicle positions** – posição GPS e velocidade do veículo em tempo real.

### 2.3.3 Ferramentas GIS *open source*

Para o suporte aos dados de trânsito com vertente geográfica, quer para armazenamento como para sua visualização, é necessário recorrer a bases de dados com extensão para dados geográficos. Atualmente existem inúmeras ofertas da Microsoft (SQL), Oracle (Spatial Locator) e IBM (DB2 Spatial), que ou requerem licenciamento ou sistemas operativos proprietários.

Em termos de *software open source*, existem opções que permitem explorar a componente geográfica dos dados, bastante utilizadas no mundo académico e empresarial.

O PostgreSQL, com suporte geográfico através do extensor PostGIS, é uma base de dados desenvolvida pelo “The PostgreSQL Global Development Group”, tem uma forte presença no mundo académico e tem licença *open source*.



O MySQL Spatial tal como o *PostGIS* (extensão geográfica para o *PostgreSQL*) têm licença *GPL*.

O *pgRouting* igualmente *open source*, é um extensor ao *PostgreSQL/PostGIS* e fornece a funcionalidade de cálculo de rotas geoespaciais, com os algoritmos mais usuais para o caminho mais curto tais como *A\**, *Dijkstra*, e suas variantes. Existem também ferramentas para importar dados do formato *OpenStreetMap* para *PostGIS*.

Como sistema de visualização de dados geográficos, o *QGIS* oferece uma das melhores ferramentas *open source*, compatíveis com *PostGIS*, a par com o *OpenStreetMap* que fornece dados cartográficos de todo o mundo.

### 2.3.4 Caso de estudo TRAVIC

Um caso de estudo que envolvem muitas das ferramentas de *software open source* referidas (*GTFS*, *PostGIS*, *OpenStreetMap*, entre outras) foi efetuado em 2014 pela empresa *geOps* [25] para mapear redes de transportes públicos em conjunto com a Universidade Freiburg na Alemanha. Deste projeto resultou a aplicação *TRAVIC* [26] - para visualização de transportes de todo o mundo em tempo real (ver Figura 7), cujos dados são alimentados via *GTFS*. O projeto pode ser acedido pela seguinte ligação:

<http://tracker.geops.ch/?z=14&s=1&x=-8235743.4976&y=4971840.9526&l=transport>

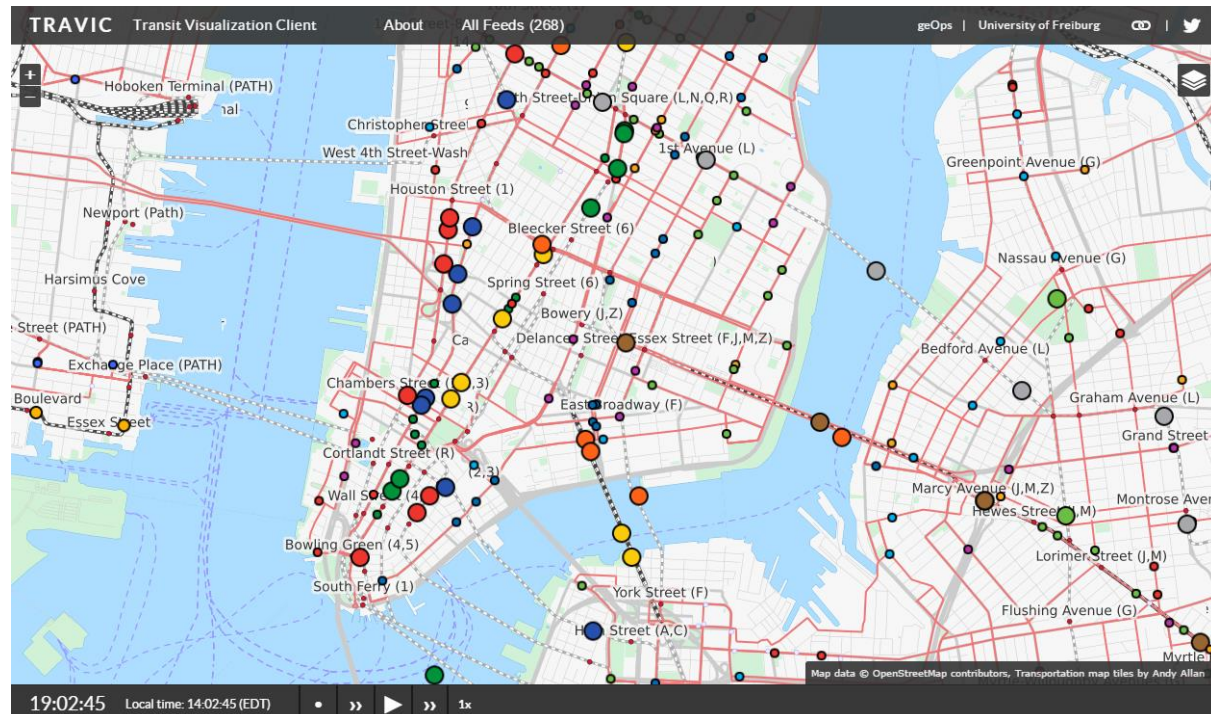


Figura 7: TRAVIC – worldwide transport movement tracker [26]

## 3 PROPOSTA DE SOLUÇÃO

---

O presente trabalho propõe a introdução de um *Access Point* (AP) WiFi em modo *monitoring* nos autocarros. Desta forma, este projeto-piloto não interfere na possível infraestrutura WiFi existente. Numa fase de exploração, esta proposta assenta no pressuposto que os autocarros disponham de infraestrutura WiFi para disponibilização de acesso à internet, de modo a recolher as maiores sinergias possíveis para rentabilização do projeto. Por este motivo decidiu-se em alavancar o projeto em torno de soluções *standard*, não modificando o *software* existente no AP, mesmo que tal fosse possível. Por outro lado, assume-se de que a maioria dos passageiros usa dispositivos móveis com o WiFi ativo.

A solução identificada consiste na introdução de um *Access Point* (AP) WiFi, em modo *monitoring*, para que seja possível obter os dados das mensagens WiFi *Probe Request* enviadas pelos *smartphones* dos utentes presentes no autocarro.

Estes dados serão correlacionados com a informação existente na base de dados XTraN, onde temos acesso à informação das carreiras e suas paragens, informação de porta aberta do autocarro e sua posição GPS.

Esta solução permite, de forma passiva, determinar os percursos efetuados pelos utentes, relações horárias e sua duração, aferir o número de passageiros construir a matriz de origem/destinos.

### 3.1 REQUISITOS GERAIS

Apos diversas reuniões com responsáveis da Tecmic foram definidos um conjunto de requisitos gerais do trabalho:

- Aferição do número de passageiros com base na solução Wi-Fi;
- Inferência dos percursos efetuados pelos passageiros;
- Integração com o sistema XTraN da Tecmic;
- Solução com baixo custo de implementação;

## 3.2 MODELO E COMPONENTES DO SISTEMA

Uma das principais vertentes deste trabalho é o de contabilizar os passageiros (através de seus terminais *smartphones*) ao longo dos percursos das carreiras da Carris usando o sistema WLAN WiFi a instalar no autocarro.

Para esta contagem, contribuem as mensagens enviadas por cada dispositivo móvel, com o objetivo de sondar ativamente as redes WiFi disponíveis, os dados dos sensores XTrAN instalado no autocarro: recetor GPS (posicionamento) e sensor de portas (aberta/fechada), os dados das paragens e percursos de cada carreira.

O diagrama do modelo do sistema proposto é ilustrado na Figura 8, onde identifico os principais componentes da solução proposta para satisfizer os requisitos:

- AP WiFi configurado em modo monitor e instalado de forma autónoma;
- XTRAN on Bus, o qual tem integrado o sistema de transferência de dados para um sistema central;
- Base de dados central, onde estão armazenados todos os dados estáticos e operacionais;
- Processamento e correlação dos dados, onde é feito o primeiro tratamento de dados com a respetiva limpeza e correlação;
- Extração de informação onde é possível obter, por exemplo, a matriz origem/destinos;

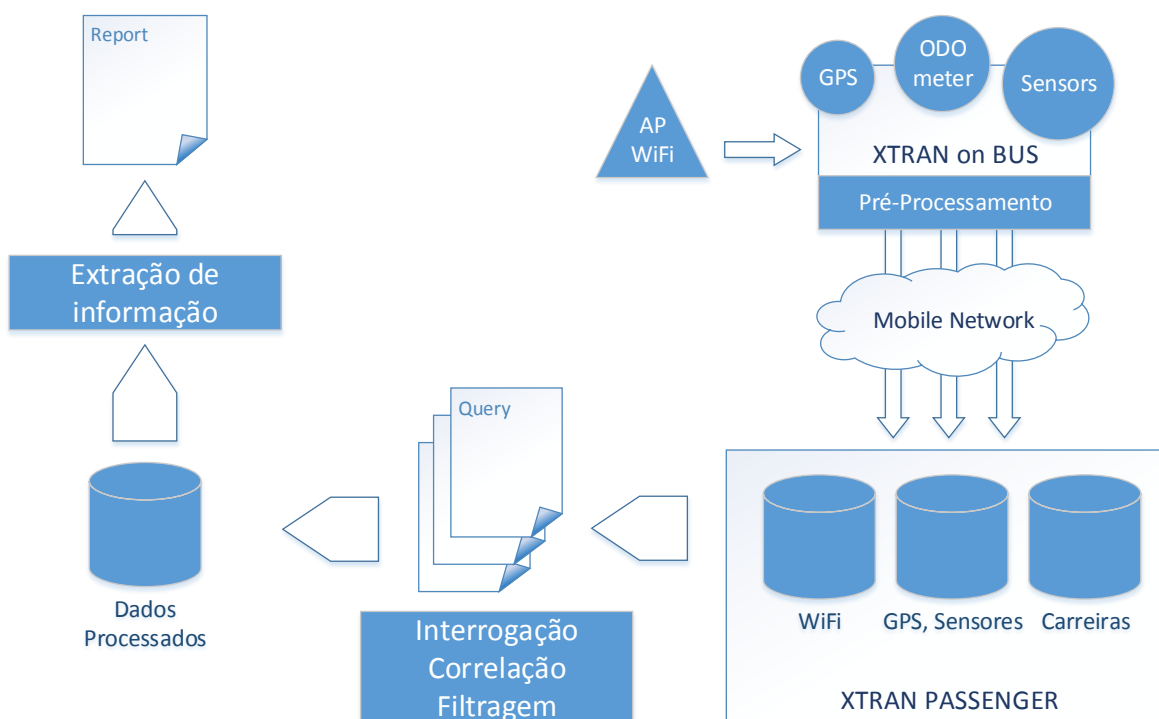


Figura 8: Diagrama geral do modelo proposto



### 3.2.1 Access Point WiFi

Para a elaboração deste trabalho e para a realização da prova de conceito optou-se pela utilização de um AP Cisco da série 1250, equipamento disponível e cedido por empréstimo pelo fabricante. Existem no entanto equipamentos Cisco mais adequados na utilização dentro de veículos, embora partilhem das mesmas versões de *software*. Este AP será configurado para enviar os dados para um servidor *Syslog* externo, que poderá ser o da unidade XTraN. Nesse caso o controlador da unidade processa os dados recebidos, filtra-os e agrega-os de modo a diminuir o volume da transferência para o sistema central, usando os meios de comunicações disponíveis *sem fio*, GPRS/3G, WiFi, etc.

No entanto, dado que este processo é completamente novo e paralelo ao sistema XTraN existente, optou-se por emular este processo através de um PC portátil, externo ao sistema. Os dados do AP são assim enviados para um processo *syslog server* instalado no PC para recolha e armazenamento dos dados. No âmbito da prova de conceito, este PC executa os vários *scripts* de *Software*, que simulam o módulo embarcado XTraN na componente do processamento necessário aos dados. Neste processamento é essencial efetuar o respetivo tratamento aos dados, (extração, filtragem, agregação e formatação) bem como inserir informação de identificação do Veículo, Carreira, Percurso, Variante de percurso e Sentido, para que a base de dados suporte dados de diferentes Carreiras.

Com este pré-processamento dos dados WIFI conseguir-se-á reduzir substancialmente o volume de dados através da agregação dos MAC *address* num único registo por cada evento. Normalmente o terminal envia múltiplas mensagens por cada evento de pesquisa por redes WIFI - uma para cada SSID existente no seu histórico. Esta agregação irá também facilitar o processo de correlação de dados.

### 3.2.2 Unidade XTraN

Esta unidade garante a centralização e agregação de todos os dados dos sensores do veículo, nomeadamente os mais relevantes para este trabalho: o tacómetro, odómetro, sensores de portas e recetor GPS, mas também para todos os restantes sensores: acelerador, travão, mudanças, consumo combustível, etc.

O módulo GPS, opcional no sistema XTraN, está integrado em conjunto com os restantes sensores, disponibilizando a informação por consulta ou por eventos. Para o propósito deste trabalho, é importante a integração deste módulo uma vez que o controlador da unidade embarcada disponibiliza a informação da posição integrada com os eventos dos sensores de portas. Este módulo efetua o tratamento dos dados (extração, filtragem, agregação e formatação) bem a inserção da informação de identificação do Veículo, Carreira, etc. de modo a serem relacionados na base de dados. Os dados GPS são armazenados sem agregação numa tabela separada.

Esta unidade garante também a transferência dos dados dos sensores para o sistema central usando os meios de comunicações móveis disponíveis sem fio, GPRS/3G, WiFi, etc.

### **3.2.3 Sistema Central de Dados**

O centro de dados XTraN é constituído por uma base de dados principal (*Passenger*) com os dados estáticos de percursos, paragens, carreiras, tempos médios de troço e horários. Além dos dados principais, esta base de dados aloja também os dados de serviço operacionais, nomeadamente: início e fim de serviço de carreira/percurso, mudanças ocasionais de percurso e horário, e dados operacionais para cada viatura.

Neste sistema são também alojados as tabelas de dados dos sensores da unidade embarcada: dados dos sensores de porta, GPS, odómetro, consumos e outros dados de suporte a todo o sistema, exemplos podem ser consultados em [1].

Com a inserção neste sistema do AP WiFi, teremos novos dados operacionais a armazenar com o registo de todos os *MAC Address* detetados ao longo do percurso para uma determinada viatura.

### **3.2.4 Processamento e correlação dos dados**

Nesta etapa são efetuadas as interrogações SQL às diferentes bases de dados com o objetivo de correlacionar os dados das diferentes fontes.

Através destas interrogações, vamos obter uma nova tabela com todos os dados necessários desnormalizados de modo a se conseguir aplicar os algoritmos para discriminar e filtrar os dados dos utentes para a contabilização pretendida em cada ligação entre paragens.

O modelo para esta discriminação pode ser dividido em duas partes principais, correspondendo aos factos de o terminal se encontrar dentro ou fora do autocarro.

Para o facto “dentro”, a principal preocupação é obter o maior número de verdadeiros positivos. Considerando a limitação temporal do envio dos “probe requests”, o modelo tem de conseguir obter a maior duração possível para janela de amostragem, sendo este o espaço ou o tempo entre cada paragem.

Para o facto “fora”, a principal preocupação é excluir o máximo número de falsos positivos. Esta função é dificultada pela grande janela de amostragem. Mas para evitar estes casos, considera-se que para a correta contabilização, o terminal tem de ser identificado em duas posições ou tempos de amostragem consecutivas.

### 3 Proposta de solução

O seguinte gráfico na Figura 9 mostra os principais casos de uso (sendo impossível prever todas as situações) para o algoritmo de decisão que irá determinar se terminal se encontra dentro ou fora do veículo, recorrendo a funções para efetuar a discriminação.

Trajeto do Autocarro		Terminais WiFi							
		dentro				fora			
casos de uso		A	B	C	D	E	F	G	H
em movimento		verdadeiros positivos falsos negativos				verdadeiros negativos falsos positivos			
Paragem	parado	V1 --	V1 --	V2 --	V2 --	X1	X1		
	Portas	V1 --	V1 --	V2 --	V2 --	X1	X1	X3	X3
	parado	V1 --	V1 --	V2 --	V2 --	X1	X1		
Paragem	em movimento	V1 --	V1 --	V2	V2			X3	X3
	parado	V1	V1	V2 --	V2 --	X2	X2	X3	X3
	Portas	V1 --	V1 --	V2 --	V2 --	X2	X2	X3	X3
Paragem	parado	V1 --	V1 --	V2 --	V2 --				
	em movimento	V1 --	V1 --	V2	V2				
	parado	V1	V1	V2 --	V2 --				
Paragem	em movimento								
	parado								
	Portas								
Paragem	parado								
	em movimento								
parado									

Sentido do percurso do autocarro ↓  
 Troço X  
 distância mín = 300m  
 Troço Y

Figura 9: Casos de discriminação de terminas

Os estados considerados para o autocarro são os seguintes:

- Em movimento;
- Parado (quer por condicionamento de trânsito ou por se encontrar numa paragem);
- Parado com portas abertas (considera-se que o autocarro apenas abre as portas numa paragem);

Os casos que correspondem aos terminais WiFi de passageiros que estão de facto dentro do autocarro são os casos A, B e C, D. Estes casos tentam representar a dificuldade da deteção devido à periodicidade das mensagens *probing request*.

Os casos E,F e G,H correspondem a casos em que os terminais estão de facto fora do autocarro, sendo que E,F representam pessoas que aguardam em paragens sendo detetadas pelo *access point* e G,H são viaturas que acompanham o autocarro durante um certo tempo do percurso sendo também detetadas.

Os terminais são representados com Vn ou Xn consoante estão de facto dentro ou fora do veículo e são identificados do seguinte modo:

- Verde = positivos
- Vermelho = negativos
- “ - - “ = Terminal não detetado

Cada par de casos está avaliado por duas funções de discriminação distintas, representadas no gráfico em paralelo e identificadas com as elipses em azul e laranja, sendo que irão efetuar a segmentação dos dados de duas formas diferentes:

- As elipses azuis representam a janela de amostragem temporal delimitada pelo evento de abertura de portas;
- As elipses laranja representam a janela de amostragem espacial delimitada pela posição geográfica da paragem;

Comparando as propriedades das duas funções, prevê-se que a função azul não discrimine muito bem os falsos positivos e a função laranja “perde” os verdadeiros positivos dado a sua limitação temporal.

Embora aplicando a restrição de o terminal ser detetado em duas janelas consecutivas, pode revelar-se muito fácil existirem falsos positivos. Os casos que se consideram mais problemáticos são os casos “E” e “F” ilustrados na Figura 9, em que o X1 e X2 representam terminais nas paragens. O terminal X2 é um falso positivo identificado pela função azul.

Este caso pode ocorrer com bastante frequência, uma vez que nos locais de paragem podem existir bastantes passageiros que não entram no autocarro.

Para melhorar a eficácia deste modelo, pode-se recorrer a uma nova função em que que diminua ligeiramente a janela de amostragem momentos antes do local de paragem. Esta função (elipse verde na Figura 10) tem o início de amostragem coincidente com o evento de abertura de portas, mas o instante final seria identificado pelo instante em que o autocarro chega à zona paragem dado pela posição GPS ou em alternativa, um valor temporal antes do início da abertura de portas.

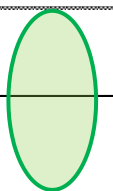
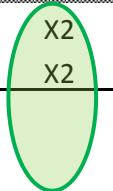
Paragem	parado	
	Portas	
	parado	
	em movimento	
Paragem	parado	X2
	Portas	X2
	parado	X2
	em movimento	

Figura 10: Função de Discriminação

No entanto, esta solução não resolve todas as situações, como se mostra nos casos “G” e “H” da Figura 9. Estes casos representam um terminal dentro de outro veículo que circula junto do autocarro durante um período de tempo indeterminado.

A escolha da melhor solução será objeto de estudo posterior, na Prova de Conceito descrita no capítulo 5.

#### 3.2.5 Extração de informação

Após o tratamento e manipulação dos dados efetuado no módulo de processamento anterior, em que se conhece a natureza específica dos dados, o seu domínio, e suas interdependências, foi disponibilizado um modelo de dados simples sob a forma de tabelas onde se poderão efetuar as interrogações diretas para se extrair a informação necessária.

Nesta última etapa serão efetuadas as diversas vistas sobre os dados de modo a se obter as informações necessárias para a inferência dos percursos efetuados pelos utentes. Esta abordagem específica tem a vantagem de permitir definir algoritmos mais ajustados ao problema específico deste projeto.

Para o problema em causa, este módulo disponibiliza interfaces para a extração de informação, onde é efetuada a manipulação dos dados para aferir informação relevante como o número de passageiros por carreira, percursos efetuados função do período do dia e da zona.

A principal informação que se pretende extrair é a matriz de Origem e Destinos dos passageiros. Com base nesta informação é possível inferir possíveis causas e efeitos correlacionando estes dados com outros acontecimentos externos, tais como efeitos climatéricos, eventos sociais, estações do ano, períodos de aulas e de férias, etc.

Uma grande mais-valia de termos os dados georreferenciados, numa base de dados com suporte geográfico, é o facto de permitir mostrar informação associada aos fluxos de mobilidade dos passageiros diretamente sobre cartografia.

## 3.3 DADOS DO SISTEMA

### 3.3.1 Dados WiFi

Pretende-se obter os dados identificadores do terminal (MAC address) através das mensagens ou *frames* de gestão 802.11 [11, 13]. Estas mensagens não encriptadas são enviadas pelos terminais aquando da pesquisa por redes WiFi disponíveis - *frames Probe Request*.

Em cada mensagem é enviado um dado único do terminal - o seu MAC address - e opcionalmente o SSID da rede a inquirir.

A pesquisa do terminal pode ser ativa ou passiva, sendo esta última sem resultado prático para o objetivo deste trabalho.

- **Pesquisa Ativa:** Este modo é o preferido pelos terminais WiFi, uma vez que permite sondar, de forma seletiva ou geral, quais as redes disponíveis. Neste modo, os terminais enviam um tipo de *frame* de gestão 802.11 para obter os SSIDs dos APs a que se poderão conectar – *Probe Request*. Os APs podem usar os SSIDs do seu histórico armazenado ou ainda colocar este campo a NULL endereçando a mensagem para todos os APs que estejam na sua zona de cobertura (mensagem *broadcast*).
- **Pesquisa Passiva:** Os terminais aguardam pelo envio da *frame* de gestão dos APs WiFi com os seus SSIDs respetivos – a *frame Beacon*. Este é o modo passivo utilizado pelo terminal para obter uma listagem do APs disponíveis. Mas nem todos os APs estão configurados para enviarem em *broadcast* os seus SSIDs.

A obtenção destes dados é conseguida através da colocação do AP em modo debug para este tipo específico de mensagens de gestão, as quais são enviadas para um servidor Syslog existente no módulo de controlo do sistema XTraN. O formato destas mensagens, geradas no AP, é o seguinte:

<b>Timestamp</b>	- String com data e hora no formato HH:MM:SS.MMM
<b>Nº sequência</b>	- Valor em hexadecimal com 3 bytes
<b>Valor 008</b>	- Não existe descrição sobre este campo
<b>Canal e rádio usado</b>	- Canal e rádio usado. Exemplo: -11b
<b>Broadcast address</b>	- MAC address hexadecimal na forma ffff.ffff.ffff

**Client address** - MAC address hexadecimal na forma aaaa.aaaa.aaaa  
**SSID** - Campo a NULL ou com SSID pretendido.  
Exemplo: FON\_ZON\_FREE\_INTERNET

Abaixo apresenta-se um excerto das mensagens *debug probe request* obtidas na consola do AP:

```
*Mar 1 00:27:04.675: 000357, 008, -11b, ffff.ffff.ffff, bcd1.d344.a696 SSID: NULL
*Mar 1 00:27:44.347: 0002AF, 008, -11b, ffff.ffff.ffff, bcd1.d344.a696 SSID: Casa-
BAMJ
*Mar 1 00:27:44.351: 0002AF, 008, -11b, ffff.ffff.ffff, bcd1.d344.a696 SSID: NULL
*Mar 1 00:28:11.367: 0001C1, 008, -11b, ffff.ffff.ffff, 98d6.f79a.9875 SSID:
FON_ZON_FREE_INTERNET
*Mar 1 00:28:11.387: 0001D5, 008, -11b, ffff.ffff.ffff, 98d6.f79a.9875 SSID:
FON_ZON_FREE_INTERNET
*Mar 1 00:28:49.003: 000153, 008, -11b, ffff.ffff.ffff, c0ee.fb25.248f SSID: NULL
*Mar 1 00:28:49.015: 00015E, 008, -11b, ffff.ffff.ffff, c0ee.fb25.248f SSID: orant01
*Mar 1 00:28:49.015: 00015F, 008, -11b, ffff.ffff.ffff, c0ee.fb25.248f SSID: NULL
```

Podemos verificar a presença do MAC *address* do terminal em todas as mensagens e opcionalmente o SSID da rede a inquirir. Este SSID é um dado importante acerca do histórico de utilização das redes WiFi, que pode ser usado como informação extra sobre aquele utilizador.

Por cada evento, o terminal envia varias mensagens por segundo, uma por cada SSID a pesquisar, incluindo uma mensagem com o campo SSID com o valor “NULL”, significando qualquer SSID.

#### 3.3.2 XTran Passenger

Esta base de dados foi disponibilizada através de uma ligação remota via VPN para um servidor SQL na Tecmic. O nome da base de dados é “FSETP\_RTP\_Passenger\_CARRIS”, onde fazem parte dezenas de tabelas com os dados XTran sobre os percursos, dados dos veículos e seus sensores, dados dos condutores, horários, etc.

Uma das estruturas de dados existente representa as várias carreiras de autocarros, suas rotas e paragens. Estes dados, de natureza estática, servirão de base ao sistema com a finalidade de serem comparados com as posições GPS obtidas em cada evento gerado pelos sensores do autocarro (concretamente os sensores de porta aberta/fechada).

O modelo de relação simplificado é o seguinte:

**CARREIRA** (ID, NR\_CARREIRA, NOME\_CARREIRA, URBANA, NR\_PTPARAGEM)  
**CARREIRAVARIANTE** (ID, FK-ID\_CARREIRA, DATA\_INI, DATA\_FIM,  
NR\_VAR\_PERC)  
**CARREIRAVARIANTESENTIDO** (ID, FK-ID\_CV, SENTIDO, FK-ID\_VARPERC\_INI, FK-  
ID\_VARPERC\_FIM)  
**VARPERC** (ID, FK-ID\_CVS, FK-ID\_TROCO, ORDEM\_TROCO, FK-NEXT)  
**TROCO** (ID, DESCRITIVO, MAXSPEED, FK-ID\_CONNECTION\_INI, TROCO\_GEO)  
**Connection** (ID, ORDEM\_PARAGEM, DIST\_PARAGEM\_SEG, DIST\_ACOMUL, FK-  
ID\_PARAGEM, FK-ID\_TROCO, FK-NEXT)  
**ConnectionTime** (ID, EPOCA, DIATIPO, TEMPO\_PERC, TEMPO-PARADO, FK-  
ID\_CONNECTION, FK-ID\_VARIANTE)  
**VARIANTE** (ID, DESCRITIVO, TRIGER\_INI, TRIGER\_FIM)  
**PTPARAGEM** (ID, NUMERO, NOME, INICIAIS, TIPO, MAXTIME\_SEC, LAT, LONG)

A tabela CARREIRA contém os dados relativos ao serviço da carreira: Número, Nome, tipo Urbana, Número de Paragens, bem como outros dados operacionais. Uma carreira pode ter variantes de percurso e de horários, pelo que existe outra tabela, CARREIRAVARIANTE para tratar esse aspeto.

A tabela CARREIRAVARIANTESENTIDO, tal como o nome indica, representa o sentido de percurso de uma carreira. A partir desta última, é efetuada a relação de ordem para cada troço de percurso, através da associação com a tabela VARPERC, que contém as referências para cada troço.

A tabela TROCO representa um conjunto de ligações entre paragens. Essas ligações são representadas pelas tabelas CONNECTION e CONNECTIONTIME: a primeira referencia a paragem destino e a próxima ligação, com informações sobre as distâncias. A segunda tem a informação sobre os tempos de acordo com determinadas variantes.

É importante salientar a única tabela que contém um elemento geométrico é a tabela troço, mas os pontos das paragens não pertencem à estrutura geométrica. Assim, os dados úteis para a inferência do local de paragem são de facto os dados Latitude e Longitude da tabela Paragem. Note-se também que a tabela PTPARAGEM tem um atributo MARGEM\_GPS o que indica que estes pontos foram obtidos por GPS e que têm uma margem de erro. O erro GPS máximo presente na tabela PTPARAGEM é 25 metros.

Para obtenção de uma vista sobre os dados das paragens para cada carreira, variante e sentido foi efetuada uma *query* SQL descrita no Anexo 1.

Na Figura 11 apresenta-se o modelo de dados parcial relativo aos dados estáticos das carreiras.



### 3 Proposta de solução

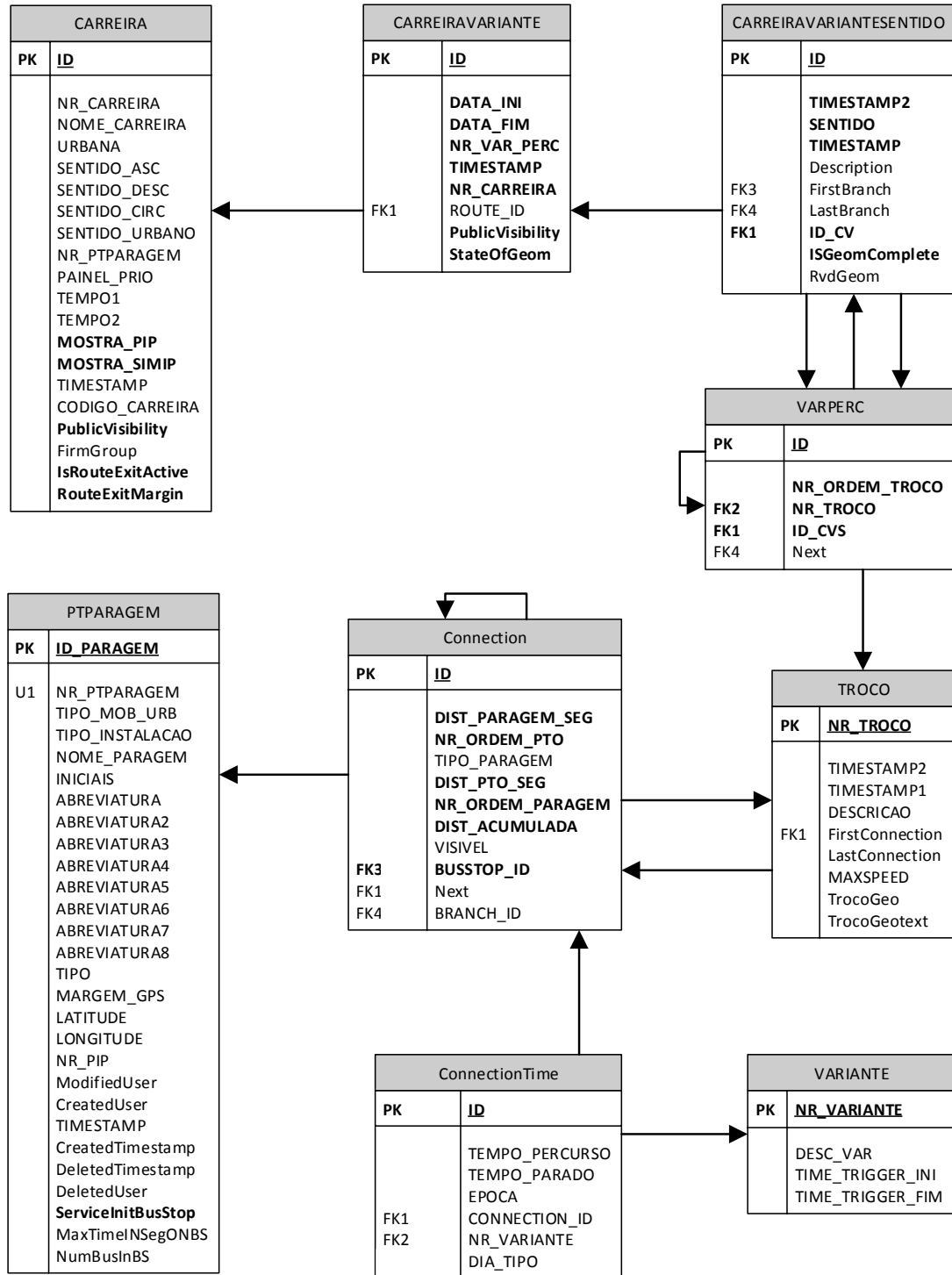


Figura 11: Base de dados XTraN Passenger (parcial)

### 3.3.3 Sensores

Os dados dos sensores são recolhidos pela unidade XTraN embarcada no veículo. Esses dados são processados e armazenados na unidade para envio posterior ou em tempo real para a base de dados sistema central.

Para a solução proposta é necessário o momento em que ocorre a abertura e fecho das portas do autocarro. Este indicador será utilizado para delimitar os momentos de validação dos utentes detetados pelo AP.

A estrutura de dados é a seguinte:

**BUS\_ID** - Identificador do veículo

**CARREIRA** - Número da carreira

**SENTIDO** - Sentido do percurso: 'ASC' ascendente, 'DESC' descendente ou 'CIRC' circular.

**VAR\_PERC** - Variante de percurso: 0, 1, 2, 3...

**Sensor** - Identificação do sensor. Para as portas, poderá ser 'Door1' ou 'Door2'.

**Event\_Type** - Tipo de evento: no caso do sensor de portas – 'OPEN' ou 'CLOSE'.

**Event\_Time** - Data e hora de ocorrência do evento no formato: DD/MM/AA HH:MM:SS.

**Distance\_Trv** - Valor instantâneo da distância percorrida desde início do percurso em metros, valor obtido pelos sensores do tacómetro da unidade no momento do evento.

**Speed** – Opcional: Valor instantâneo da velocidade obtido pelo GPS. Este dado pode não estar disponível \*.

**Latitude** - Opcional: Posição GPS. Este dado pode não estar disponível \*.

**Longitude** - Opcional: Posição GPS. Este dado pode não estar disponível \*.

Os dados dos sensores são essenciais para a inferência correta do número de passageiros, pois definem as janelas temporais que possibilitam a entrada / saída dos mesmos. Será dentro destas janelas temporais que poderão existir ocorrências de registo de presenças dos identificadores os dispositivos móveis WiFi, pelo respetivo *Access Point* instalado no veículo.

Como a localização do veículo via GPS pode ser falível, o recurso aos valores do tacómetro são também indispensáveis para inferência da posição do mesmo, comparando este valor com os dados de distâncias entre paragens da base de dados.

\* Estes dados apenas estão disponíveis caso a unidade XTraN integre o módulo GPS.

### 3.3.4 Dados GPS

Os dados GPS serão a forma como obteremos a localização real do veículo ao longo do tempo e da sua trajetória. A forma de integração de uma unidade recetora GPS é através de um porto série EIA-422, obedecendo a este *standard* de interface e ao protocolo NMEA - National Marine Electronics Association [31], para a receção das mensagens GPS.

A unidade XTraN, após a interpretação e processamento das mensagens NMEA, coloca em memória interna a última posição conhecida, sua precisão, hora e data desse evento. Estes dados são por sua vez inseridos numa base de dados local por cada evento dos sensores, de modo a identificar a posição do autocarro nesse momento. Tipicamente o ritmo de mensagens de posição NMEA uma por cada segundo, mas depende do ritmo de comunicação serie definido na interface com o recetor. Para um ritmo série de 4800bps esta frequência de mensagens é garantida.

O problema associado a estes dados são o facto de possuírem um erro de localização intrínseco. Para aplicações de posicionamento em tempo real este erro pode atingir várias dezenas de metros, variando normalmente entre 4 e 50 metros dependendo do número de amostras recolhidas, e do número de satélites visíveis.

Como exemplo, na Tabela 3 mostra-se um excerto de dados do recetor GPS de um *smartphone android* HTC com a aplicação GPSlogger, o qual foi configurado para obter a melhor amostra em cada 5 segundos.

Neste intervalo de tempo, obteremos posições entre cada 80 metros no caso de um veículo circular a cerca de 60Kmh, o que para algumas aplicações pode ser inviável.

Timestamp	Latitude	Longitude	Precisão [m]
2015-05-31T16:24:20Z	38,811603	-9,189344	5,00
2015-05-31T16:24:25Z	38,811519	-9,189169	43,31
2015-05-31T16:24:31Z	38,811502	-9,189449	4,00
2015-05-31T16:24:36Z	38,811467	-9,189385	46,94
2015-05-31T16:24:42Z	38,811501	-9,189606	4,00
2015-05-31T16:24:47Z	38,81147	-9,189496	24,00
2015-05-31T16:24:52Z	38,811404	-9,189492	33,82
2015-05-31T16:24:58Z	38,811388	-9,189745	5,00
2015-05-31T16:25:03Z	38,81149	-9,189862	38,02
2015-05-31T16:25:09Z	38,811393	-9,189731	3,00
2015-05-31T16:25:14Z	38,811533	-9,189814	51,37
2015-05-31T16:25:19Z	38,811608	-9,18973	45,83
2015-05-31T16:25:24Z	38,811639	-9,189701	48,95
2015-05-31T16:25:31Z	38,811391	-9,189882	5,00

Tabela 3: Excerto de dados GPS

Podemos concluir que a precisão da localização depende fortemente da cobertura de satélites GPS e do número de amostras obtidas. Num veículo em movimento, se quisermos obter a sua posição a cada segundo, não existirá a possibilidade de se obter amostras suficientes para diminuição do erro. Terá de ser encontrado um compromisso entre o tempo de amostragem versus o número de amostras necessárias para se obter um erro reduzido.

## 3.4 POTENCIAIS PROBLEMAS

No decorrer da análise dos dados e do modelo do sistema, foram identificados diversos problemas, como por exemplo: a discriminação dos terminais que não estejam de facto dentro do autocarro, validação do percurso do utente (perceber quem entra e sai do autocarro e quando), problemas de cobertura, privacidade de dados, precisão e disponibilidade de sinal GPS, os quais serão descritos em seguida. Esta proposta de solução irá apresentar nos capítulos seguintes os métodos para resolver os problemas identificados.

### 3.4.1 Frequência das mensagens de *Probing*

Esta solução baseia-se na iniciativa das mensagens de *Probe Request* de cada um dos terminais WiFi, em média com uma frequência de uma mensagem por minuto, mas não existe um *standard* que obrigue a respeitar estes tempos – é dependente do fabricante, do modelo do terminal e versão do Sistema Operativo.

Por outro lado, caso o terminal estiver em modo *sleep* com opção de poupança de energia, o módulo WiFi pode ser desligado, dependendo do seu modelo ou versão do S.O.

As mensagens *Probe Request* (em modo *broadcast*) servem para o terminal descobrir ativamente quais os *Access Point* que existem na área de cobertura e seus respetivos SSID.

Outro tipo de mensagens *Probe Request* podem ser enviadas, para interrogar diretamente determinados APs presentes na lista memorizada pelo terminal. Estas mensagens diferem das anteriores apenas no campo SSID, enviado com o SSID pretendido em vez de NULL, indicando qualquer SSID.

### 3.4.2 Ofuscação do endereço MAC

Para tentar garantir aos seus clientes alguma privacidade, a Apple ([www.apple.com.pt](http://www.apple.com.pt)) tomou a iniciativa de alterar frequentemente o endereço MAC, dos terminais com versão igual ou superior a iOS 8.

Esta manipulação do endereço MAC apenas é efetuada quando o terminal está em modo *sleep* e caso não exista alguma aplicação a usar o acesso *wireless* [18].

Por esta razão, e caso os terminais não estejam com alguma aplicação que consuma dados via internet ativa, haverá a possibilidade de serem identificados como vários terminais diferentes.

### 3.4.3 Terminal WiFi invisível

Alguns modelos de terminais podem nunca enviar mensagens de *probing* caso estejam em modo *sleep*. Este comportamento é fortemente dependente das versões de SW do Sistema Operativo e das configurações de poupança de bateria.

No entanto, devido aos interesses comerciais instalados, esta tendência tenderá a diminuir, não sendo verificando este comportamento na maioria dos terminais mais recentes.

### 3.4.4 Ligação entre o MAC e o indivíduo

Embora não exista uma ligação direta entre o endereço MAC e o utilizador, que o identifique, é possível extrair informação relativamente aos hábitos do indivíduo, potenciais ligações sociais, etc. apenas pelos registos dos SSIDs que são enviados nas mensagens de *probing*, obtidos da lista memorizada no terminal em questão. Cruzando esta informação obtida temporalmente com a localização física de cada SSID, temos facilmente por onde e quando passou o utilizador. De facto várias organizações recolhem e registam o mapeamento geográfico dos SSID de cada AP em todo o mundo, por exemplo: Wigle (<https://wisle.net>) [27], Google Geolocation API [29] ou Google Maps Timeline (<https://www.google.pt/maps/timeline>) [28].

Embora este tipo de monitorização já esteja a ser efetuado por muitas empresas, cada vez mais existe a preocupação acerca da privacidade dos dados e do indivíduo.

### 3.4.5 Discriminação dos terminais WiFi

Admitindo que o ganho e diagrama de radiação das antenas do AP a instalar no autocarro pode ser dimensionado previamente, de acordo com a cobertura pretendida para o espaço interior do autocarro, este irá certamente registar mensagens de outros terminais

que não estão no interior do autocarro, mas que ainda estejam na zona de cobertura do AP.

#### **3.4.5.1 Autocarro parado ou quase parado**

Na zona de paragem ou mesmo fora desta (exemplo: passagem de peões), é possível estarem pessoas nas proximidades com terminais com sinal suficiente para que as mensagens sejam recebidas pelo AP. É necessário um mecanismo que exclua estes terminais móveis face aos que entraram no autocarro.

#### **3.4.5.2 Autocarro em marcha normal ou lenta**

No sentido de circulação do Autocarro, na mesma via ou numa via paralela, é possível ao AP receber mensagens de outros dispositivos móveis de outros veículos que acompanhem o autocarro. É necessário um mecanismo que exclua estes dispositivos face aos que se encontram dentro do autocarro.

### **3.4.6 Precisão GPS**

Os recetores GPS comerciais têm um erro de localização que depende do número de amostras recolhidas. Normalmente um recetor GPS tem uma taxa de amostragem de 1Hz, ou seja, tem uma atualização de dados de posicionamento em cada segundo (existem no entanto recetores mais precisos com taxas de amostragem nas dezenas de Hz). Para um elevado número de amostras, após tratamento estatístico, o erro fica limitado a poucos metros com uma confiança de 95% [30].

Um dos fatores que mais contribui para o erro de posicionamento é a cobertura de satélites GPS, ou seja, o número de satélites visíveis em cada momento. Para a obtenção de coordenadas em 3D, são necessários no mínimo de 4 satélites visíveis. O erro tende a diminuir com o aumento do número de satélites visíveis.

Os erros dependem também da qualidade da implementação do recetor, que afetam a precisão do sinal de saída *Pulse Per Second* (PPS) derivado dos osciladores internos.

## 4 IMPLEMENTAÇÃO

---

No capítulo 3 foram apresentados os requisitos gerais, o modelo da solução, descritos os seus componentes e caracterizados os dados sistema. Neste capítulo são descritos os detalhes técnicos da sua implementação.

### 4.1 CONFIGURAÇÃO DO AP WIFI

Para a contabilização dos terminais WiFi, pretende-se ter acesso ao identificador MAC *address* de cada terminal dentro da zona de cobertura de um determinado AP, através do acesso ao *logging* das mensagens *Probe Request* para esse AP.

Nos capítulos seguintes, descreve-se a infraestrutura a instalar e sua configuração para obtenção dessas mensagens.

#### 4.1.1 Sistema com 1 AP em modo autónomo (*standalone*)

##### 4.1.1.1 Versão de SW para AP autónomo

Para instalar o AP em modo autónomo foi necessário colocar a imagem de SW adequada, uma vez que os APs emprestados já tinham sido anteriormente instalados em modo *lightweight*.

Para a série 1250, a Cisco disponibiliza duas versões diferentes:

- A versão *Lightweight*: com a designação k9w8
- A versão *Autonomous*: com a designação k9w7

A versão de SW que obtivemos para AP autónomo foi a c1250-k9w7-tar.124-21a.JY

O processo para modificar o modo de funcionamento do AP de *lightweight* para *autonomous* não é direto e requer execução do procedimento adequado descrito no [Anexo 2](#).

##### 4.1.1.2 Configuração AP autónomo

Após a instalação da nova imagem de SW, o AP foi configurado com os parâmetros mínimos para o seu funcionamento que essencialmente foram a data, hora e interface BVI.

```
interface BVI1
ip address 10.100.100.10 255.255.255.0
no ip route-cache
```

A configuração deste AP autônomo está descrita em detalhe no [Anexo 3](#).

Para obtermos o *logging* das mensagens *Probe Request*, configuramos o AP para enviar as mensagens de *debug* (nível 7) para um servidor *syslog* externo (10.100.100.2), com os números de sequência e a hora de cada mensagem. Os comandos a efetuar na consola do equipamento são:

```
configure terminal
service sequence-numbers
service timestamps log datetime
logging 10.0.0.1
logging trap 7
logging facility local7
end
```

Para termos acesso aos dados das mensagens de *Probing*, ativamos o *debug* para esse tipo de mensagens pretendidas, não esquecendo de configurar também o serviço cliente NTP para manter o relógio interno síncrono com o servidor:

```
configure terminal
clock set <hh:mm:ss day month year>
ntp server 10.100.100.2
end
debug dot11 rxprobe
```

Excerto de *logs* recebidos no servidor *syslog* (retirados os cabeçalhos inseridos pelo processo *syslog* que são sempre "Fri Jul 24 18:45:35 2015: <191>229: 003100: \*"):

```
Jul 24 18:45:35: ToCtrl: 00008D, 004, -11b, ffff.ffff.ffff, e850.8bc9.cead SSID: NULL
Jul 24 18:45:36: ToCtrl: 0006B6, 004, -11b, ffff.ffff.ffff, 4c74.0316.0c50 SSID: NULL
Jul 24 18:45:36: ToCtrl: 0006B7, 004, -11b, ffff.ffff.ffff, 4c74.0316.0c50 SSID: NULL
Jul 24 18:45:47: ToCtrl: 000585, 004, -11b, ffff.ffff.ffff, b407.f970.9330 SSID: NULL
Jul 24 18:45:47: ToCtrl: 0005AB, 004, -11b, ffff.ffff.ffff, b407.f970.9330 SSID: NULL
Jul 24 18:45:58: ToCtrl: 00077E, 004, -11b, ffff.ffff.ffff, 60d9.c735.2578 SSID: NULL
Jul 24 18:45:58: ToCtrl: 000792, 004, -11b, ffff.ffff.ffff, 60d9.c735.2578 SSID: NULL
Jul 24 18:46:02: ToCtrl: 00020F, 004, -11b, ffff.ffff.ffff, 848e.df1e.9076 SSID: NULL
Jul 24 18:46:02: ToCtrl: 000227, 004, -11b, ffff.ffff.ffff, 848e.df1e.9076 SSID: NULL
Jul 24 18:46:07: ToCtrl: 00017A, 004, -11b, ffff.ffff.ffff, e899.c46e.98e3 SSID: NULL
Jul 24 18:46:07: ToCtrl: 0001CC, 004, -11b, ffff.ffff.ffff, e899.c46e.98e3 SSID: NULL
Jul 24 18:46:07: ToCtrl: 0001F4, 004, -11b, ffff.ffff.ffff, e899.c46e.98e3 SSID: NULL
```



## 4.2 MÓDULO DE PROCESSAMENTO DE DADOS

O esquema seguinte (Figura 12) representa o modelo de processamento a implementar, responsável por preparar um novo modelo de dados, com o qual será possível extrair e inferir as contagens de passageiros nos diversos percursos.

Este modelo é constituído por 3 blocos de preparação e transformação de dados, identificados na Figura 12 pelos retângulos 1, 2 e 3.

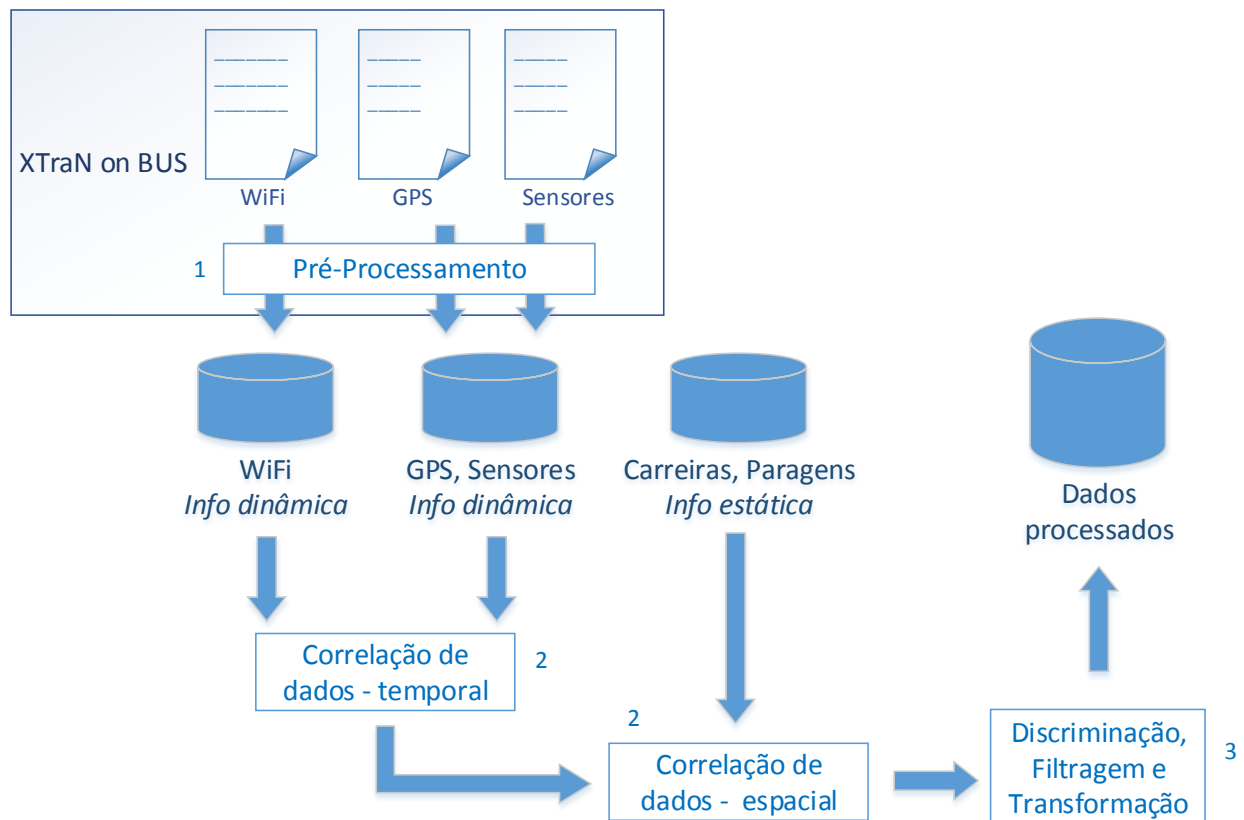


Figura 12: Diagrama do modelo de processamento de dados

### 4.2.1 Pré-processamento

Este primeiro processo, efetuado pela unidade embarcada, tem como função a preparação dos dados obtidos dos ficheiros de logs WiFi, GPS e dos sensores de portas.

Esta preparação consiste em filtrar e agregar os dados operacionais de cada fonte de modo a diminuir o volume de dados e para adaptar para um formato conveniente, para facilitar a correlação dos dados a nível temporal e espacial. O resultado deste processo

é a criação e a inserção numa BD local de tabelas operacionais para processamento posterior. O código implementado para este processo pode ser consultado no [Anexo 5](#).

#### **4.2.1.1 Dados Operacionais Sensores - Eventos**

Nesta fase de pré-processamento temos de garantir que apenas os dados dos sensores de porta aberta (um único evento de abertura por paragem) são transferidos para a tabela operacional. Para tal, são considerados como evento de porta aberta o primeiro evento que ocorre em cada paragem. Este processamento é efetuado através de scripts em Unix Shell com recurso a PGSQL para filtragem e armazenamento local.

No excerto apresentado na Tabela 4, estão os dados na nova tabela operacional, após filtragem para o evento de porta aberta para a viatura 536, seguem o formato descrito em 3.3.3. (para esta apresentação foram retirados os campos VARIANTE e BUS\_SPEED os quais assumiram sempre o valor zero).

BUS_ID	VIAGEM	NR_CARREIRA	SENTIDO	EV_Time	EV_Type	gPonto	Stop_ID
536	8	702	Desc	06-08-2015 12:09:36	OPEN	(38,73068; -9,17321)	Null
536	8	702	Desc	06-08-2015 12:11:31	OPEN	(38,73238; -9,17183)	Null
536	8	702	Desc	06-08-2015 12:12:27	OPEN	(38,73465; -9,17021)	Null
536	8	702	Desc	06-08-2015 12:13:52	OPEN	(38,73316; -9,169)	Null

*Tabela 4: Dados Operacionais Sensores – Eventos de Porta*

Após criação da tabela Eventos, adicionou-se dois novos campos: um ponto geográfico “gponto” através das coordenadas dos atributos Lat e Long, e o Bus\_Stop\_ID preenchido a NULL. Este campo será atualizado no bloco de correlação espacial por inferência da localização geográfica da paragem descrito em 4.4.2.

#### **4.2.1.2 Dados Operacionais WiFi**

Tal como explicado em 3.2.1, estes dados são enviados para o servidor *syslog* por cada evento de *probe request* de cada terminal, o qual cria um ficheiro de “log” por cada viagem.

Após o término duma viagem, estes ficheiros são processados de modo a se extrair a informação relevante (*timestamp* e *MAC address*) e para se inserir os campos identificadores da viatura, viagem, carreira e sentido.

## 4 Implementação

De modo a não ser armazenada informação privada do passageiro e para resolver o problema identificado em 3.4.4, o MAC do terminal é transformado por uma função de HASH, ficando protegida informação sensível do passageiro. O resultado é gravado num novo ficheiro CSV, e a fonte de dados (ficheiro *log*) é apagado.

Numa segunda fase, este ficheiro CSV é processado para ser ingerido numa base de dados local e temporária. Para evitar transferências de enormes volumes de informação, estes dados WiFi são filtrados e agregados para limitar a uma unidade o número de eventos por segundo de cada terminal. Este processamento é efetuado através de scripts em Unix Shell com recurso a PGSQL para filtragem e armazenamento local.

A terceira fase é a transferência da informação para a base de dados central. Após o sucesso da transferência, todos os dados e ficheiros de log e CSV são apagados.

A Tabela 5 apresenta o formato dos dados após filtragem os quais são inseridos numa tabela operacional.

BUS_ID	VIAGEM	NR_CARREIRA	NR_VAR_PERC	SENTIDO	Timestamp	passenger_id
536	8	702	0	desc	06-08-2015 12:09:51	575a82258c89f14c 4b1f69bac0949317
536	8	702	0	desc	06-08-2015 12:10:13	8a18e1e63828719b 06e0a34e40a7f374
536	8	702	0	desc	06-08-2015 12:10:22	575a82258c89f14c 4b1f69bac0949317
536	8	702	0	desc	06-08-2015 12:10:53	575a82258c89f14c 4b1f69bac0949317
536	8	702	0	desc	06-08-2015 12:11:02	8a18e1e63828719b 06e0a34e40a7f374

Tabela 5: Dados Operacionais WiFi

### 4.2.1.3 Dados Operacionais GPS

Tal como efetuado para a tabela Eventos dos sensores de portas, os dados GPS de cada posição são também processados para incluir um novo campo “g\_ponto” geográfico, através das coordenadas dos atributos Lat e Long. Para identificação da viagem efetuada perante a base de dados central, são também adicionados os atributos: BUS\_ID, NR\_VIAGEM, NR\_CARREIRA, VARIANTE, e SENTIDO. Os restantes dados seguem o formato já descrito no Capítulo 3.3.4.

### 4.2.1.4 Dados Estáticos de Carreiras

Por último construiu-se uma nova tabela operacional a partir dos dados estáticos das carreiras, troços e paragens, para se obter uma relação direta da ordem das paragens para cada carreira, sentido e variante de percurso. Estes dados foram obtidos invocando a *query* SQL na base de dados XTraN passenger (a consultar no Anexo 1) e são mantidos sob a forma desnormalizada na tabela operacional “Paragens”.

Dado a ausência de um atributo geométricos para cada paragem, foi necessário criar um novo campo de geometria “gponto” na tabela operacional, a partir dos valores de LATITUDE e LONGITUDE dos dados fonte.

Este passo é importante para ser possível correlacionar estes dados espaciais com os dados do GPS e com os Eventos dos sensores. Este processamento é efetuado com recurso a *scripts* em PGSQL.

Um excerto dos dados são apresentados na Tabela 6 para o percurso da carreira nº 702 no Sentido descendente e na Variante 0 (atributo omitido):

BUS_STOP_ID	LATITUDE	LONGITUDE	NR_CARREIRA	SENTIDO	NR_TROCO	NR_ORDEM_TROCO	NR_ORDEM_PARAGEM
50086	38,730748	-9,173326	702	DESC	2	1	1
2709	38,732515	-9,1718	702	DESC	2	1	2
2709	38,732515	-9,1718	702	DESC	3	2	1
2708	38,734679	-9,170232	702	DESC	3	2	2
2706	38,733084	-9,169057	702	DESC	3	2	3
2704	38,730844	-9,168605	702	DESC	3	2	4
2704	38,730844	-9,168605	702	DESC	1	3	1
8606	38,726199	-9,170424	702	DESC	1	3	2
6112	38,726365	-9,16767	702	DESC	1	3	3
6110	38,728512	-9,16599	702	DESC	1	3	4
6108	38,726812	-9,165069	702	DESC	1	3	5
6101	38,727598	-9,161784	702	DESC	1	3	6
6101	38,727598	-9,161784	702	DESC	4	4	1
6128	38,726692	-9,158576	702	DESC	4	4	2
1224	38,724903	-9,154848	702	DESC	4	4	3
50197	38,725055	-9,151962	702	DESC	4	4	4
50188	38,725436	-9,151416	702	DESC	4	4	5

Tabela 6: Dados Estáticos Carreiras

**NOTA:** Os valores do atributo Variante [0, 1, 2] correspondem a:

- 0 - Madrugada
- 1 - Diurno
- 2 - Noturno

#### 4.2.2 Correlação de Dados

Este segundo bloco de processamento tem como função a correlação espacial e temporal entre os dados das diferentes tabelas operacionais, executado com o auxílio de *scripts* em PGSQL. O código que foi produzido pode ser consultado no [Anexo 6](#).

### 4.2.2.1 Correlação Espacial

Como correlação espacial, temos a inferência da paragem para cada evento de abertura de porta, relacionando os dados GPS de cada evento com a posição geográfica da paragem da tabela de paragens.

Esta relação é obtida facilmente através da função `ST_Distance` do PostGIS, que devolve as distâncias geométricas entre os pontos de cada paragem com o ponto GPS do evento sensor de porta. O primeiro resultado obtido após ordenação ascendente, será a paragem com a distância mais curta ao ponto GPS, sendo adotado esse valor para atualização na tabela Eventos.

### 4.2.2.2 Correlação Temporal

A correlação temporal é necessária para inferir quais os passageiros (terminais WiFi) que estão presentes entre paragens, sendo o evento de abertura o principal delimitador temporal.

Em 3.2.4 foi apresentado o modelo de amostragem temporal dos passageiros em cada segmento. Um segmento é o espaço entre duas paragens consecutivas, e esta definição é necessária para a correta contabilização dos passageiros, ao invés duma contabilização por cada paragem, devido ao maior intervalo temporal que o segmento apresenta, beneficiando as amostragens das mensagens de *probe request*. Este método visa resolver a problemática da frequência das mensagens de *probing* descrita em 3.4.1, que podem ser superiores a 60 segundos, não obedecendo a uma norma *standard*.

Em termos de intervalo de amostragem, este pode ser definido de três formas relativamente ao tempo final, que correspondem a três diferentes funções de discriminação:

- a) Instante de abertura de porta na paragem seguinte;
- b) Instante de abertura na paragem seguinte, ao qual se subtrai um  $\delta^2$ ;
- c) Instante correspondente à posição GPS mais próxima do limite da área<sup>2</sup> da paragem seguinte;

<sup>2</sup> Valor a definir.

A escolha destes valores e funções de discriminação irão influenciar as amostragens dos passageiros, e visam resolver os problemas enumerados em 3.2.4 e 3.4.5.

De acordo com o tipo de percurso de cada carreira ou das condições de trânsito, a escolha da função mais indicada irá resultar em valores mais aproximados aos dados reais.

Para a determinação do tempo em c) foi necessário recorrer a duas funções do PostGIS: `ST_Buffer` e `ST_Disjoint`.

`ST_Buffer` é indicado para estabelecer uma área na zona de influência da paragem, criando um círculo com ponto central na própria paragem.

ST\_Disjoint é a função que devolve todos os pontos que não estão na área da paragem.

O cálculo do tempo foi obtido com o ponto GPS mais próximo em termos temporais do evento de abertura de porta na paragem seguinte, mas não pertencendo a essa área da paragem. Este código foi implementado na função “segment\_passengers”, do Anexo 6.

Após este processamento é produzida uma nova tabela com a contabilização de passageiros segmento em segmento em que os atributos principais são: Segmento, Passenger\_ID, FromStop e ToStop. A nova tabela terá essencialmente o seguinte formato apresentado na Tabela 7 (omitindo os campos de identificação de percurso, viagem, sentido, etc):

Segmento	Passenger_ID	FromStop	ToStop
1	575a82258c89f14c4b1f69bac0949317	50086	2709
1	8a18e1e63828719b06e0a34e40a7f374	50086	2709
2	575a82258c89f14c4b1f69bac0949317	2709	2708
2	8a18e1e63828719b06e0a34e40a7f374	2709	2708
2	1f19f2e96c9e55802315c4aa622c4271	2709	2708
2	0463ba41d3bc53fd24f5f2bf8232a07e	2709	2708
3	575a82258c89f14c4b1f69bac0949317	2708	2706
3	8a18e1e63828719b06e0a34e40a7f374	2708	2706
3	1f19f2e96c9e55802315c4aa622c4271	2708	2706

*Tabela 7: Passageiros por Segmento*

### 4.2.3 Filtragem e Agregação de dados

O terceiro bloco consiste na filtragem dos passageiros segundo as regras que foram determinadas em 3.2.4 para excluir os falsos positivos. O código produzido para este bloco pode ser consultado no [Anexo 7](#).

#### 4.2.3.1 Filtragem de falsos passageiros

As regras a implementar no algoritmo de filtragem são:

- Excluir os passageiros que não tenham efetuado pelo menos um percurso com dois segmentos consecutivos;
- Excluir os passageiros que tenham efetuado percursos em segmentos não consecutivos;

A partir da estrutura de dados anterior, que essencialmente tem a informação exemplificada na Tabela 7, consegue-se facilmente filtrar e contabilizar o número de passageiros em cada segmento.

## 4 Implementação

Assim foi implementada a função “validate\_passengers”, que percorre a tabela identificada na Tabela 7 para contabilizar os passageiros de acordo com as regras de filtragem, marcando na tabela de passageiros no campo booleano “valido” *True* ou *False* caso este seja válido ou não.

O resultado desta validação é apresentado na Tabela 8 (omitindo os campos de identificação de percurso, viagem, sentido, etc para simplificação da apresentação dos dados).

<b>passenger_ID</b>	<b>valido</b>
575a82258c89f14c4b1f69bac0949317	True
8a18e1e63828719b06e0a34e40a7f374	True
1f19f2e96c9e55802315c4aa622c4271	True
0463ba41d3bc53fd24f5f2bf8232a07e	False

*Tabela 8: Validação de Passageiros*

### 4.2.3.2 Origem-Destino por passageiro

Foi ainda implementada uma função “create\_od\_matrix” que a partir das últimas tabelas produzidas “Passageiros por Segmento” e “Passageiros Únicos com Validação” ilustradas nas Tabelas 7 e 8, efetua a exclusão dos passageiros de acordo com as regras de obrigatoriedade definidas, nomeadamente, a permanência em dois segmentos consecutivos.

A Tabela 9 exemplifica o resultado desta função:

<b>Passenger_ID</b>	<b>FromStop</b>	<b>ToStop</b>
575a82258c89f14c4b1f69bac0949317	50086	2706
8a18e1e63828719b06e0a34e40a7f374	50086	2706
1f19f2e96c9e55802315c4aa622c4271	2709	2706

*Tabela 9: Tabela de Origem e Destinos por passageiro*

## 4.3 EXTRAÇÃO DE INFORMAÇÃO

Neste módulo são implementadas as interfaces do sistema para extração de informação.

Após ter disponível o novo modelo de dados implementado pelo módulo de processamento anterior, com informação georreferenciada e numa base temporal é possível extrair informação sobre a utilização deste sistema de transportes.

A título de exemplo, estes dados permitem inferir facilmente:

- O número de passageiros por cada segmento num determinado por percurso;
- Obter a matriz Origem-destino para determinado percurso;
- Identificar os percursos com mais utilizados;
- Identificar as paragens com maior intercâmbio entre carreiras;
- Identificar percursos efetuados com intercâmbio entre carreiras;
- Relacionar o tempo de abertura e fecho das portas com o número de passageiros que entraram e saíram;

Estas *queries* são implementadas sob a forma de evocação de serviços com geração automática de informação, que se suportam nas funções implementadas do módulo de processamento, as quais têm acesso à base de dados, por exemplo:

- **PassageirosPorSegmento** (Carreira, Variante, Sentido, BusID, Viagem)  
Esta função devolve uma tabela com as contagens de passageiros por segmento. Os dados são filtrados pelos parâmetros de entrada, em que apenas a Carreira é obrigatória.
- **MatrizViagens** (Carreira, Variante, Sentido, BusID, Viagem)  
Esta função devolve uma tabela com as contagens de passageiros por percurso origem e destino para uma determinada carreira. Os dados são filtrados pelos parâmetros de entrada, sendo obrigatória a Carreira e o Sentido.

O resultado desta função é apresentado na Tabela 10:

count	segment	fst_stop	lst_stop
5	1	79108	79111
1	1	79108	2410
1	1	79108	7825
1	2	79110	7825
1	2	79110	79112
2	3	79111	7822
4	3	79111	78102
3	4	79112	7824
7	5	78102	7822
1	6	7824	2511
4	7	7822	2511
3	7	7822	7805
14	8	7825	2511
2	9	7805	2401
2	9	7805	2511
1	10	2511	2403
1	10	2511	2405
1	11	2401	2405
1	13	2405	2409
2	14	2406	2410

Tabela 10: Informação sobre Origem – Destinos para um percurso

O código produzido neste módulo pode ser consultado no [Anexo 8](#).



## 5 PROVA DE CONCEITO

---

De modo a validar a presente solução, é necessário efetuar vários ensaios reais com autocarros, efetuando vários percursos em determinadas carreiras da Carris. Mas desde logo foram identificados algumas restrições de carácter técnico e de disponibilidade de recursos que impossibilitaram em tempo útil de obter autorização para instalação dos sistemas para esta prova de conceito dentro do autocarro.

Essas restrições prendiam-se com o facto de ser necessário instalar o AP e o Laptop os quais requerem um ponto de alimentação de 220VAC, além de que seria necessário um local para fixação do AP. Em alternativa seria necessário adquirir uma UPS que suportasse várias horas de autonomia. Por outro lado, para obter os dados reais da unidade embarcada do sistema XTraN (sensores GPS e portas) seria necessário obter expressa autorização da Carris, intermediado pela Tecmic, a qual teria também de disponibilizar um recurso para apoio ao projeto durante o mês de Agosto, o que por corresponder com o mês de férias dificultou este ensaio. A Carris, por seu lado está a passar por um processo de privatização, o que dificulta a obtenção de autorizações e decisões.

### 5.1 ENSAIO DE CAMPO

Assim, dado as presentes limitações, elaborei as seguintes iniciativas:

1) Utilizando a minha viatura pessoal:

Como dispunha de uma viatura pessoal com tomada de 220V, tomei a resolução de efetuar uma simulação do percurso de uma carreira, seguindo imediatamente atrás do autocarro em algumas viagens. O problema da alimentação e autonomia foi resolvido, mas a situação não era a ideal uma vez que o AP se encontrava fora do autocarro.

2) Utilizando um Autocarro:

O cenário ideal deve ter o AP instalado dentro do autocarro. Recorrendo à ajuda do mesmo motorista que foi alvo de perseguição, este decidiu ajudar após uma conversa que tivemos no final do último percurso, utilizando a minha viatura. Assim planeamos efetuar no dia seguinte, dois percursos com o equipamento a bordo. Dada esta excelente oportunidade, procurei por uma UPS suficientemente pequena para ser transportável, que tivesse autonomia suficiente para alimentar o AP durante as duas viagens.

De referir que estes ensaios apenas foram possíveis com a participação de um colega de trabalho, de modo a ajudar na condução e transporte do equipamento.

Os equipamentos a utilizar neste ensaio são:

- AP Cisco 1252AG (AP autónomo em modo monitor);
- PC portátil HP ProBook 5330 (Syslog server + simulador sensor portas);
- *Smartphone* HTC com sistema operativo *Android* (*receptor GPS*);

De salientar que nestes cenários de teste, para além de recolher os dados do AP no PC portátil, terei de simular os dados dos sensores operacionais do autocarro (sensor de porta e sensor GPS) que num cenário real seriam disponibilizados pela Tecmic através dos dados do sistema XTraN embarcado.

Para o registo da posição GPS, recorri à aplicação GPSlogger instalada no *Smartphone* HTC. Esta aplicação foi ensaiada previamente para avaliar a sua precisão. Os seus dados são armazenados em ficheiro CSV no cartão de memória do telefone. Os dados obtidos estão em conformidade com o exposto em 3.3.4 (posição em Latitude/Longitude, *timestamp* UTC, Velocidade, Elevação e Precisão). Esta aplicação foi configurada para gerar uma posição em cada 5 segundos, sendo a amostragem a cada segundo. Desta forma, a posição gerada é a melhor de 5 em termos de precisão.

Para o registo temporal das paragens, construí um pequeno programa em *Python* que corre no Laptop HP e regista num ficheiro CSV o *timestamp* em que ocorre determinado evento a seleccionar pelo utilizador: Abertura ou Fecho de portas, além da marca de Início ou Final de viagem.

Neste PC portátil está também instalado um servidor *Syslog* e um servidor NTP para recolher as mensagens enviadas pelo AP com data e horas corretas, o qual está ligado diretamente por cabo Ethernet e configurado para a mesma rede local. O AP foi configurado como descrito em 4.1.1.

A carreira escolhida para o ensaio foi a 400 “Rossio Levante – Parque das Nações Sul”, por motivos de proximidade do local de trabalho, garantia de afluência de passageiros e transeuntes nas zonas de paragem. Os dados das rotas e paragens sobre esta carreira podem ser consultados no [Anexo 4](#).

O percurso da carreira é mostrado na Figura 13, sendo os percursos dos sentidos ascendente e descendente muito idênticos.



Figura 13: Mapa Carreira 400 Ascendente [<http://carris.transporteslisboa.pt/>]

O primeiro ensaio de campo foi efetuado como o previsto, utilizando a viatura pessoal para perseguir o autocarro em duas viagens. Por cada viagem marquei os tempos de abertura e fecho de portas sempre que o autocarro entrava na paragem, independentemente da porta acionada. Em cada paragem tentei contabilizar o número real de passageiros que saía e entrava, para comparação posterior com os resultados obtidos. Esta tarefa revelou-se especialmente difícil nas paragens com maiores concentrações de passageiros, embora no geral seja uma boa base para aferir os resultados.

No segundo dia efetuei o ensaio dentro do autocarro, com a colaboração do motorista, efetuando uma viagem por percurso. Todos os dados foram gerados e armazenados nos respetivos ficheiros e seguindo as mesmas metodologias usadas no primeiro ensaio.

## 5.2 INSERÇÃO DE DADOS NA BD

De modo a trabalhar e visualizar os dados obtidos, foi utilizado o PostgreSQL com o extensor PostGIS e criada uma base de dados local de nome “xtran”. Para processamento e carregamento dos dados, foram produzidos diversos scripts em PGSQL e/ou Python com módulo psycopg2, interface para PostGIS. Todos os scripts efetuados para a adaptação dos dados fonte e inserção na base de dados de trabalho estão presentes no Anexo 5.

Os dados foram carregados em diversas tabelas locais a partir da BD do sistema XTrAN Passenger descrita em 3.3.2.

Para a tabela `t_paragem` (Tabela 11), foi adicionado um campo geométrico como o ponto da paragem, construído a partir dos campos `LAT` e `LONG`, muito importante para efetuar relacionamentos geográficos com outros dados.

PARAGEM_NR	PARAGEM_NOME	CARREIRA	SENTIDO	ORDEM	LONG	LAT
79108	"Rossio de Levante"	"400"	"ASC"	1	-9.095	38.78323
79110	"Passeio Garcas"	"400"	"ASC"	2	-9.09548	38.78228
79111	"R. Ilha Amores"	"400"	"ASC"	3	-9.09656	38.78064
79112	"Cais Olival"	"400"	"ASC"	4	-9.09546	38.7788
78102	"Cais Olival"	"400"	"ASC"	5	-9.09514	38.77607
7824	"Av. Boa Esperanca"	"400"	"ASC"	6	-9.09741	38.77448
7822	"Av. D. Joao II"	"400"	"ASC"	7	-9.09771	38.77139
7825	"Est. Oriente"	"400"	"ASC"	8	-9.09896	38.76708
7805	"Est. Oriente"	"400"	"ASC"	9	-9.09876	38.76856
2511	"R. Caribe"	"400"	"ASC"	10	-9.09817	38.76594
2401	"Av. Mediterraneo"	"400"	"ASC"	11	-9.09841	38.76245
2403	"Oceanario Lisboa"	"400"	"ASC"	12	-9.09636	38.76116
2405	"Av. Fernando Pessoa"	"400"	"ASC"	13	-9.09731	38.75857
2406	"Hosp. Descobertas"	"400"	"ASC"	14	-9.09818	38.75739
2409	"Parque Nacoes Sul"	"400"	"ASC"	15	-9.09683	38.75531
2410	"Parque Nacoes Sul"	"400"	"ASC"	16	-9.0966	38.75504

Tabela 11: Tabela Paragens (Carreira 400 ascendente)

Os dados *probe request* dos terminais WiFi ficaram armazenados no servidor *Syslog*, cuja sintaxe teve de ser processada e filtrada para se extrair a informação relevante: o *timestamp* e *MAC* do terminal. Para tal recorri ao módulo *PyParsing* para filtrar esses dois campos e utilizei uma função de *HASH* do módulo *Hashlib*, *MD5*, para transformar de forma unívoca o *MAC* numa *string* hexadecimal de modo a não revelar dados privados do utilizador, resolvendo o problema identificado em 3.4.5. Foi assim criada e preenchida a tabela `t_wifi` com esses dois campos (ver Tabela 12) e também com os campos de identificação unívoca do percurso efetuado (`bus_id`, `viagem`, `carreira`, `variante`, `sentido`).

TIMESTAMP	HASH_MAC
07-09-2015 19:01:40	e29eb50b762d4b7bcdf2cd1c4635f641
07-09-2015 19:01:40	e29eb50b762d4b7bcdf2cd1c4635f641
07-09-2015 19:01:41	b7bcb467ef889408315b9c7cd4183a5c
07-09-2015 19:01:41	b7bcb467ef889408315b9c7cd4183a5c
07-09-2015 19:01:41	b7bcb467ef889408315b9c7cd4183a5c
07-09-2015 19:01:41	09e6498964d075acaf197a62c36d41ec
07-09-2015 19:01:41	09e6498964d075acaf197a62c36d41ec
07-09-2015 19:01:45	49b0f5bef4a1943614c8b36cce315e8b
07-09-2015 19:01:45	49b0f5bef4a1943614c8b36cce315e8b

Tabela 12: Tabela WiFi após função HASH

Pela amostra da Tabela 12, verifica-se que existem vários registos por segundo de um único terminal.

Num cenário de exploração desta solução, seria essencial reduzir ao máximo o volume de dados a transferir para a base de dados do sistema central XTraN. Estes dados podem ser substancialmente reduzidos através duma *querie* SQL, agregando as entradas de cada terminal ao segundo, recorrendo para tal à função “date\_trunc” do Postgresql (ficheiro “03\_agregate\_aplogs.sql” do Anexo 5):

```
Select ... From t_wifi
Group by date_trunc('second',w_timestamp), w_hash_mac;
```

Em relação aos dados de simulação dos sensores de porta, bem como os de GPS, ambos provenientes de ficheiros delimitados CSV, todos os campos foram carregados respetivamente para as tabelas t\_events e t\_gps, aos quais foram também adicionados os campos de identificação unívoca do percurso. Foi ainda previsto um campo alfanumérico para associar o código da paragem ao evento dos sensores de portas. Esta associação será efetuada após o carregamento de todos os dados.

Na Tabela 13 são mostrados os dados dos eventos de portas para um percurso ascendente.

EVT_INI	TIME_INI	EVT_END	TIME_END	SENTIDO	PARAGEM_NR
"ABERTO"	"2015-09-07 19:10:40"	"FECHADO"	"2015-09-07 19:10:47"	"ASC"	null
"ABERTO"	"2015-09-07 19:11:20"	"FECHADO"	"2015-09-07 19:11:42"	"ASC"	null
"ABERTO"	"2015-09-07 19:12:36"	"FECHADO"	"2015-09-07 19:12:52"	"ASC"	null
"ABERTO"	"2015-09-07 19:14:04"	"FECHADO"	"2015-09-07 19:14:21"	"ASC"	null
"ABERTO"	"2015-09-07 19:15:17"	"FECHADO"	"2015-09-07 19:15:41"	"ASC"	null
"ABERTO"	"2015-09-07 19:18:07"	"FECHADO"	"2015-09-07 19:18:32"	"ASC"	null
"ABERTO"	"2015-09-07 19:19:21"	"FECHADO"	"2015-09-07 19:19:22"	"ASC"	null
"ABERTO"	"2015-09-07 19:22:47"	"FECHADO"	"2015-09-07 19:22:49"	"ASC"	null
"ABERTO"	"2015-09-07 19:23:51"	"FECHADO"	"2015-09-07 19:24:16"	"ASC"	null
"ABERTO"	"2015-09-07 19:25:49"	"FECHADO"	"2015-09-07 19:26:01"	"ASC"	null
"ABERTO"	"2015-09-07 19:27:23"	"FECHADO"	"2015-09-07 19:27:31"	"ASC"	null
"ABERTO"	"2015-09-07 19:28:14"	"FECHADO"	"2015-09-07 19:28:30"	"ASC"	null
"ABERTO"	"2015-09-07 19:29:05"	"FECHADO"	"2015-09-07 19:29:15"	"ASC"	null
"ABERTO"	"2015-09-07 19:29:31"	"FECHADO"	"2015-09-07 19:29:33"	"ASC"	null
"ABERTO"	"2015-09-07 19:31:10"	"FECHADO"	"2015-09-07 19:31:14"	"ASC"	null
"ABERTO"	"2015-09-07 19:32:05"	"FECHADO"	"2015-09-07 19:32:16"	"ASC"	null

Tabela 13: Eventos de Porta por Paragem (Carreira 400 ascendente)

Depois de comparados os pontos GPS do percurso efetuado com os pontos geométricos das paragens, verifiquei que o autocarro efetuou um percurso diferente do estabelecido (ver Figuras 13 e 14). Considerei assim desatualizados os dados da Carris e corriji as localizações das paragens 79110, 79111 e 79112 do sentido ascendente. No outro sentido não registei qualquer alteração.

Uma vez que a precisão dos dados GPS era elevada (erro predominante na ordem dos 4 metros), foi possível inferir qual paragem corresponde com o evento de abertura de porta. Este processo foi automatizado com um script em PGSQL com o nome “03\_associate\_positions.sql” disponível para consulta no Anexo 5.

O algoritmo utilizado foi o seguinte: por cada intervalo de tempo compreendido entre porta aberta e fechada, obter os pontos GPS registados nesse intervalo de tempo, selecionando aqueles que apresentam precisões com erros inferiores a 10 metros. O ponto escolhido para representar o evento é o que tem a distância inferior ao ponto da paragem, obtido ordenando o resultado da função ST\_Distance por ordem ascendente.

Numa primeira instância, não se conseguiu relacionar todas as paragens aos eventos dos sensores de porta. Isto deveu-se ao facto de existirem troços de percurso em que a velocidade atingida foi maior e, por conseguinte, obtiveram-se menos registos de pontos GPS. Ouve casos limite em que não se registou qualquer ponto numa paragem entre abertura e fecho de portas devido às seguintes condições:

- Tempo entre abertura e fecho inferior a 5 segundos;
- GPS configurado para registar a localização com maior precisão em intervalos de 5 segundos (cerca de 5 amostragens por intervalo);

A resolução passou por aumentar o intervalo de tempo em +/- 3 segundos, tendo-se conseguido associar todas as paragens aos eventos de portas.

O Mapa da Figura 14 mostra o percurso medido pelos pontos GPS, com pontos azuis, no troço em que o autocarro efetuou um percurso diferente (sentido ascendente), já com a nova localização das três paragens. A figura mostra também a localização dos eventos de abertura de porta, identificados a laranja e a localização das paragens com o símbolo vermelho.

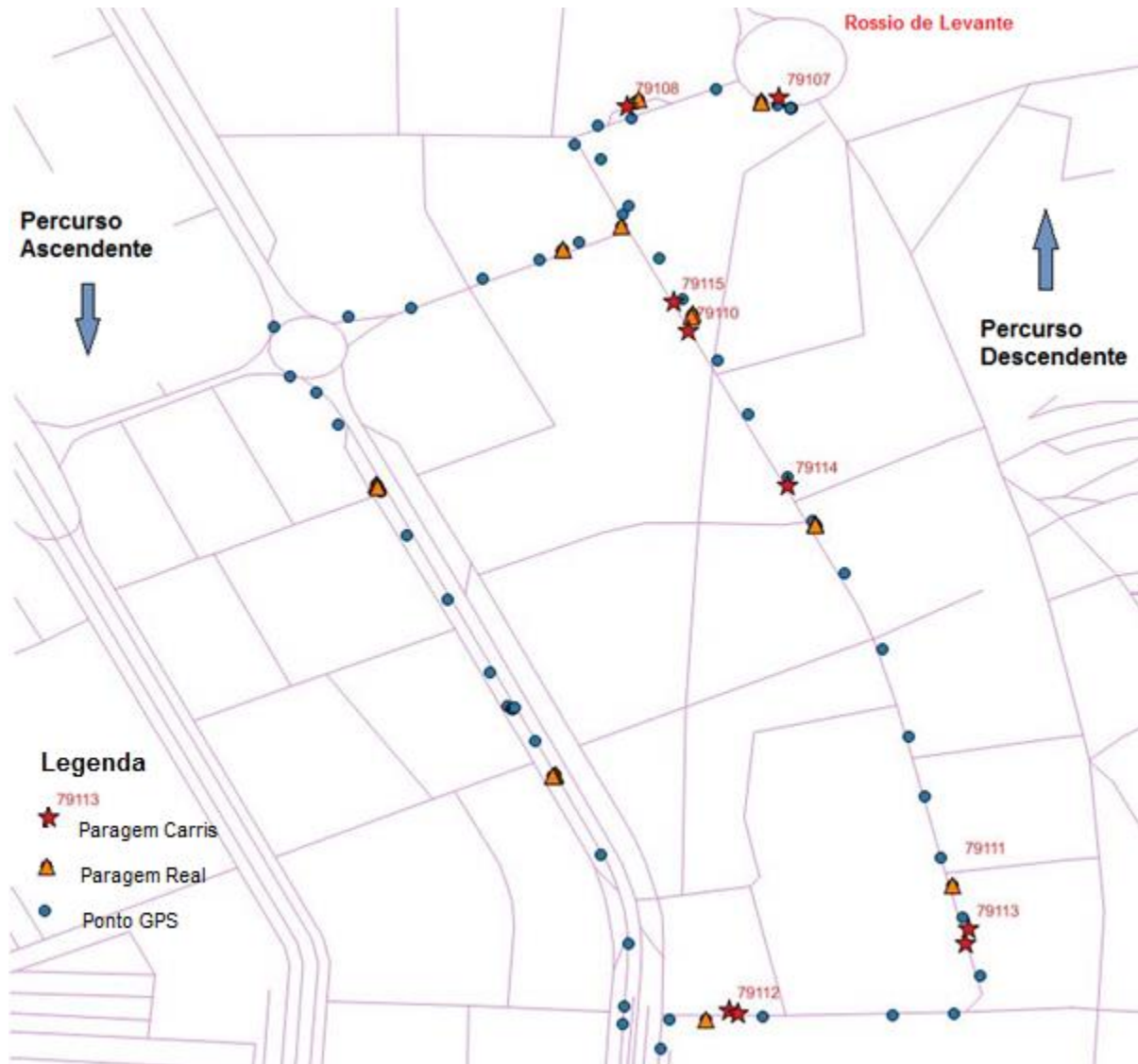


Figura 14: Alteração da posição das paragens (Carreira 400 ascendente)

### 5.3 CORRELAÇÃO DE DADOS

Após a validação dos dados, seu processamento e carregamento da base de dados, o passo seguinte é a segmentação dos terminais pelos intervalos de tempo entre paragens consecutivas. Como foi já referido, as marcas de tempo são o momento de abertura de porta em cada paragem. O código produzido para a correlação e filtragem de dados pode ser consultado no [Anexo 6](#).

Prevendo a necessidade de implementar diferentes funções de discriminação – que dependem da dimensão da janela de amostragem, o código SQL foi preparado para



adaptar o final da janela de amostragem para uma de três opções, relacionado com as funções descritas em 3.2.4, na Figura 9:

- a) Função A: segmentação pelo tempo de abertura de portas (paragem origem e destino);
- b) Função B: variante da função A, em que se subtrai uma constante (ex: 10s) ao tempo de abertura de porta da paragem destino;
- c) Funções C e D: variantes em que o tempo final é o momento em que o autocarro entra na zona da paragem, estabelecida com raio = 50m ou 100m, respetivamente;

Nesta janela temporal parametrizável, são contabilizados os terminais (passageiros) diferentes para cada segmento, sendo este segmento o intervalo de espaço/tempo entre duas paragens consecutivas. Esta definição será usada ao longo do restante trabalho.

Para a determinação do tempo de fim de amostragem para as funções C e D, recorreu-se às funções geométricas `ST_Disjoint` e `ST_Buffer`. A última define a área da influência da paragem destino, com raio de X metros. A primeira indica quais os pontos que estão fora da área da paragem, alimentada pelos pontos GPS que pertencem ao intervalo de tempo do segmento. O ponto GPS escolhido (concretamente o *timestamp* desse ponto) é o que está mais próximo em termos temporais ao tempo de abertura de porta da paragem destino, mas que seja obrigatoriamente inferior a este. Este código está implementado no ficheiro “04\_segment\_passengers\_2.sql” no Anexo 6.

Estas funções serão úteis para testar o sistema e o algoritmo face às situações previstas em 3.4.5 “Discriminação dos terminais”.

O resultado traduz-se numa nova tabela `t_passengers`, organizada da seguinte forma como se mostra na Tabela 14:

<b>PASSENGER_ID</b>	<b>SEGMENT</b>	<b>STOP1</b>	<b>STOP2</b>
db4ab1b0823760112c21e174d0ae15ae	1	79108	79110
db4ab1b0823760112c21e174d0ae15ae	1	79108	79110
db4ab1b0823760112c21e174d0ae15ae	1	79108	79110
db4ab1b0823760112c21e174d0ae15ae	1	79108	79110
21cc3311b5a051bed362435c5d0cd887	1	79108	79110
21cc3311b5a051bed362435c5d0cd887	1	79108	79110
36ad0f1612df0c0ba7fdd5720a14f8e3	1	79108	79110
36ad0f1612df0c0ba7fdd5720a14f8e3	1	79108	79110
90105daa6c4ae9fe408edf410131a677	1	79108	79110
90105daa6c4ae9fe408edf410131a677	1	79108	79110
90105daa6c4ae9fe408edf410131a677	1	79108	79110

*Tabela 14: Passageiros por segmento*

Na Tabela 14 podemos verificar a repetição do identificador do passageiro em cada percurso. Para termos os passageiros por segmento é necessário filtrar os ID distintos.



Após esta segmentação, o próximo passo é inferir quais os passageiros que estão dentro do autocarro.

As regras de exclusão dos passageiros *outliers* são:

- presentes em apenas 1 segmento;
- presentes em vários segmentos não consecutivos;

A primeira regra aplica-se para excluir os terminais que foram contabilizados dentro do raio de cobertura do AP, por exemplo enquanto aguardam na paragem.

A segunda regra serve para excluir eventuais contabilizações em que os terminais acompanham o autocarro em movimento, por exemplo um automóvel ou um outro autocarro.

Para o efeito, a própria tabela `t_passengers` dispõe de um campo binário “valido”, com o intuito de se registar a decisão *True/False* após inferência se este está dentro ou fora do autocarro. Esta validação está implementada numa função em SQL, no ficheiro “05\_validate\_passengers\_2.sql” presente no Anexo 7.

Para a implementação da matriz origem-destinos, foi criada uma outra função SQL “06\_create\_od\_matrix\_3.sql”, presente também no Anexo 7. O resultado é apresentado na Tabela 15, organizada por paragem origem e destino de viagem por cada passageiro. Esta tabela tem o nome `od_matrix` e pretende ser a base de informação para o estado seguinte: Extração de informação.

<b>passenger_id</b>	<b>fst_stop</b>	<b>lst_stop</b>
b970587bb7b15b1dfe8601a25868a9c2	79108	79111
db4ab1b0823760112c21e174d0ae15ae	79108	79111
ded8c494bbaa31f34d19914c215ff516	79108	79111
3f02787d00969b4239cb5da9449af147	79108	79111
6854257ccaef3fc5b02af5f9bc978c1b	79108	79111
b45d020b9e28de7025ea96a73ed4461b	79110	79112
9d4eaaa9030fb49a1cd7c8b631c5aa14	79111	7822
68bfc561a3337591254e52e16f6081b7	79111	7822
21cc3311b5a051bed362435c5d0cd887	79108	7825
e7d9191b1683fc52b81c41405321c2b1	79110	7825
49b0f5bef4a1943614c8b36cce315e8b	79108	2410

Tabela 15: Matriz Origem-Destino por passageiro

## 5.4 RESULTADOS DO ENSAIO COM VIATURA PRÓPRIA

Com base nos mecanismos criados descritos em 4.3 para extração de informação, produziu-se os seguintes resultados sobre os ensaios efetuados, os quais foram adaptados para a forma de gráfico ou tabela para serem apresentados. Para a obtenção destes dados, foram usadas as *queries* SQL descritas no Anexo 8.

### Viagem 1 - sentido ascendente, ensaio em 7-Set-2015 iniciado às 19:10h.

Para avaliar a melhor função de discriminação para os terminais *outliers*, apresenta-se na Tabela 16 os dados de contagens de passageiros para a primeira viagem (no sentido ascendente), aplicando as funções discriminantes A, B, C e D. Estas funções efetuam a segmentação dos dados em janelas temporais entre paragens (segmentos), sendo que as três últimas implementam variantes da primeira função em relação ao tempo final da janela: respetivamente a 10 segundos antes, a 50 metros ou a 100 metros da paragem destino, como descrito em 5.3.

Viagem-1 ASC Segmentos	Paragem Origem	Funções de Discriminação			
		A	B (-10s)	C (-50m)	D (-100m)
1	Rossio de Levante	7	6	7	6
2	Passeio Garcas	8	8	9	8
3	R. Ilha Amores	10	10	10	6
4	Cais Olival	19	15	12	8
5	Cais Olival	27	18	15	7
6	Av. Boa Esperanca	22	13	13	6
7	Av. D. Joao II	15	10	11	9
8	Est. Oriente	23	22	23	22
9	Est. Oriente	25	24	24	21
10	R. Caribe	8	6	5	4
11	Av. Mediterraneo	4	4	4	4
12	Oceanario	6	3	3	3
13	Av. Fernando Pessoa	5	2	2	2
14	Hosp. Descobertas	5	4	4	2
15	Parque Nacoes Sul	4	3	3	1
		<b>188</b>	<b>148</b>	<b>145</b>	<b>109</b>

Tabela 16: Passageiros por segmento, viagem 1

Neste ensaio, devido à utilização de viatura própria, não foi possível efetuar uma contabilização real do número de passageiros dentro do autocarro. Apenas foi possível

perceber o volume de entradas e saídas de forma grosseira. No entanto, analisando a Tabela 16 e o respetivo gráfico da Figura 15 podemos verificar que os resultados obtidos pela função mais discriminante converge para o volume de passageiros detetado visualmente em cada paragem e ao longo do percurso efetuado.

A primeira parte do gráfico, nos segmentos do 4 ao 7, notou-se maior circulação de pessoas em redor das paragens, pelo que as diferenças acentuam-se pela discriminação das funções.

Entre o segmento 7 e 10 que corresponde à avenida com maior concentração de pessoas (D. João II) na zona da Gare do Oriente, não existem grandes diferenças pelo facto que nenhuma das 4 funções conseguiu excluir um caso particular não previsto que se descreve em seguida. Após o segmento 10, o volume de passageiros diminui substancialmente, tal como o número de pessoas em redor das paragens. O tempo durante as paragens diminuiu e a velocidade do autocarro aumentou. As funções tiveram comportamentos similares pelo facto de existir muito menos “ruído”.

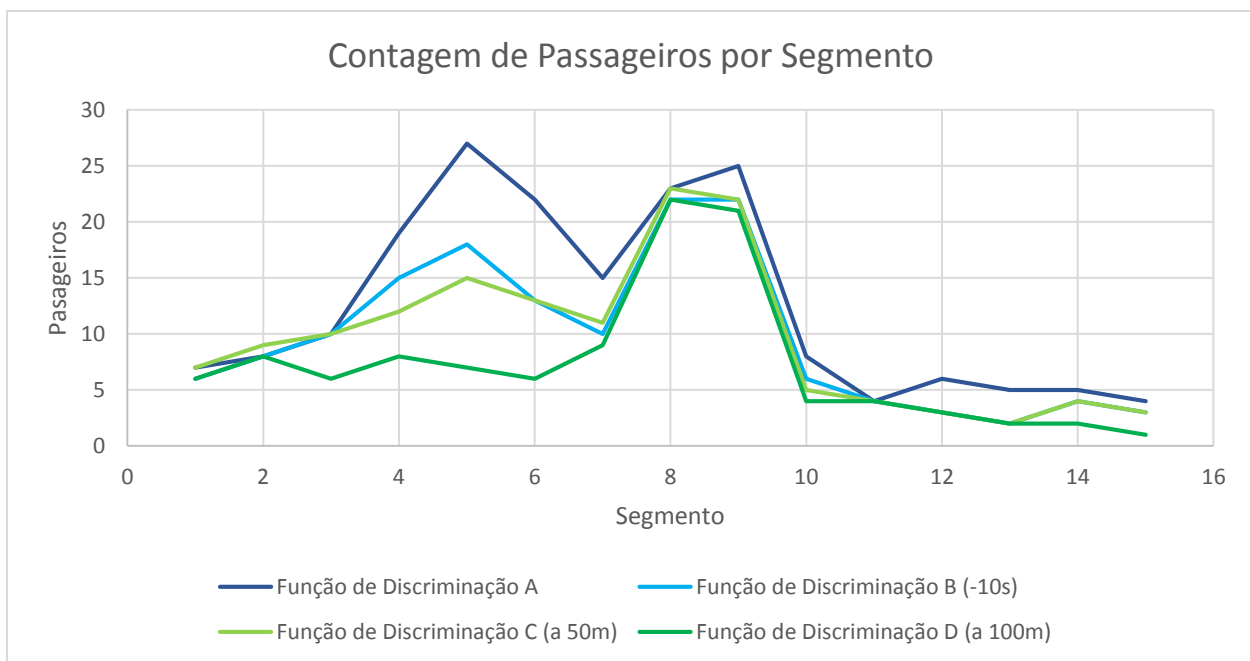


Figura 15: Distribuição de passageiros por segmento, viagem 1

Para explicar o fenómeno que ocorreu entre o segmento 7 e 10, entre as paragens Av. D. João II (7822) e Rua Caribe (2511) onde o elemento central é a Gare do Oriente), é necessário recorrer ao mapa seguinte ilustrado na Figura 16:

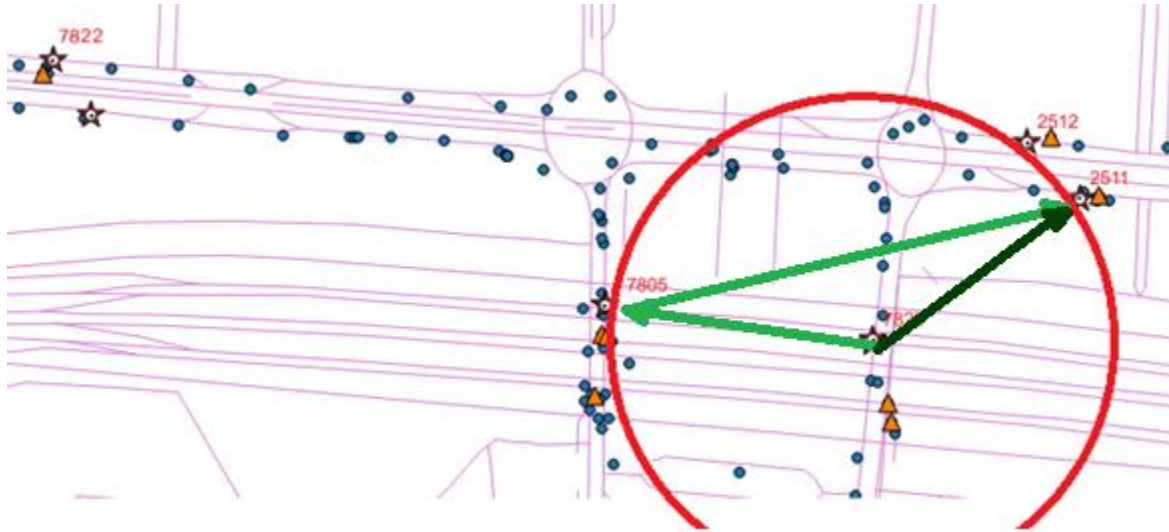


Figura 16: Paragens na zona da Gare Oriente

O círculo a vermelho representa o alcance do sinal WiFi do AP, um raio de 150m em torno da primeira paragem da Gare – 7825 (sentido ascendente) vindo da paragem 7822. A viagem segue o seu percurso para a segunda paragem na Gare – 7805, e depois prossegue novamente pela Av. D. João II para a paragem 2511.

Dado que as paragens 7805 e 2511 estão muito próximas da primeira 7825: cerca de 150m em linha reta, e sendo esta zona bastante povoada, o AP pode conseguir detetar facilmente pessoas nas zonas destas três paragens dentro do seu raio de cobertura. Neste caso particular, a regra de validação de passageiros presentes em dois segmentos consecutivos não resulta.

Para contemplar este caso particular, tem de ser incluída uma nova regra com base na distância linear entre os extremos resultantes da soma entre dois segmentos consecutivos. Caso essa distância seja menor que a soma das distâncias individuais, deve estender-se para 3 o número de segmentos consecutivos de presença obrigatória para que a deteção seja válida.

A solução passa por se parametrizar o número mínimo de segmentos de presença obrigatória, função da distância em linha reta total para dois segmentos consecutivos. Considerando cada segmento como um vetor (vetores verde claro), a distância total é dada pelo vetor soma (verde escuro) dos dois segmentos. Pode-se considerar inicialmente o valor empírico de 300m (média da distância entre paragens) como indicador mínimo para a dimensão do vetor soma. Em princípio o valor de 300m será suficiente para atenuar totalmente o sinal WiFi, para que o terminal não seja detetado. Este alcance depende essencialmente dos ganhos das antenas, potência de emissão e dos fatores de propagação, mas tipicamente o valor anunciado de alcance entre um terminal WiFi e o AP é de 150m em espaços abertos.

A Tabela 17 mostra a aplicação da filtragem de 3 segmentos para todo o trajeto, em que a redução nas zonas de maior densidade de pessoas (entre os segmentos 6 e 10) é bastante notória. Como este filtro foi aplicado a todo o trajeto, eventualmente reduziu também alguns passageiros que efetuaram pequenos trajetos no autocarro, embora esses trajetos curtos sejam atípicos.

Viagem-1 ASC Segmentos	Paragem Origem	Aplicação do filtro por nº de segmentos			
		A (2 segm.)	A (3 segm.)	D (2 segm.)	D (3 segm.)
1	Rossio de Levante	7	2	6	2
2	Passeio Garcas	8	3	8	3
3	R. Ilha Amores	10	5	6	3
4	Cais Olival	19	5	8	4
5	Cais Olival	27	5	7	4
6	Av. Boa Esperanca	22	6	6	5
7	Av. D. Joao II	15	9	9	7
8	Est. Oriente	23	7	22	5
9	Est. Oriente	25	7	21	5
10	R. Caribe	8	2	4	2
11	Av. Mediterraneo	4	2	4	2
12	Oceanario	6	2	3	2
13	Av. Fernando Pessoa	5	1	2	1
14	Hosp. Descobertas	5	1	2	1
15	Parque Nacoes Sul	4	1	1	1
		<b>188</b>	<b>58</b>	<b>109</b>	<b>47</b>

Tabela 17: Comparação entre filtros utilizados (viagem1)

Conclui-se assim que esta filtragem adicional (presença obrigatória em 3 segmentos consecutivos) é muito eficiente e pode ser aplicada após qualquer função de discriminação, mesmo a mais simples (A). Os valores resultantes aproximam-se bastante dos valores obtidos com a função D, significando que as contagens convergem para o que seria a contagem real.

## Viagem 2 - sentido descendente, ensaio em 7-Set-2015 iniciado às 19:35h

Fazendo o mesmo ensaio para a segunda viagem, obtiveram-se os resultados que se apresentam na Tabela 18 e Figura 17.

Viagem-2 DESC Segmentos	Paragem Origem	Função de Discriminação			
		A	B (-10s)	C (-50m)	D (-100m)
1	Parque Nacoes Sul	2	2	2	2
2	Hosp. Descobertas	6	2	2	2
3	Av. Fernando Pessoa	7	3	3	3
4	Oceanario	4	4	4	4
5	Av. Mediterraneo	10	5	6	4
6	R. Caribe	20	10	9	5
7	Est. Oriente	25	18	16	13
8	Est. Oriente	20	15	15	13
9	Av. D. Joao II	13	7	7	5
10	Av. Boa Esperanca	16	8	7	5
11	Cais Olival	10	4	2	2
12	R. Ilha Amores	4	3	2	2
13	R. Moscavide	4	2	2	2
14	Passeio Garcas	4	2	2	2
		<b>145</b>	<b>85</b>	<b>79</b>	<b>64</b>

Tabela 18: Distribuição de passageiros por segmento, viagem 2

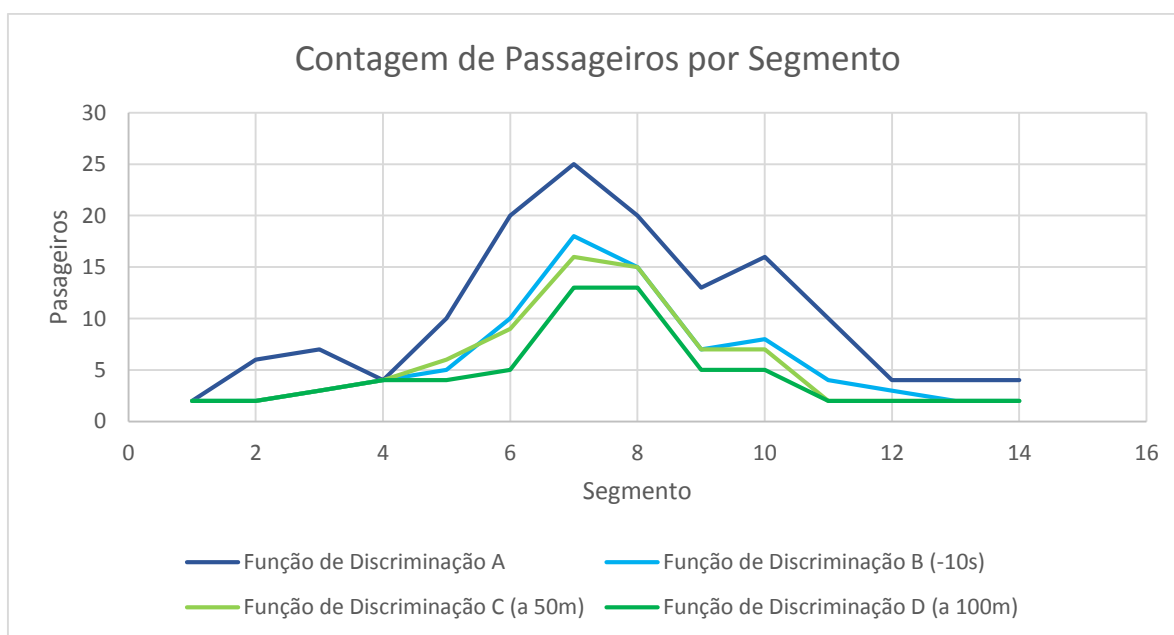


Figura 17: Distribuição de passageiros por segmento, viagem 2

Analisando a Tabela 18 e Figura 17 verifica-se que neste ensaio existe um maior equilíbrio entre as funções B, C e D embora se mantenha o problema evidenciado no ensaio anterior para a zona da Gare Oriente. Esta maior aproximação pode ser explicada pelo menor “ruído” envolvente, ou seja, menor número de transeuntes e de maior fluidez no trânsito devido à hora de realização deste ensaio.

Aplicando o filtro dos três segmentos, que obriga a uma validação do terminal ao longo de três segmentos sucessivos, é possível discriminar um maior número de falsos positivos, aproximando ainda mais às contagens reais.

Viagem-2 DESC Segmentos	Paragem Origem	Aplicação do filtro por nº de segmentos			
		A (2 segm.)	A (3 segm.)	D (2 segm.)	D (3 segm.)
1	Parque Nacoes Sul	2	2	2	2
2	Hosp. Descobertas	6	2	2	2
3	Av. Fernando Pessoa	7	2	3	2
4	Oceanario	4	2	4	2
5	Av. Mediterraneo	10	2	4	2
6	R. Caribe	20	3	5	2
7	Est. Oriente	25	3	13	2
8	Est. Oriente	20	5	13	4
9	Av. D. Joao II	13	4	5	4
10	Av. Boa Esperanca	16	4	5	4
11	Cais Olival	10	2	2	2
12	R. Ilha Amores	4	2	2	2
13	R. Moscavide	4	2	2	2
14	Passeio Garcas	4	2	2	2
		<b>145</b>	<b>37</b>	<b>64</b>	<b>34</b>

Tabela 19: Comparação entre filtros utilizados (viagem2)

Em relação a estes resultados, a conclusão é semelhante ao do ensaio anterior e revela uma convergência dos resultados após a filtragem com 3 segmentos.

## 5.5 RESULTADOS DO ENSAIO COM AUTOCARRO

No dia seguinte e no mesmo horário, adotou-se a metodologia utilizada no ensaio anterior, mas com a diferença de que neste ensaio, todo o equipamento foi colocado dentro do autocarro com recurso a uma pequena UPS de 600VA. Esta UPS tem autonomia para alimentação do AP durante uma viagem em cada sentido. Segundo as especificações, a UPS suporta 120W entre 10 a 15 minutos. O AP tem um consumo máximo de 18,5W, pelo que teremos uma autonomia entre 1 a 1,5 horas.

### Viagem 3 - sentido ascendente, ensaio em 8-Set-2015 iniciado às 19:22h.

Na Tabela 20 apresentam-se as inferências automatizadas para as funções de discriminação A e B, com aplicação posterior do filtro de 2 e 3 segmentos.

Viagem-3 ASC Segmento	Paragem Origem	Filtro			
		2 segm.	3 segm.	2 segm.	3 segm.
		Função			
		A	A	B	B
1	Rossio de Levante	4	2	2	2
2	Passeio Garcas	5	2	2	2
3	R. Ilha Amores	4	2	3	2
4	Cais Olival	3	2	3	2
5	Cais Olival	5	3	5	3
6	Av. Boa Esperanca	10	3	5	3
7	Av. D. Joao II	28	10	19	8
8	Est. Oriente	31	10	25	8
9	Est. Oriente	24	10	17	7
10	R. Caribe	11	3	7	3
11	Av. Mediterraneo	9	3	5	3
12	Oceanario	15	5	8	4
13	Av. Fernando Pessoa	10	4	6	4
14	H. Descobertas	7	4	6	4
15	Parque Nacoes Sul	6	3	5	3
		<b>172</b>	<b>66</b>	<b>118</b>	<b>58</b>

Tabela 20: Distribuição de passageiros por segmento, viagem 3



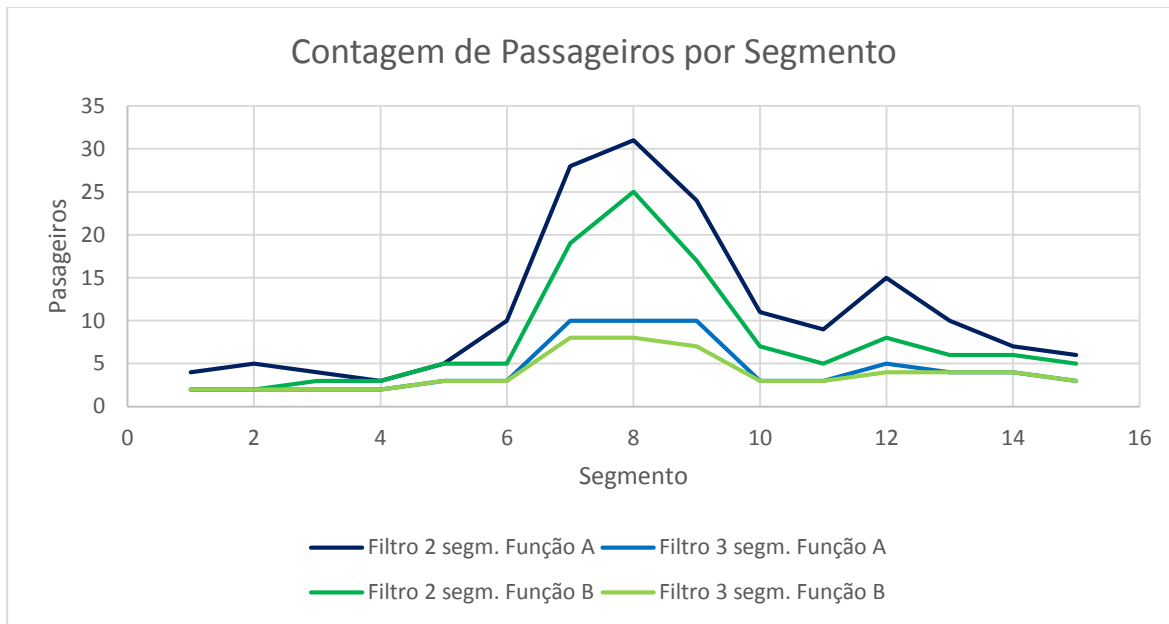


Figura 18: Distribuição de passageiros por segmento, viagem 3

#### Viagem 4 - sentido descendente, ensaio iniciado às 19:42h.

De igual modo se procedeu ao ensaio para o percurso descendente dentro do autocarro, cujos resultados se apresentam na Tabela 21 e Figura 19.

Viagem-4 ASC Segmento	Paragem Origem	Filtro			
		2 segm.	3 segm.	2 segm.	3 segm.
		Função			
		A	A	B	B
1	Parque Nacoes Sul	2	2	2	2
2	Hosp. Descobertas	7	2	3	2
3	Av. Fernando Pessoa	11	3	5	3
4	Oceanario Lisboa	7	3	4	3
5	Av. Mediterraneo	12	4	7	3
6	R. Caribe	21	4	16	4
7	Est. Oriente	21	5	18	5
8	Est. Oriente	23	6	19	6
9	Av. D. Joao II	16	4	13	4
10	Av. Boa Esperanca	8	4	6	4
11	Cais Olival	5	2	3	2
12	R. Ilha Amores	2	2	2	2
13	R. Moscavide	7	2	2	2
15	Passeio Garcas	7	2	2	2
		<b>149</b>	<b>45</b>	<b>102</b>	<b>44</b>

Tabela 21: Distribuição de passageiros por segmento, viagem 4

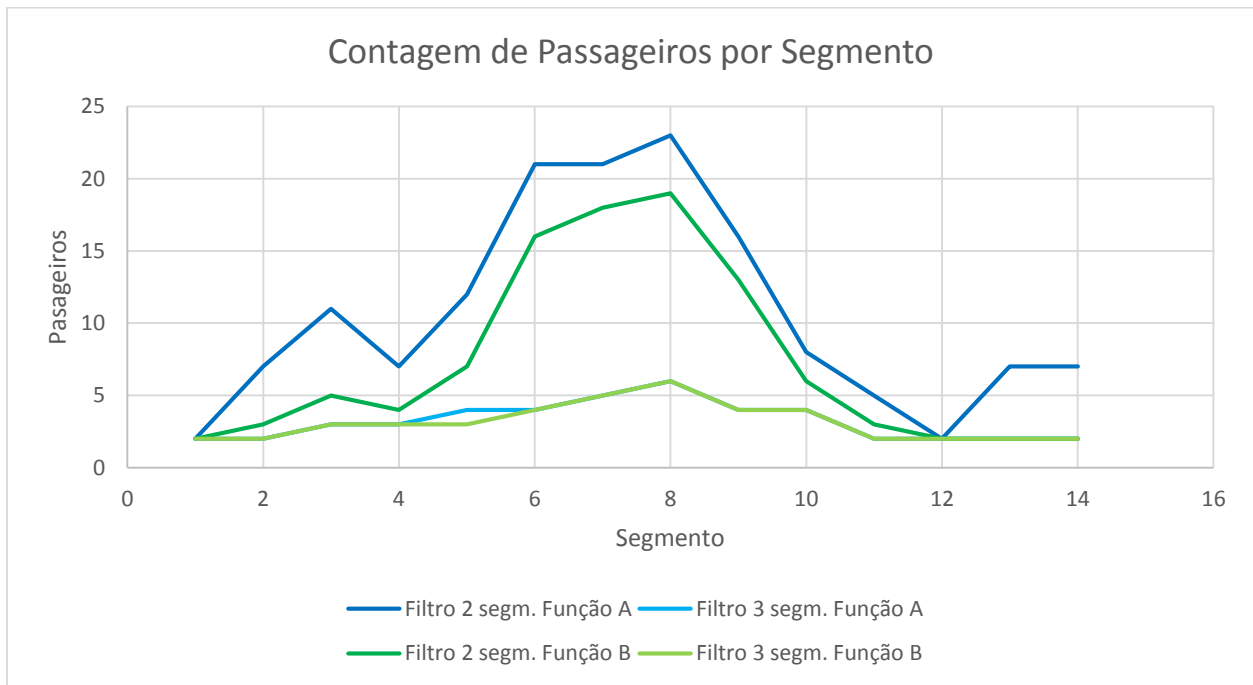


Figura 19: Distribuição de passageiros por segmento, viagem 4

Comparando os valores obtidos nestas duas viagens, para as duas funções discriminantes e para os dois tipos de filtragem, constata-se uma situação semelhante ao do ensaio anterior, utilizando a viatura própria, em que a escolha do filtro com 3 segmentos foi eficaz para resolver o problema do percurso circular na zona da Gare do Oriente - segmentos 6 a 10.

## 5.6 ANÁLISE DOS RESULTADOS

Embora os ensaios tenham sido executados em dias diferentes, utilizaram-se os mesmos procedimentos práticos de forma a não interferir nos resultados. As viagens foram efetuadas na mesma carreira e no mesmo horário (embora no segundo dia o autocarro se tenha atrasado cerca de 10 minutos), correspondendo a dois dias consecutivos de trabalho, respetivamente segunda e terça-feira.

De forma a comparar os resultados dos ensaios com viatura própria e com o autocarro, os gráficos das Figuras 20 e 21 apresentam as duas funções A e B com os filtros para 3 segmentos, para os percursos ascendente e descendente, respetivamente.

Nas legendas das figuras, temos para as quatro funções:

VP = Viatura Própria;

AC = Autocarro;

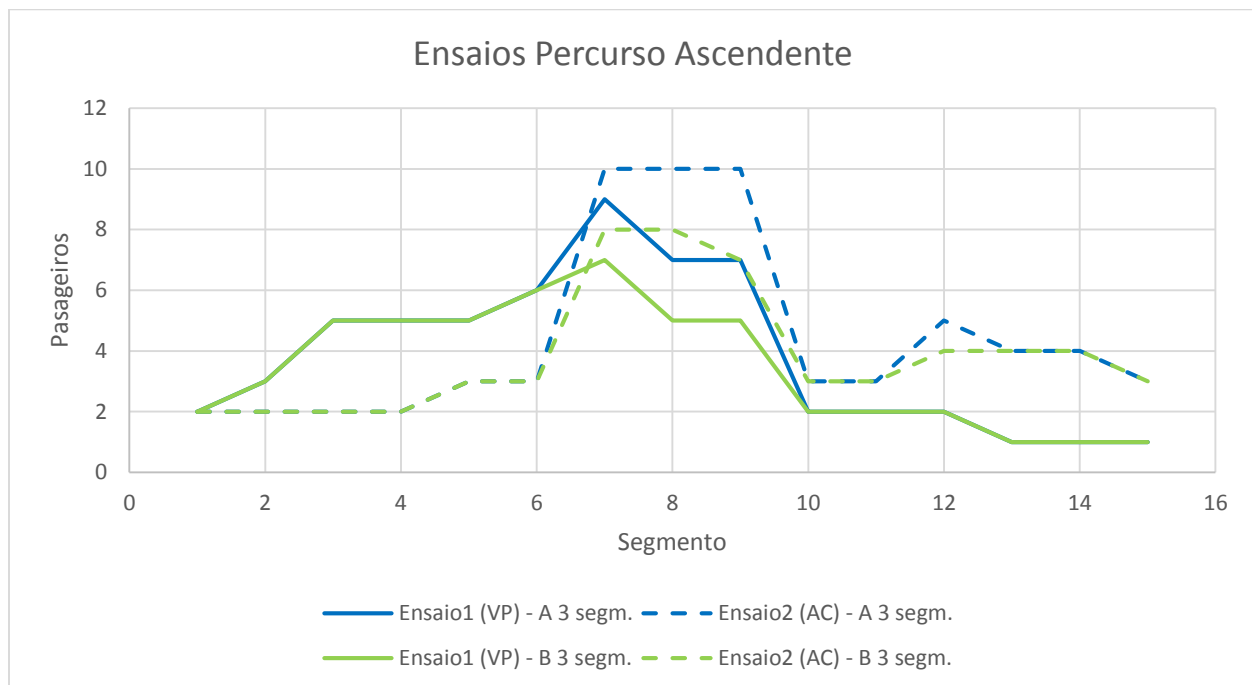


Figura 20: Comparação de ensaios – percurso ascendente

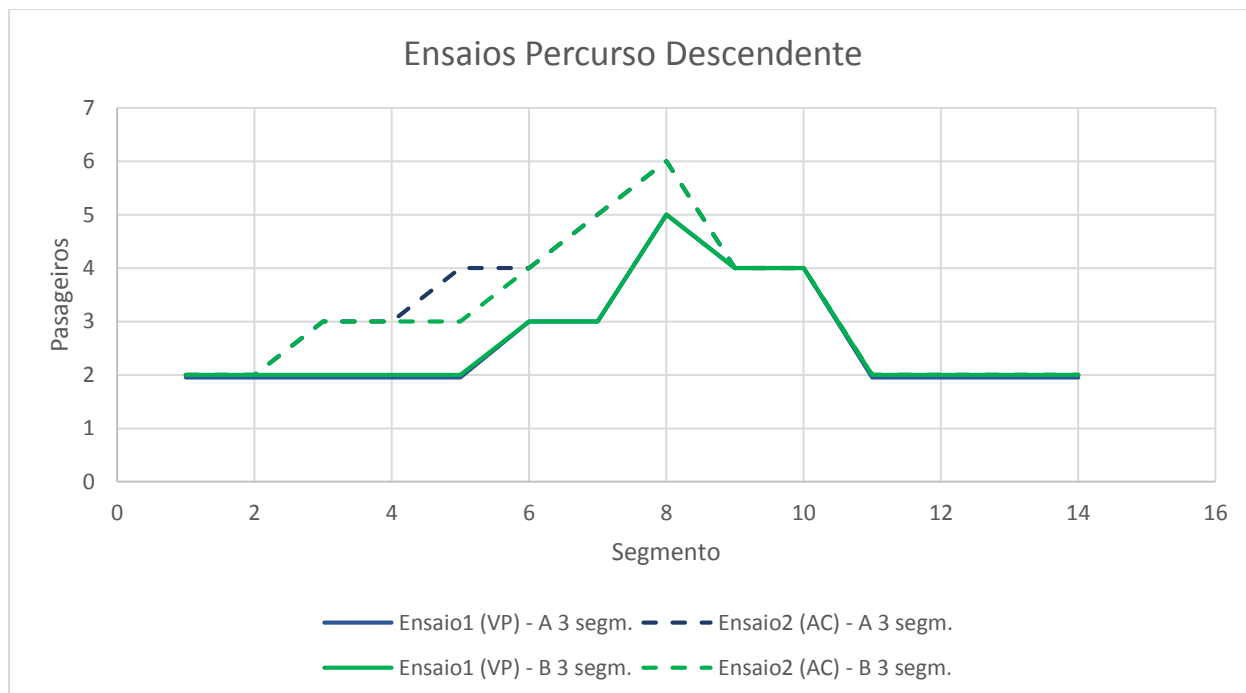


Figura 21: Comparação de ensaios – percurso ascendente

Analisando os gráficos para o 1º e 2º ensaios, verifica-se que, devido ao índice reduzido de deteções, uma pequena variação de 1 ou 2 unidades torna o erro muito significativo. Esta situação não permite retirar conclusões muito assertivas, embora se possa explicar alguns fenómenos entre os dois ensaios:

O facto da localização do AP estar dentro do autocarro no segundo ensaio, enquanto no primeiro ter estado no exterior, é certamente uma das razões para a diferença de resultados (Figuras 20 e 21). Usando a viatura própria, com o AP no exterior do autocarro, o registo de terminais móveis era mais abrangente na periferia do autocarro sendo necessário recorrer aos filtros para se obter resultados mais fidedignos, como se evidencia pela diferença entre as curvas a cheio vs tracejado. No caso do AP no interior do autocarro, é possível que o ruído causado pelos equipamentos do exterior tenha diminuído ligeiramente, mas é certo que o registo dos terminais no seu interior melhorou substancialmente.

Para comparação dos dados inferidos com a contagem real, a mesma apenas pode ser feita com os dados do segundo ensaio (AP dentro do Autocarro).

Nas Figuras 22 e 23, apresenta-se os gráficos com a comparação das inferências face à contagem manual, para os ensaios efetuados com o autocarro em ambos os sentidos.

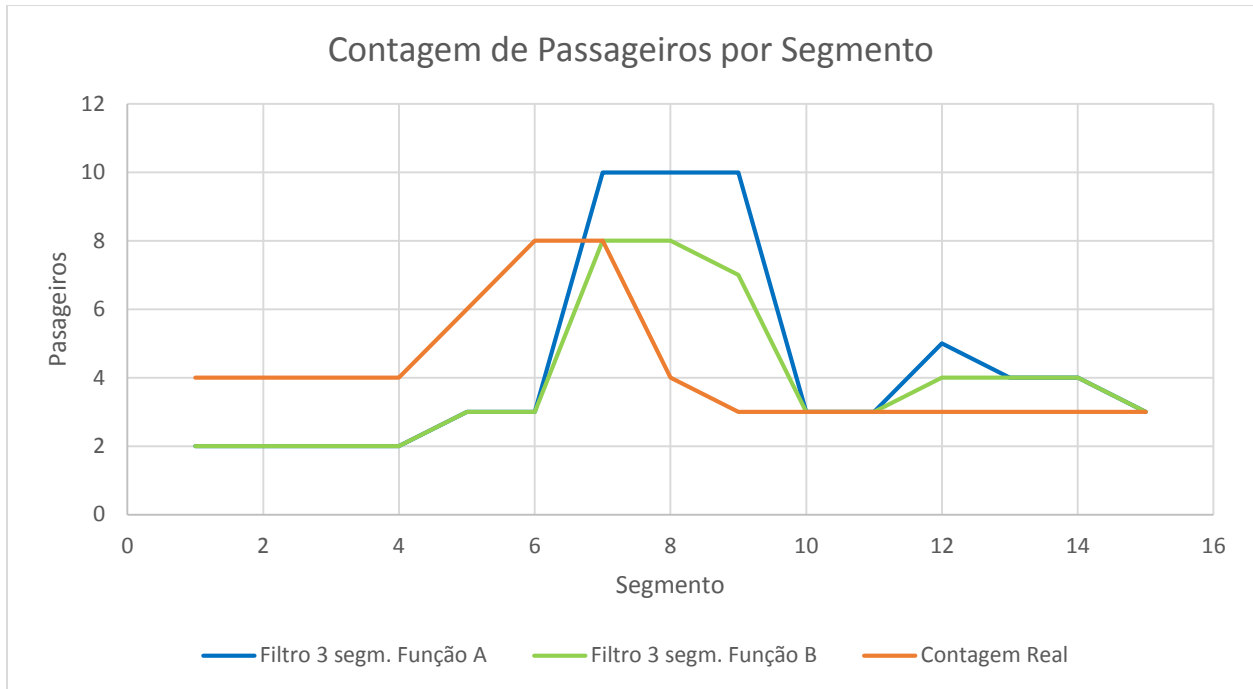


Figura 22: Comparação com Contagem Real – percurso ascendente

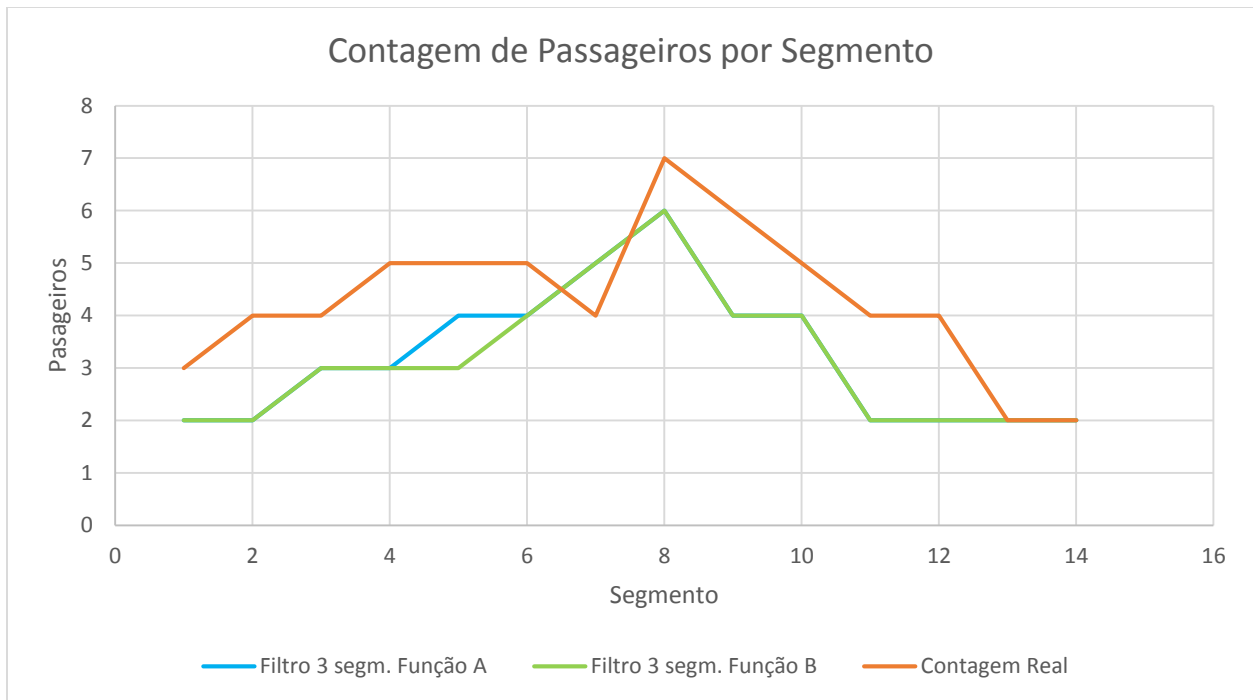


Figura 23: Comparação com Contagem Real – percurso descendente

Uma vez dentro do autocarro, foi possível ter uma ideia mais clara do número real de passageiros a bordo, no entanto em situações de maior afluência nas entradas e saídas, não foi possível registar quem de facto foi o passageiro que entrou/saiu devido à rápida movimentação simultânea dos passageiros. Assim, registou-se apenas o número de entradas e saídas referente a cada paragem que resulta no total por segmento.

Após análise dos valores das contagens e inferências, conclui-se que após o processo de filtragem, o número total de passageiros inferidos aproximam-se da contagem efetuada, mas existe uma variação do erro das inferências ao longo do percurso, sendo visível, por exemplo, um desfasamento da curva das inferências do percurso ascendente.

Este erro pode ser explicado pela não periodicidade das mensagens *Probe Request*, provocando a deteção tardia dos terminais, nos segmentos seguintes.

Por outro lado, caso uma viatura acompanhe o autocarro, esse terminal pode ser contabilizado, gerando erros de inferência em excesso.

De notar que devido ao reduzido número de passageiros, uma não deteção pelo simples facto do utente ter o terminal com WiFi desligado, provoca um grande erro total.

Adicionalmente, para uma análise estatística coerente, seria necessário um volume de dados maior, só possível com a realização de um maior número de ensaios em percursos de maior utilização, usando métodos de contagem automáticos para servir de base de comparação.

Na impossibilidade de realização de tais ensaios como já evidenciado pela ausência de recursos da Tecmic/Carris, prossegue-se com a análise com os dados existentes.

Normalmente utiliza-se a matriz de confusão e curvas ROC para medir o desempenho de um sistema de predição. Mas devido às características do projeto e natureza dos ensaios, este não pode ser considerado um ensaio de classificação supervisionado, uma vez que se desconhece os números individuais de Verdadeiros Positivos e Falsos Positivos. As inferências das contagens são sempre a soma de VP + FP.

Assim, para aferir a precisão usarei a seguinte fórmula simplificada:

$$Precisão = 1 - \frac{|N^{\circ} \text{ Contagens Reais} - N^{\circ} \text{ Inferências}|}{N^{\circ} \text{ Contagens Reais}}$$

Apresenta-se nas Figuras 24 e 25 as matrizes de confusão simplificadas, para os percursos ascendente e descendente, onde a precisão média total calculada para cada percurso é respetivamente 60% e 71%

## 5 Prova de Conceito

Percurso Ascendente Autocarro			Inferência por Segmento - Função Verde a 100m														
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
			2	2	2	2	3	3	8	8	7	3	3	4	4	4	3
Contagem Manual por Segmento	1	4	50%														
	2	4		50%													
	3	4			50%												
	4	4				50%											
	5	6					50%										
	6	8						38%									
	7	8							100%								
	8	4								0%							
	9	3									0%						
	10	3										100%					
	11	3											100%				
	12	3												67%			
	13	3													67%		
	14	3														67%	
	15	3															100%

Figura 24: Matriz de confusão para percurso Ascendente no Autocarro

Percurso Descendente Autocarro			Inferência por Segmento - Função Verde a 100m													
			1	2	3	4	5	6	7	8	9	10	11	12	13	14
			2	2	3	3	3	4	5	6	4	4	2	2	2	2
Contagem Manual por Segmento	1	3	67%													
	2	4		50%												
	3	4			75%											
	4	5				60%										
	5	5					60%									
	6	5						80%								
	7	4							75%							
	8	7								86%						
	9	6									67%					
	10	5										80%				
	11	4											50%			
	12	4												50%		
	13	2													100%	
	14	2														

Figura 25: Matriz de confusão percurso Descendente no Autocarro

Em complemento à abordagem atual desta prova de conceito, de utilizar processos de filtragem dos terminais por segmentos, propõe-se uma alternativa como trabalho futuro, com a utilização dos valores de potência de sinal recebido RSSI.

Os terminais seriam discriminados comparando o valor RSSI com um valor limite de referência.

Com esta nova abordagem, não será possível resolver o problema de fundo que é a não obrigatoriedade do *probing* pelos terminais, mas a discriminação pode ser melhorada sem penalizar os terminais que efetuam muito poucas mensagens de *probing*, os quais seriam excluídos das contagens inferidas pelo método anterior.

De modo a obter estes valores de forma viável, com baixos custos de implementação, a solução ideal seria a utilização equipamento WiFi com *Software* de código aberto.

Com os equipamentos utilizados neste projeto, estas medidas apenas estariam disponíveis utilizando o AP em modo *Lightweight* em conjunto com um controlador de rede WiFi Cisco. Através do controlador, utilizando o processo Syslog de modo a extrair as mensagens de debug com os valores RSSI, estes poderão ser utilizados como fator principal de discriminação. Inicialmente esta solução foi preterida devido aos custos de *hardware* e licenciamento envolvidos.

No entanto, pretendeu-se a estudar a viabilidade desta nova abordagem. Verifiquei o tipo de dados que poderia obter efetuando um ensaio com o AP configurado em modo *Lightweight*, com recurso a um controlador virtual WLAN.

No essencial, a configuração do syslog no controlador WLAN foi a seguinte:

```
config logging syslog host aaa.bbb.ccc.ddd
config logging syslog facility local7 #{0..7} 0=emergency 7=debug
config logging syslog level 7          #{0..7} 0=emergency 7=debug
config logging debug syslog enable
```

Para se obter os dados pretendidos, ativa-se o debug para as mensagens LBS:

```
debug location client enable
```

```
<175>Cisco_b4:91:5a: *apfProbeThread: Sep 12 21:35:08.251: e8:99:c4:6e:98:e3
LBS Client e8:99:c4:6e:98:e3 pkt from AP 00:24:14:86:06:f0(1) RSSI (A -52, B -
57), SNR 46, Chan 0, Pwr -128
```



De modo a avaliar o tipo de sinal obtido, submeteram-se dois terminais num ensaio de 30 minutos com distâncias fixas do AP, num espaço fechado: respetivamente 6 metros e 20 metros, este último com obstáculos (paredes). As antenas utilizadas no AP são dipolares e com um ganho de 2,5dBi.

Nos gráficos seguintes ilustrados pelas Figuras 26 e 27 apresenta-se as curvas de variação do sinal RSSI, onde se incluiu também para análise o cálculo da média móvel acumulada do valor instantâneo.

Gráfico para terminal no alcance do AP, a 6 metros durante 30 minutos:

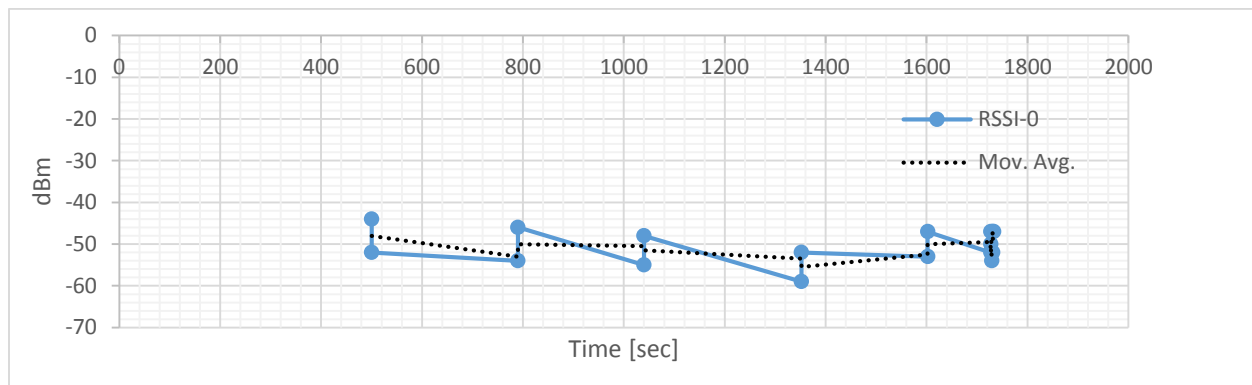


Figura 26: Gráfico do valor RSSI para um terminal fixo a 6 metros do AP

Gráfico para terminal no limite do alcance do AP, durante 30 minutos:

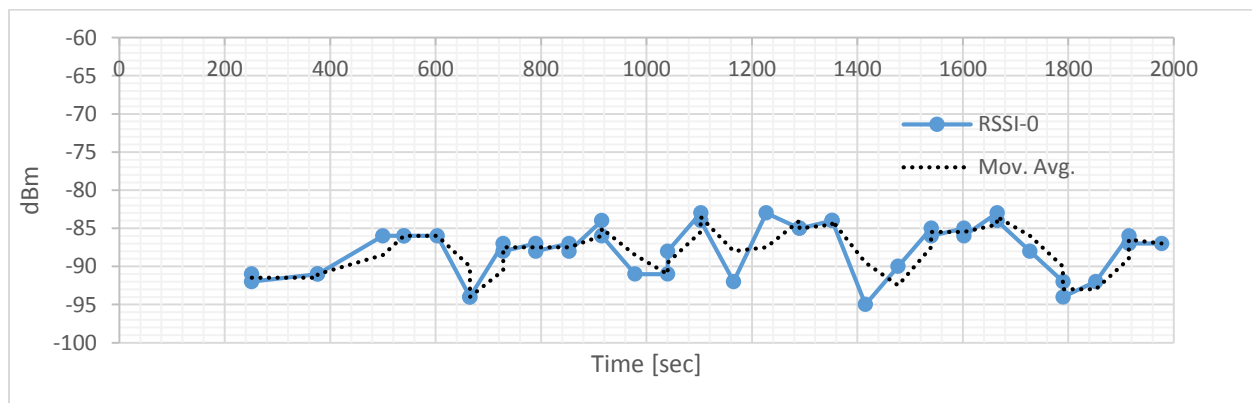


Figura 27: Gráfico do valor RSSI para um terminal no limite do alcance do AP

Normalmente utiliza-se o valor limite de referência de -70dBm, como potência mínima de receção para transmissão de dados, mas no entanto o terminal pode ser detetado com

valores inferiores pelas mensagens de gestão. Este valor deve ser utilizado como limite de referência inicial.

Mas para ajustar melhor este valor à realidade do espaço dentro do autocarro, seria necessário efetuar um ensaio de medição de modo a garantir uma correta discriminação para passageiros fora do autocarro. Este valor dependerá naturalmente da localização do AP e do ganho das antenas.

Embora o valor RSSI tenha oscilações grandes num curto espaço de tempo, este método afigura-se viável utilizando uma função simples para estabilizar o sinal tal como a media móvel.

Para esse cálculo, seria necessário acrescentar um novo campo na tabela WiFi para alojar o valor RSSI do terminal para cada mensagem recebida. No momento da agregação dos dados para a janela de tempo pretendida, calcular-se-ia a média dos valores RSSI desse intervalo de tempo. Esse valor seria então registado na base de dados. Os passageiros poderiam assim ser validados caso o valor médio do sinal RSSI se encontre dentro do valor de referência estabelecido.

## 6 CONCLUSÕES

---

O objetivo principal deste trabalho resultou de uma necessidade real duma empresa que presta serviços de informação na área dos sistemas de transportes, a Tecmic.

A necessidade real é encontrar uma solução de contabilização de baixo custo, escalável, e de rápida implementação. Os sistemas de contabilização automáticos que se baseiam em sensores físicos (sensor de pressão, sensores fotoelétricos) ou por visão artificial (luz visível ou por infravermelhos), normalmente são dispendiosos e requerem processos de ajuste e adaptação ao meio, de modo as evitar interferências da luz solar, quer por reflexão, quer por incidência direta no campo de abertura da lente do sensor. Este tipo de sensores são utilizados com bastante sucesso em espaços indoor em que a iluminação é artificial e controlada. Em espaços exteriores e em especial num veículo móvel como o autocarro, existe sempre a possibilidade de interferência pela luz solar que provoca a impossibilidade de deteção dos passageiros no momento exato da abertura da porta para entrada e/ou saída dos passageiros.

Como alternativa, a Tecmic procura uma solução alternativa de baixo custo baseada na deteção dos sinais WiFi como forma de contabilização de passageiros. Embora assente no pressuposto que a maioria dos passageiros usa dispositivos móveis com Wi-Fi ativo, o valor desta solução será muito inferior ao de outras soluções existentes no mercado.

Uma das aplicações diretas deste trabalho é o melhoramento do ranking dos condutores da Carris tendo em conta o trabalho da Eco-Condução desenvolvido nos últimos anos e referido anteriormente em 2.2 [1]. O número de passageiros afeta diretamente o consumo e esta informação é importante para a determinação desse ranking.

Assim, o âmbito deste trabalho foi o de criar uma prova de conceito para servir de base a uma possível implementação no terreno. Esta proposta prevê a interligação com a base de dados XTraN Passenger, a introdução de um Access Point WiFi no autocarro, o qual poderá ser integrado com o sistema embarcado por via de mensagens Syslog.

A componente principal deste trabalho foi descrever os principais casos de utilização, prevendo os principais problemas a resolver, entre os quais a discriminação dos terminais dentro e fora do autocarro. Consequentemente foram implementados os algoritmos de inferência e regras de filtragem para uma contabilização o mais precisa possível.

Como ferramentas de trabalho, foram escolhidas ferramentas de código aberto, nomeadamente a base de dados geográfica PostgreSQL/PostGIS e o QuantumGIS para visualização de dados. Toda a programação assenta essencialmente *scripts* em PGSQL, com as exceções em que havia necessidade de conversão de dados para ingestão, onde foi utilizado Python com o extensor PsyCpg2.

## 6.1 RESULTADOS OBTIDOS:

Através dos ensaios efetuados no terreno, foi possível aferir a viabilidade desta solução. Os resultados das comparações entre as inferências automáticas por monitorização WiFi, com as medidas reais efetuadas manualmente permitiram introduzir alguns ajustes nas funções de discriminação de modo a resolver casos não previstos inicialmente, como foi o caso do percurso circular numa zona de elevada aglomeração de pessoas.

O erro provocado por este fenómeno foi eficazmente diminuído, após uma filtragem por segmentação dos passageiros (definido em 5.3 e ilustrado na Figura 16), em relação ao número de segmentos consecutivos de presença obrigatória do passageiro, de forma a este poder ser validado como tal.

No entanto, após comparação com as contagens reais e pela observação dos fluxos de entradas e saídas, verifica-se que as inferências pela deteção WiFi passiva sofre maioritariamente com o problema da frequência das mensagens de *probing* do terminal, essenciais para a correta deteção. Este problema pode ter sido amplificado pela filtragem adicional efetuada.

Em alternativa ou em complemento, ajustando o ganho das antenas do AP para o caso real do autocarro, diminuindo o alcance do sinal WiFi, quer por via do diagrama de radiação, quer por atenuação do ganho das antenas, pode-se reduzir substancialmente a deteção dos falsos positivos.

Adicionalmente, caso as medidas RSSI do terminal estivessem disponíveis, seria possível melhorar a discriminação dos terminais falsos positivos. Para que isso fosse possível, seria necessário utilizar um AP que disponibilizasse essa informação nos ficheiros de coleta de dados ou, em alternativa, usar um AP com *software* de código aberto.

Assumindo que o terminal que está dentro do autocarro tem uma posição relativa ao AP predominantemente estática, seria suposto que o sinal RSSI desse terminal tivesse igualmente um comportamento estável. Tal como descrito em 2.2.1, Figura 4 [18], estas medidas podem apresentar uma grande variância. No entanto, recorrendo a um mecanismo simples para amortecer estas variações, por exemplo, o cálculo da média móvel ou mesmo da média linear, podemos obter um sinal mais estável, passível de ser comparado com um valor pré-definido e validarmos se esse terminal estaria dentro ou fora do autocarro.

No entanto, para efetuar estes ajustes, seria necessário realizar um número superior de ensaios durante um período de tempo suficientemente longo para se recolher um volume de dados que permita uma análise estatística com maior confiança. Esses ensaios seriam realizados de preferência nos autocarros que já disponham de um sistema de

contagem automática por visão artificial. Assim seria possível aferir e comparar os resultados dos dois sistemas sem recurso a contabilizações manuais.

Uma das maiores dificuldades foi exatamente obter as respetivas autorizações da Tecmic e Carris, bem como obter os recursos para efetuar os ensaios necessários, os quais, fruto do processo de privatização recente da Carris, originou atrasos na tomada de algumas decisões o que inviabilizou a realização dos ensaios necessários.

### 6.2 TRABALHO FUTURO

Como trabalho futuro está previsto durante o próximo ano uma parceria entre a Tecmic e o ISEL para suporte à implementação desta solução no sistema XTraN, tendo como base o presente trabalho como prova de conceito, de modo a realizar ensaios em maior número e duração, com a finalidade de aferir o impacto da integração no seu sistema.

Adicionalmente, quando na posse de dados suficientes sobre as contagens de passageiros na rede da Carris, será possível estender o trabalho para inferir possíveis relações com acontecimentos externos ao sistema, nomeadamente fenómenos climatéricos, eventos desportivos, períodos sazonais, etc.

Outra via é o estudo da inferência do número de utentes em cada componente de uma carreira de autocarro ao longo do dia, da semana, fins de semana ou feriados, informação importante para o planeamento de rotas e horários dos transportes.

Essa informação é claramente uma mais-valia para a Tecmic e para a Carris.



## **7 ANEXOS**

---

### **7.1 ANEXO 1 - XTRAN SQL QUERY VIEW**

Por favor consultar CD do Projeto na pasta Anexos1

### **7.2 ANEXO 2 – MODIFICAÇÃO AP DE *LIGHTWEIGHT* PARA *AUTÓNOMO***

Por favor consultar CD do Projeto na pasta Anexos2

### **7.3 ANEXO 3 - CONFIGURAÇÃO DO AP1252 EM MODO AUTÓNOMO**

Por favor consultar CD do Projeto na pasta Anexos3

### **7.4 ANEXO 4 – ROTAS CARREIRA Nº400 CARRIS**

Por favor consultar CD do Projeto na pasta Anexos4

### **7.5 ANEXO 5 – CÓDIGO PARA MANIPULAÇÃO E INGESTÃO DE DADOS**

Por favor consultar CD do Projeto na pasta Anexos5

### **7.6 ANEXO 6 – CÓDIGO PARA CORRELAÇÃO DE DADOS**

Por favor consultar CD do Projeto na pasta Anexos6

### **7.7 ANEXO 7 – CÓDIGO PARA FILTRAGEM DE DADOS**

Por favor consultar CD do Projeto na pasta Anexos7

## **7.8 ANEXO 8 – QUERIES PARA EXTRAÇÃO DE INFORMAÇÃO**

Por favor consultar CD do Projeto na pasta Anexos8



## 8 BIBLIOGRAFIA

---

1. João Ferreira, José Almeida, Alberto Silva. “The Impact of Driving Styles on Fuel Consumption: A Data Warehouse and Data Mining based Discovery Process”, IEEE Transactions on Intelligent Transportation Systems Magazine on Intelligent Transportation Systems in March 2015. DOI 10.1109/TITS.2015.2414663.
2. J. Almeida and J. Ferreira, “BUS Public Transportation System Fuel Efficiency Patterns,” in proceedings of the 2nd International Conference on Machine Learning and Computer Science (IMLCS'2013), 2013, Malaysia
3. Tecmic - XTran Gestão de Frotas, acessido em Fevereiro 2015 pela ligação <http://www.tecmic.pt/pt-br/portfolio/xtran/>
4. ANACOM - Dados estatísticos de serviços moveis 3T2014, acessido em Maio 2015 pela ligação [http://www.anacom.pt/render.jsp?contentId=1339776#.VeDza\\_IViko](http://www.anacom.pt/render.jsp?contentId=1339776#.VeDza_IViko)
5. MARKTEST - Estudo Barómetro de Telecomunicações, acessido em Maio 2015 pela ligação <http://www.marktest.com/wap/a/n/id~1dba.aspx>
6. Cisco Location Based Services - Solution Architecture, acessido pela ligação <http://www.cisco.com/en/US/docs/solutions/Enterprise/Mobility/emob30dg/Locatn.html>, em Março 2015
7. Cisco Compatible Extension (CCX) Overview, acessido em Março 2015 pela ligação [https://www.cisco.com/web/partners/downloads/765/ccx/Comp\\_Ext\\_Cust\\_Preso.pdf](https://www.cisco.com/web/partners/downloads/765/ccx/Comp_Ext_Cust_Preso.pdf)
8. Cisco Wireless LAN Controller Software Documentation, acessido em Março 2015 pela ligação <http://www.cisco.com/c/en/us/support/wireless/wireless-lan-controller-software/tsd-products-support-series-home.html>
9. Lightweight AP (LAP) Registration to a Wireless LAN Controller (WLC), Algoritmos de Descoberta, Seleção e Ligação, e acessido em Abril 2015 pela ligação <http://www.cisco.com/c/en/us/support/docs/wireless-mobility/wireless-lan-wlan/70333-lap-registration.html>
10. Cisco Wireless LAN Controller Configuration Guide – Release 7.3, acessido em Março 2015 pela ligação [http://www.cisco.com/c/en/us/td/docs/wireless/controller/7-3/configuration/guide/b\\_cg73.html](http://www.cisco.com/c/en/us/td/docs/wireless/controller/7-3/configuration/guide/b_cg73.html)
11. Configuring Probe Request Forwarding (from AP to WLC), Enable, Disable and Filtering, acessido em Março 2015 pela ligação [http://www.cisco.com/c/en/us/td/docs/wireless/controller/7-6/configuration-guide/b\\_cg76/b\\_cg76\\_chapter\\_01110101.pdf](http://www.cisco.com/c/en/us/td/docs/wireless/controller/7-6/configuration-guide/b_cg76/b_cg76_chapter_01110101.pdf)
12. Cisco Aironet 1250 Series Access Point Quick Start Guide, acessido pela ligação [http://www.cisco.com/c/en/us/td/docs/wireless/access\\_point/1250/quick/guide/ap1250qs.html](http://www.cisco.com/c/en/us/td/docs/wireless/access_point/1250/quick/guide/ap1250qs.html), em Março 2015

13. Cisco IOS Software Configuration Guide for Cisco Aironet Access Points – for Cisco IOS Releases 12.4(10b)JA and 12.3(8)JEC, acessido em Março 2015 pela ligação [http://www.cisco.com/c/en/us/td/docs/wireless/access\\_point/12-4\\_10b\\_JA/configuration/guide/scg12410b.html](http://www.cisco.com/c/en/us/td/docs/wireless/access_point/12-4_10b_JA/configuration/guide/scg12410b.html)
14. GTFS - General Transit Feed Specification – Static Transit, acessido em Julho 2014 pela ligação <https://developers.google.com/transit/gtfs/reference>
15. GTFS - General Transit Feed Specification – Realtime Transit, acessido em Julho 2014 pela ligação <https://developers.google.com/transit/gtfs-realtime/>
16. GTFS - Static Transit - UML diagram, and Data examples - acessido pelas ligações [http://www.google.com/help/hc/images/transitpartners\\_1106431\\_objecttablelarge\\_en.gif](http://www.google.com/help/hc/images/transitpartners_1106431_objecttablelarge_en.gif) e <https://developers.google.com/transit/gtfs/examples/gtfs-feed>, em Julho 2014
17. Zhuliang Xu et All, Pedestrian Monitoring System using Wi-Fi Technology And RSSI Based Localization - Active Sniffing method using Probe Responses - University of Technology Sydney, Australia, 2013
18. Lorenz Schauer, Martin Werner, Philipp Marcus - Estimating Crowd Densities and Pedestrian Flows Using Wi-Fi and Bluetooth, Mobile and Distributed Systems Group, LMU Munich, Germany 2014
19. M. V. Barbera, A. Epasto, A. Mei, V. C. Perta, and J. Stefa. Signals from the crowd: Uncovering social relationships through smartphone probes - In Proceedings of the conference on Internet measurement conference, pages 265–276. ACM, 2013
20. B. Bonn´e, A. Barzan, P. Quax, and W. Lamotte. Wifipi: Involuntary tracking of visitors at mass events - In World of Wireless, Mobile and Multimedia Networks (WoWMoM), IEEE 14th International Symposium and Workshops on a, pages 1–6, 2013
21. A. Musa and J. Eriksson. Tracking unmodified smartphones using wi-fi monitors - In Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, pages 281–294, 2012
22. Y. Wang, J. Yang, H. Liu, Y. Chen, M. Gruteser, and R. P. Martin. Measuring human queues using wifi signals - In Proceedings of the 19th annual international conference on Mobile computing & networking, pages 235–238. ACM, 2013
23. Serviço Net Bus, implementado em algumas carreiras de autocarros da Carris, acessido em Junho 2015 pela ligação <http://carris.transporteslisboa.pt/pt/netbus/>
24. HELLA – precision stereo camera for people counting, acessido pela ligação [https://www.hella.com/hella-com/assets/media\\_global/People\\_Counter\\_GB\\_V2.pdf](https://www.hella.com/hella-com/assets/media_global/People_Counter_GB_V2.pdf)
25. geOps - Geoinformatics Service Provider, acessido pela ligação <http://geops.de/blog/mapping-public-transit-networks?language=en>
26. TRAVIC - Realtime movement visualization of transit data, acessido pela ligação <http://tracker.geops.ch/?z=14&s=1&x=-8235743.4976&y=4971840.9526&l=transport>

## 8 Bibliografia

27. WiGLE - Wireless Geographic Logging Engine: Making maps of wireless networks since 2001. Acedido pela ligação <http://www.wigle.net>
28. Google Maps Timeline, acedido pela ligação <https://www.google.pt/maps/timeline>
29. Google Maps Web Services – Geolocation API, acedido pela ligação <https://developers.google.com/maps/documentation/geolocation/intro>
30. Official U.S. Government information about the Global Positioning System (GPS), acedido pela ligação <http://www.gps.gov/systems/gps/performance/accuracy/> em Agosto2015
31. NMEA - International Marine Electronics Association, acedido em Agosto2015 pela ligação [http://www.nmea.org/content/nmea\\_standards/nmea\\_0183\\_v\\_410.asp](http://www.nmea.org/content/nmea_standards/nmea_0183_v_410.asp)
32. Vassilis Kostakos “Wireless detection of passenger trips on public transport buses” LabUSE, University of Madeira, Carnegie Mellon University, 2008