



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA

**USO CONJUNTO DE SMARTCARD Y  
DETECCIONES WIFI PARA ESTIMAR  
INFORMACIÓN OPERACIONAL DE UN  
SISTEMA DE TRANSPORTE PÚBLICO  
EN TIEMPO REAL**

**PABLO IGNACIO GUAJARDO RAMOS**

Tesis para optar al grado de  
Magíster en Ciencias de la Ingeniería

Profesor Supervisor:  
**JUAN CARLOS HERRERA**

Santiago de Chile, abril, 2017.

© 2017, Pablo Guajardo Ramos



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA

# **USO CONJUNTO DE SMARTCARD Y DETECCIONES WIFI PARA ESTIMAR INFORMACIÓN OPERACIONAL DE UN SISTEMA DE TRANSPORTE PÚBLICO EN TIEMPO REAL**

**PABLO IGNACIO GUAJARDO RAMOS**

Tesis presentada a la Comisión integrada por los profesores:

**JUAN CARLOS HERRERA**

**SEBASTIÁN RAVEAU**

**MARCELA MUNIZAGA**

**JORGE VERA**

Para completar las exigencias del grado de  
Magíster en Ciencias de la Ingeniería

Santiago de Chile, abril, 2017

A mis seres queridos, que me apoyaron mucho.

## **AGRADECIMIENTOS**

Agradezco a mis padres, mi hermana, mi pareja y a todos los que formaron parte de esta tesis de alguna u otra forma.

Me gustaría agradecer a todas las empresas externas que hicieron posible que tuviéramos datos reales con los que probar nuestra metodología. Pudimos entregar un trabajo mucho más robusto gracias a las gestiones que hicimos en conjunto para probar la metodología desarrollada en la tesis. En particular, me gustaría agradecer al Centro de Desarrollo Urbano Sustentable (CEDEUS, CONICYT/FONDAP 15110020), por su apoyo financiero y colaboración.

Gracias a los integrantes del Departamento de Ingeniería de Transporte de la Pontificia Universidad Católica de Chile por sus comentarios, tanto de académicos como de alumnos de postgrado, que fueron un gran aporte al desarrollo de esta tesis desde las primeras ideas en conversaciones de pasillo hasta la defensa de la tesis.

Destaco la participación en este trabajo de Juan Carlos Herrera por su rol como profesor guía y como consejero. Varias barreras tanto internas como externas a la investigación fueron superadas por la buena comunicación que tuvimos, su capacidad de mantener la calma y optimismo.

## INDICE GENERAL

	Pág.
DEDICATORIA .....	ii
AGRADECIMIENTOS .....	iii
INDICE DE TABLAS .....	vi
INDICE DE FIGURAS.....	vii
RESUMEN.....	viii
ABSTRACT .....	ix
1. Introducción.....	1
2. Contexto bibliográfico y tecnológico. ....	4
2.1. Situación actual. ....	4
2.2. Descripción de la operación de un WS. ....	5
2.2.1. WS versus BTS.....	8
2.3. Uso de WS para análisis de tráfico peatonal y vehicular. ....	10
2.4. Uso de otras tecnologías en transporte.....	11
2.4.1. Bluetooth.....	11
2.4.2. Señal de celulares. ....	12
2.4.3. Smartcard.....	13
3. Metodología.....	15
3.1. Presentación de las variables.....	17
3.2. Procesos que componen la metodología para estimar matriz OD y carga por arco.....	19
3.2.1. Inicialización de un viaje. ....	19
3.2.2. Actualización de un viaje.....	21
3.2.3. Finalización de viajes. ....	22
3.2.4. Estimación de los ponderadores. ....	23
3.3. Metodología para la estimación de otros resultados. ....	25
3.3.1. Tiempos de permanencia en estación. ....	26

3.3.2. Viajes diarios por dispositivo. ....	26
3.3.3. Días especiales. ....	27
4. Implementación y Prueba Piloto. ....	28
4.1. Descripción de la red. ....	29
4.2. Instalación en terreno. ....	30
4.3. Implementación computacional de la metodología. ....	31
4.3.1. Descripción del programa en el lenguaje de BD PostgreSQL. ....	31
5. Resultados y Discusión. ....	39
5.1. Ponderadores y penetración. ....	39
5.1.1. Penetración a través del tiempo. ....	41
5.1.2. Penetración a través del espacio. ....	47
5.2. Matriz OD. ....	50
5.3. Carga por arco. ....	52
5.4. Tiempos de permanencia en estación. ....	54
5.5. Viajes diarios por dispositivo. ....	56
5.6. Días especiales. ....	57
6. Conclusiones y Futuras aplicaciones. ....	59
6.1. Conclusiones. ....	59
6.2. Futuras Aplicaciones. ....	61
BIBLIOGRAFIA. ....	64
A N E X O S. ....	66
Anexo A: Diferencias de penetración de <i>smartphone</i> por país y características socioeconómicas. ....	67
Anexo B: Programa en C# para controlar el flujo de información. ....	68

## INDICE DE TABLAS

	Pág.
Tabla 2-1: Obstáculos que se encuentran en ambientes de interior que afectan las señales inalámbricas. Fuente: Harwood (2009).....	7
Tabla 5-1: Proporción de viajes realizados por estudiantes para E1, E2, y E3 agrupadas por periodo del día. ....	46
Tabla 5-2: Proporción de frecuencia de viajes para una semana laboral. ....	57

## INDICE DE FIGURAS

	Pág.
Figura 1-1: Porcentaje de adultos que posee <i>smartphone</i> por país. ....	1
Figura 2-1 Comparación de popularidad de uso de Wifi y Bluetooth en lugares peatonales. Fuente: Abedi, Bhaskar y Chung (2013).....	9
Figura 4-1: Ejemplo de datos obtenidos con los WS. ....	28
Figura 4-2: Ejemplo de datos obtenidos con las SC. ....	29
Figura 4-3: Diagrama con flujo de información a través de tablas de base de datos. ....	33
Figura 5-1: Afluencias sumadas de las estaciones E1, E2 y E3 cada 15 minutos para una semana laboral obtenidas a través de la SC. ....	40
Figura 5-2: Afluencias sumadas de las estaciones E1, E2 y E3 para una semana laboral obtenidas a través de los WS.....	41
Figura 5-3: Penetración conjunta de E1, E2 y E3 estimada cada 15 minutos.....	42
Figura 5-4: Detecciones totales de dispositivos y afluencias de SC para las estaciones E1, E2 y E3 agrupadas por periodo del día lunes de la semana analizada. ....	44
Figura 5-5: Detecciones por dispositivo y penetración del día lunes. ....	45
Figura 5-6: Penetración para el lunes de cada estación analizada. ....	48
Figura 5-7: Afluencias de SC por estación por periodo para el lunes.....	50
Figura 5-8: (a) matriz OD de un día martes obtenida con la metodología del operador del STP usando las SC. (b) matriz OD de un día martes obtenida con la metodología de esta tesis.....	51
Figura 5-9: Carga por arco estimada con la metodología de esta tesis y con la del operador del STP para el periodo un día martes entre 17:00 y 20:00. ....	53
Figura 5-10: Tiempo de permanencia en minutos estimado con los WS para el lunes. ....	55
Figura 5-11: Afluencias de SC agrupadas de E1, E2 y E3 de 6 días jueves consecutivos. .	58

## RESUMEN

Existen *scanners* (*Wifi Scanner* o *WS*) capaces de captar las señales Wifi enviadas por los *smartphones* de las personas. Con una red de transporte equipada con dichos *WS*, es posible observar patrones de movimiento a lo largo de la red. Por otro lado, los sistemas de transporte público del mundo están rápidamente adoptando las *Smartcard* (*SC*) como método de pago ya que es una manera barata y segura de recolectar la tarifa.

Este trabajo busca usar conjuntamente estas dos fuentes de datos para estimar información valiosa sobre los viajes de un sistema de transporte público. Se desarrolló una metodología para expandir los patrones de viaje obtenidos por los *WS* usando los conteos por estación de las *SC*. Luego, la metodología fue probada en una prueba piloto con datos reales en un sistema de transporte público.

La expansión debe depender tanto del instante como de la estación por la que el usuario entra al sistema. Las penetraciones observadas de la tecnología Wifi fueron de entre 30% y 70% del total de usuarios, dependiendo del tiempo y el espacio. Utilizar esta metodología permite, en tiempo real, estimar la matriz OD, estimar cargas por arco, analizar días en que ocurre alguna contingencia, observar la ruta que eligen usuarios entre un mismo par OD, estimar tiempos de espera, entre otros.

Palabras Claves: Wifi, Smartcard, Intelligent Transport System, transporte público.

## **ABSTRACT**

There are scanners (Wifi scanner or WS) capable of detecting Wifi signals sent by smartphones owned by people. With a WS equipped transport network, it is possible to observe movement patterns throughout the network. On the other hand, public transport networks around the world are rapidly adopting Smartcard (SC) as payment method due to it being a cheap and secure way of collecting fare.

This work seeks to fusion these two sources of data to estimate valuable information about the trips of a public transport system. A methodology was developed to expand the travel patterns retrieved by the WS using counts by station from the SC. The methodology was proven in a pilot test, with real data from a public transport system.

The expansion should depend as much on the instant as on the station through which the user enters the system. The penetrations observed with Wifi technology were between 30% and 70% of the total of users, depending on time and space. Using this methodology allows to, in real time, estimate the OD matrix and link load, analyze days where some contingency happens, observe the route followed by users from the same OD pair, estimate waiting times, among others.

Keywords: Wifi, Smartcard, Intelligent Transport System, public transport.

## 1. INTRODUCCIÓN.

La tecnología celular se transforma cada día más en parte esencial de la vida cotidiana de los individuos, ya no solo para comunicarse directamente con otra persona, sino que para enterarse de lo que sucede en el mundo, publicar fotografías en redes sociales, comentar sobre algún hecho de interés, entre otros. Pew Research Center (2016) realizó un estudio a nivel global donde se puede ver el porcentaje de adultos por país que posee *smartphone* (ver Figura 1-1). Chile es uno de los países en vías de desarrollo con mayor penetración de *smartphone* (65%), muy cerca de países desarrollados como Estados Unidos (72%) o el Reino Unido (68%) y superando a otros como Alemania (60%) y Japón (39%). Además, en conjunto con un estudio anterior (Pew Research Center, 2014), se puede ver que Chile aumentó su penetración de *smartphone* de un 39% en 2013 a un 65% en 2015.

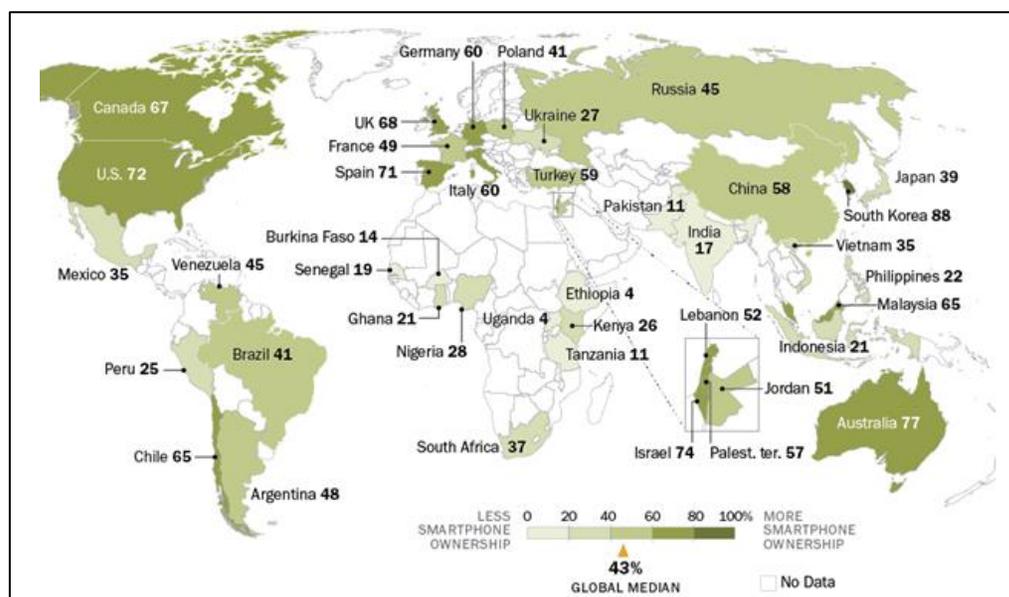


Figura 1-1: Porcentaje de adultos que posee *smartphone* por país. Fuente: Pew Research Center, 2016.

Estos dispositivos<sup>1</sup> están constantemente entregando y recibiendo señales ya sea porque el usuario ejecutó una acción o porque el dispositivo está pasivamente enviando señales protocolares. Estas señales pueden ser captadas por detectores de señales Wifi o *Wifi Scanner* (WS) y luego identificar al mismo dispositivo en otros WS, ya que la señal incluye un identificador único del dispositivo. Así, se puede hacer seguimiento en tiempo real a personas que posean dispositivos Wifi, capturando patrones de movimiento de una muestra de la población.

Por otro lado, muchos de los sistemas de transporte público (STP) masivos del mundo utilizan una tarjeta inteligente o *smartcard* (SC) para el pago de ciertos medios de transporte (Bip! en Santiago, Oyster en Londres, Octopus en Hong Kong, entre otros). Cada validación de la SC, en general, posee información del lugar de la validación, instante en el que ésta ocurre y el identificador único de la tarjeta. En la mayoría de los sistemas, la validación se realiza solo al momento de ingresar a él, lo cual impide conocer directamente información sobre el viaje realizado (destino y ruta para llegar a destino, por ejemplo). La penetración de esta tecnología en los usuarios del STP puede llegar a ser completa, dependiendo de los niveles de evasión que se tenga, ya que se utiliza como sistema de pago.

En STP cada vez más complejos y congestionados, urge la necesidad de tener información de calidad, de manera rápida y capaz de captar la variabilidad de los sistemas. Esta información es necesaria tanto para la toma de decisiones operacionales del día a día como para decisiones estratégicas de largo plazo. Esta investigación se enfocará en grandes STP donde existan estaciones en que los pasajeros deban marcar su entrada al sistema, como por ejemplo *Bus Rapid Transit* (BRT), metro, tren o tranvía de alta capacidad. En grandes ciudades, estos STP son algunos de los principales modos de transporte dado que tienen alta capacidad y circulan en una red segregada del resto de los vehículos y por ende evitan su congestión.

---

<sup>1</sup> En este trabajo se usa la palabra ‘dispositivo’ para referirse a aparatos electrónicos del tipo *smartphone* que tienen capacidad Wifi.

El objetivo principal de esta investigación es usar de forma conjunta los patrones de viaje obtenidos con los WS y los conteos de las SC para estimar, en tiempo real, información relevante para la planificación y operación de un STP. Dentro de los objetivos específicos se encuentran:

- Proponer una metodología que pueda usar de forma conjunta los datos de los WS y las SC para estimar en tiempo real la matriz origen-destino (OD), carga por arco y otros indicadores relevantes de un STP.
- Investigar cómo varía la penetración de la tecnología Wifi en la población en el tiempo y en el espacio geográfico, y cuáles pueden ser las causas de esta variación y sus consecuencias en nuestra estimación.
- Estudiar el uso de modelos simples que usen fuentes de datos masivas y disponibles en tiempo real, en vez de modelos analíticos sofisticados, estáticos y difíciles de calibrar.

Además de este capítulo introductorio, esta tesis cuenta con otros cinco capítulos. En el segundo capítulo explican las variables de transporte a obtener y la tecnología que se usó para ello, junto con un análisis bibliográfico del estado del arte en cuando a la tecnología usada, otras similares y sus relaciones con el transporte. El tercer capítulo explica la metodología, sus variables y consideraciones más importantes. El cuarto capítulo describe el diseño de la prueba piloto realizada y la implementación computacional de la metodología. Los resultados y análisis obtenidos a partir de la prueba piloto se presentan en el quinto capítulo. Finalmente, el sexto capítulo discute las conclusiones principales de la investigación y las futuras aplicaciones.

## **2. CONTEXTO BIBLIOGRÁFICO Y TECNOLÓGICO.**

En esta sección se presenta la investigación previa y paralela al desarrollo de la metodología, tanto de la tecnología que se está utilizando como de otras tecnologías similares.

### **2.1. Situación actual.**

Para planificar y operar de buena manera cualquier STP, se necesita saber cuántos usuarios quieren ir de cada origen a cada destino de la red. Esto se conoce como la matriz OD. Actualmente, la principal forma de obtener la matriz OD es a través de encuestas realizadas a una muestra de usuarios en varios puntos y horarios de la red, para luego a través de modelos analíticos probabilísticos estimar la matriz OD más probable en el futuro. Otras metodologías han emergido a medida que la tecnología avanza, algunas de las cuales se discuten más adelante en este mismo capítulo.

Las encuestas en general son de alto costo, tienen baja penetración y requieren de estimación de modelos con muchos supuestos para la expansión de los datos. Esto genera que la encuesta se realice con poca frecuencia y que se tenga incertidumbre sobre la precisión de la información, que además es estática y se calcula solo para ciertos periodos de operación.

Otra pieza de información relevante para entender y planificar un STP es la carga por arco, la cual se entiende como la cantidad de usuarios que viaja de una estación<sup>2</sup> a otra estación vecina durante cierto periodo de tiempo. Este es un indicador relacionado directamente con el nivel de servicio del sistema que perciben los usuarios en cuanto a comodidad. Para una carga por arco dada, se pueden entregar diferentes niveles de frecuencias de vehículos. Mientras más baja sea la frecuencia, es mayor el hacinamiento que sufren los usuarios pero menor el costo de operación del sistema.

---

<sup>2</sup> “Estación” es la palabra que se utilizará en este trabajo para referirse al espacio físico donde los usuarios abordan o abandonan el vehículo.

Para encontrar una frecuencia que equilibre el hacinamiento que experimentan los usuarios y un buen uso del sistema, el primer paso es conocer la carga por arco. Hoy en día en general se realiza asignando los viajes de la matriz OD (obtenidos a partir de una encuesta) a la red y luego expandiendo esa muestra con diversos métodos. Existen casos en que, para calcular la carga por arco, el vehículo es pesado en algunos puntos de la red. Asumiendo un peso promedio de la población y sabiendo el peso del vehículo vacío se estima la cantidad de personas que hay dentro. Estos métodos requieren de muchos supuestos y modelos que son difíciles de calibrar, por lo que esta calibración es poco frecuente.

## **2.2. Descripción de la operación de un WS.**

Los dispositivos electrónicos que tienen su capacidad Wifi activada emiten señales constantemente al ambiente. Estas señales pueden ser de transmisión de datos, dado que se encuentran conectadas a un punto de acceso de Wifi, o también pueden ser señales de búsqueda de puntos de acceso de Wifi. Los dispositivos emiten regularmente estas señales de búsqueda cuando no están conectados a un punto de acceso para así conectarse a alguno que el usuario ya haya tenido guardado en su configuración. Estas señales son enviadas al ambiente y pueden ser captadas por aparatos tecnológicos capaces de esta tarea, como los WS usados en esta investigación.

El WS capta el identificador único e instante de detección de los dispositivos electrónicos que emitieron la señal dentro de una determinada área de detección, la cual es una característica del WS. El identificador único de cada dispositivo electrónico con capacidad de emitir señales Wifi es el código MAC (*Media Access Control*). Este código es asignado por la IEEE (*Institute of Electrical and Electronics Engineers*), donde la primera mitad del código representa a la empresa que está adquiriendo el código y la segunda mitad representa un indicador único para un dispositivo de esa empresa (Abedi *et al.*, 2013). Existen otras variables que pueden obtener los WS como la fuerza de la señal o el canal inalámbrico en que está operando la señal, pero para los efectos de esta investigación no son relevantes.

Es importante destacar que a pesar que en este trabajo se hable de dispositivos electrónicos con capacidad Wifi (que podrían incluir *smartphones*, *laptop*, *tablets*, entre otros), la investigación se enfoca principalmente en los *smartphones* dada su gran penetración en el mundo y porque los usuarios en general los llevan con ellos cuando viajan en transporte público.

La frecuencia con que se detecta un dispositivo no depende del WS sino que del dispositivo emisor de la señal. Dependiendo de la tecnología y cómo está programado cada dispositivo, la frecuencia con que las señales son enviadas varía, por lo que un dispositivo podría ser detectado varias veces por el mismo WS. Los dispositivos más nuevos generalmente envían señales con más frecuencia que los viejos, pero también depende del uso que se le esté dando al dispositivo (Musa y Eriksson, 2012). Un dispositivo que está siendo usado activamente por una persona envía más señales que el mismo dispositivo inactivo, sobre todo si es que se está utilizando alguna aplicación que requiera uso de datos móviles (Musa y Eriksson, 2012).

Si el dispositivo no envía señales mientras está en el área de detección entonces no será detectado. Por lo tanto, puede ocurrir que se tengan trayectorias incompletas que deben ser reconstruidas a partir de la información de otros WS. Más aún, el dispositivo podría haber enviado una señal pero ésta no llega al WS dado que el medio físico por el que viaja la señal no se lo permitió. En la Tabla 2-1 de Harwood (2009) se muestra cómo obstruyen diferentes materiales que se encuentran típicamente en ambientes de interior a las señales inalámbricas:

Tabla 2-1: Obstáculos que se encuentran en ambientes de interior que afectan las señales inalámbricas. Fuente: Harwood (2009)

<b>OBSTRUCCIÓN</b>	<b>SEVERIDAD</b>	<b>EJEMPLOS DE USO</b>
Madera	Baja	Paredes de interior o puertas.
Paneles	Baja	Paredes de interior.
Muebles	Baja	Sillones o muebles de oficina.
Vidrio transparente	Baja	Ventanas.
Vidrio opaco	Media	Ventanas.
Personas	Media	Áreas de alto tráfico de personas.
Cerámica	Media	Paredes.
Concreto	Media/Alta	Paredes de exterior.
Espejo	Alta	Espejos o vidrio reflectante.
Metal	Alta	Muebles de oficina o puertas.
Agua	Alta	Acuarios, lluvia, fuentes.

Para el caso del transporte público, las obstrucciones de concreto y metal son comunes debido a que en general la obra gruesa y los objetos dentro de las estaciones están hechos de concreto y metal. Además, los vehículos están compuestos principalmente de metal. En momentos de alta demanda de usuarios no solo aumenta la cantidad de señales en el ambiente sino que la cantidad de obstrucciones debido a personas circulando, que están compuestas principalmente por agua.

Debido a las razones anteriores, al trabajar con este tipo de tecnología se deben considerar las limitaciones anteriores. Las variables relacionadas directamente con la tecnología son un factor importante al momento de analizar posibles discrepancias entre la información estimada según esta metodología y la realidad.

Como se está trabajando con datos capaces de identificar el movimiento de un dispositivo electrónico, la privacidad es un tema a considerar. Inicialmente, no hay problemas de privacidad ya que los WS saben qué dispositivo está en qué lugar, pero no son capaces de asociar ese dispositivo a una persona natural. Si este dato se cruza con la información de quién es la persona natural que compró cada dispositivo móvil (dato que las compañías telefónicas poseen) sería posible, asumiendo que el dueño del teléfono es también su

usuario, hacer un seguimiento a personas naturales. Esto último sí atenta contra la privacidad de las personas, aunque en general los WS tienen un sistema de encriptación del código MAC que logra que este siga siendo único, pero que no sea posible asociarlo a un dispositivo electrónico.

Para una revisión sobre las características de las detecciones de señales Wifi, sus ventajas y desafíos, el lector puede revisar el trabajo de Abedi *et al.* (2013). Este trabajo resume de muy buena manera la investigación que ha llevado a cabo ese grupo de académicos. Parte de su trabajo está relacionado con detectar no solo la presencia sino que la ubicación física de un dispositivo Wifi de dos maneras posibles: la primera es utilizando la correlación que existe entre la distancia entre un dispositivo y el WS, y la fuerza de la señal enviada por el dispositivo; la segunda, es usando tríos de WS triangulados para estimar la zona en que se encuentra el dispositivo según cuáles WS lo detectan y cuáles no. Estas metodologías no serán usadas en este trabajo directamente, pero es importante entender que la tecnología permite estimar este tipo de variables para futuras aplicaciones.

Así como existen los WS, también existen los *Bluetooth Scanners* (BTS). Su operación es similar al del WS con la diferencia que detecta dispositivos con su capacidad *Bluetooth* activada, y por lo tanto, captura un identificador (número MAC) distinto al obtenido por un WS. La tecnología *Bluetooth* ha estado más enfocada en el uso de vehículos ya que muchos de ellos incluyen un sistema de *Bluetooth* interno, aunque también la mayoría de los *smartphones* tienen la capacidad de enviar señales *Bluetooth*. La tecnología *Bluetooth* es más antigua que Wifi y primero se desarrollaron investigaciones en transporte usando *Bluetooth*, pero enfocado principalmente en vehículos particulares y no en contextos peatonales.

### **2.2.1. WS versus BTS**

La tecnología Wifi comienza a aumentar de forma explosiva su uso en la población y se empieza a trabajar con ella para reemplazar el *Bluetooth* en temas, principalmente, de flujo de peatones. Abedi *et al.* (2013) compararon mediante experimentos reales usando detectores de Wifi y *Bluetooth* las diferencias entre las penetraciones de las tecnologías. Se

eligieron seis lugares peatonales y se detectó señales Wifi y *Bluetooth* simultáneamente durante aproximadamente dos horas. Luego de filtrar las detecciones de dispositivos repetidos, se compararon la cantidad de dispositivos únicos que emiten señal Wifi versus la cantidad de dispositivos únicos que emiten señal *Bluetooth*, y se tiene que en los seis casos más del 90% de los dispositivos únicos corresponde a Wifi. Los resultados se resumen en la Figura 2-1:

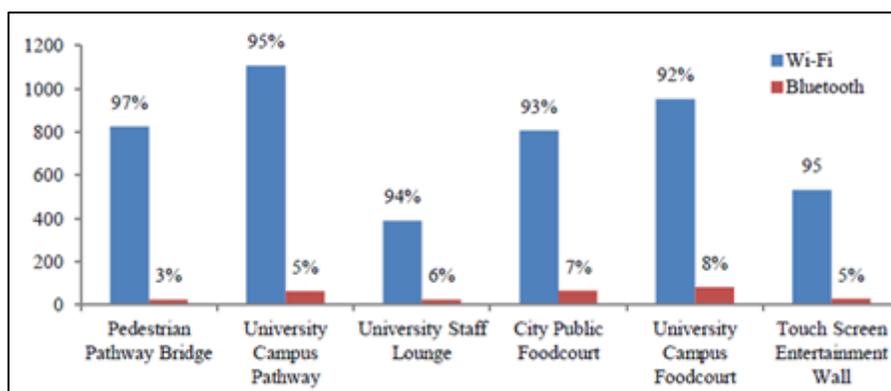


Figura 2-1 Comparación de popularidad de uso de Wifi y Bluetooth en lugares peatonales.

Fuente: Abedi, Bhaskar y Chung (2013).

Por otro lado, la investigación de Abedi *et al.* (2013) comprobó que las señales Wifi son detectadas en mucho menos tiempo que el *Bluetooth* debido a que la tecnología Wifi funciona escuchando señales, es decir, no tiene que comunicarse de vuelta con el dispositivo que se está detectando como sí debe hacerlo la tecnología *Bluetooth*. Ellos midieron experimentalmente el tiempo que un dispositivo se demoraba en ser descubierto por el scanner. El promedio de 1000 experimentos fue de 10,577 segundos para *Bluetooth* y 1,365 segundos para Wifi, es decir que un dispositivo promedio usando Wifi se demora 8 veces menos en ser detectado que el mismo dispositivo usando *Bluetooth*.

### **2.3. Uso de WS para análisis de tráfico peatonal y vehicular.**

Considerando una red de WS y la marca de tiempo que cada dispositivo deja en cada WS visitado, es posible inferir una trayectoria aproximada de cada uno de ellos. Asumiendo que cada dispositivo corresponde a una persona, distintos investigadores han utilizado WS para inferir información de movimiento de personas en distintos modos de transporte. Sin embargo, aún es un área poco desarrollada con pocos grupos de investigadores trabajando en estos temas. Varios de estos investigadores pertenecen a escuelas de Ingeniería Eléctrica o Ingeniería en Computación, por lo que el foco de sus investigaciones tiene más relación con cómo se producen las detecciones, las posibilidades de detectar a alguien o las interferencias que afectan las señales. A pesar de que esos resultados son valiosos, esta revisión bibliográfica se enfocará en los avances en cuanto a estimación de información de movimientos de personas en distintos modos de transporte.

Danalet *et al.* (2012, 2014) enfocaron su investigación en hacer un seguimiento de flujo de peatones en ambientes peatonales como el campus de una universidad. Desarrollaron una metodología para estimar la ruta seguida y el tiempo en cada destino por un peatón utilizando como input un mapa con la geografía y actividades de la red, y señales Wifi obtenidas de puntos de acceso Wifi múltiples.

Abedi *et al.* (2014) analizaron el flujo de personas de una oficina durante tres semanas, logrando entre otras cosas detectar movimientos grupales y la fidelidad de los individuos para mantenerse en el mismo grupo. Este análisis se podría realizar en contextos de transporte público para identificar cómo se comportan los usuarios al viajar en cuanto a qué tan regulares son sus viajes y si es que viajan en grupo regularmente o no.

En cuanto a otros modos de transporte no motorizado, la investigación de Abedi *et al.* (2015) estimó tiempos de viaje de personas que circulan caminando, corriendo o en bicicleta, logrando identificar qué dispositivo pertenece a qué grupo en una buena proporción de los casos. Se enfocaron en averiguar cuál es la importancia de la antena usada en los WS y cómo ésta afecta la estimación de velocidad de viaje. La capacidad de la antena debe ser lo suficientemente grande para captar a dispositivos que circulan

rápidamente por la zona de detección, pero no tan grande como para generar que las zonas de detección se sobrepongan y por ende se produzcan anomalías.

Las investigaciones que se enfocan en aplicar tecnología Wifi para la estimación de movimientos de usuarios de un STP masivo son escasas. Los trabajos están más enfocados en entender características de las señales, hacer un seguimiento de peatones dentro de un espacio físico y en el tiempo, estimar velocidades de traslado entre algunos puntos, entre otras. Este trabajo busca abrir una nueva línea de investigación donde se usen de forma conjunta la información de señales Wifi emitidas por los usuarios y la información de SC, para así obtener patrones de movimientos de usuarios de un STP masivo.

Otros investigadores han utilizado la tecnología Wifi para detectar patrones de movimiento de vehículos particulares. Musa y Eriksson (2012) desarrollaron modelos probabilísticos para estimar las rutas y la velocidad de vehículos particulares usando WS. Sus experimentos mostraron que la precisión de usar esta metodología depende en gran medida de la cantidad de WS que se tenga y de la ubicación física de los mismos. Los autores confirman que utilizar WS es una forma práctica, económicamente viable y precisa de recolectar información del sistema de transporte.

#### **2.4. Uso de otras tecnologías en transporte.**

En este subcapítulo se describen y analizan algunos de estudios relacionados al transporte que han utilizado tecnologías con fines similares a *Wifi*.

##### **2.4.1. Bluetooth.**

Por lo mencionado en la sección 2.2.1, el uso de BTS se enfoca principalmente en el tráfico de vehículos particulares. Esta tecnología ha aumentado su penetración en vehículos en los últimos años debido a que muchos aparatos digitales que están dentro del vehículo se conectan a través de BT (Michau *et al.*, 2013). Los análisis en vehículos particulares son variados: velocidades o tiempos de viaje (Abbott-Jard *et al.*, 2013), matrices origen destino (Barceló *et al.*, 2012, Michau *et al.*, 2013), entre otras.

El uso de BTS en STP es escaso. Kostakos (2008) utilizó esta tecnología para estimar los patrones de viajes de usuarios que subían a un bus equipado con BTS. El muestreo de viajes logrado fue bajo dado que la penetración de *Bluetooth* en usuarios del transporte público era baja hace casi una década. Además, el muestreo estaba sesgado al comportamiento de viaje de las personas que tienen un dispositivo con conexión *Bluetooth* activada, pero es una primera aproximación al uso de tecnologías inalámbricas para estimar flujos de personas en transporte público.

#### **2.4.2. Señal de celulares.**

Las señales que emiten los celulares al llamar o enviar un mensaje de texto pueden ser trianguladas con bastante precisión dentro de la ciudad. Varios autores han aprovechado esta fuente de información para estudiar los patrones de movilidad de una ciudad (Caceres *et al.*, 2007, Calabrese *et al.*, 2013, Alexander *et al.*, 2015, entre otros).

El reciente estudio de Alexander *et al.* (2015), al igual que este trabajo, tiene como uno de sus objetivos obtener la matriz OD de una red de transporte. Ellos realizaron un experimento en Boston durante 60 días con datos reales de las llamadas de 2 millones de usuarios. Las empresas telefónicas triangularon las llamadas utilizando antenas de señal de celular para encontrar la ubicación de cada celular que hizo una llamada con una precisión de entre 200 a 300 metros. Luego de filtrar y aplicar un algoritmo probabilístico, los investigadores fueron capaces de obtener diariamente y por periodo del día los viajes realizados en la ciudad, segregando por propósito de viaje. El propósito del viaje lo obtuvieron utilizando la información de varios días del mismo usuario para poder definir, bajo algunos supuestos, cuál es su lugar de trabajo y cuál es su hogar.

La hora exacta en que una persona viaja se obtiene como una probabilidad que depende de las horas de su última llamada en el primer lugar y la primera llamada del segundo lugar. Entonces, no es posible estimar viajes en tiempo real ni de forma dinámica ya que se necesita estudiar los patrones de cada usuario a través del tiempo.

A pesar de lo anterior, esta metodología arroja buenos resultados para viajes promedio diarios en una ciudad. La ventaja de utilizar datos de llamadas es que el campo de detección de la metodología es prácticamente toda la ciudad sin tener que construir nueva infraestructura, a diferencia de los WS que son necesarios instalarlos en lugares estratégicos para obtener datos. Para una planificación a nivel de ciudad es razonable utilizar esta información, pero cuando se busca analizar el funcionamiento de un sistema de transporte en detalle, se necesita mejor información de cuándo y dónde ingresan y salen los usuarios del sistema.

### **2.4.3. Smartcard.**

El uso de la SC para obtener información operacional de un sistema de transporte se ha utilizado desde hace unos años. En redes de transporte público donde no solo se debe validar el ingreso al sistema con la SC sino que también el egreso, como en Londres o Tokyo, es posible encontrar patrones Origen Destino directamente. Por otro lado, en sistemas como Santiago, Paris, Buenos Aires o Nueva York los usuarios deben validar su entrada al sistema pero no su salida, por lo que la estimación de la matriz OD no es trivial ya que no se conoce dónde ni cuándo termina el viaje de cada usuario.

Varios investigadores han desarrollado metodologías para utilizar la información obtenida con las SC. Para una revisión bibliográfica de cómo se fueron desarrollando estas metodologías y cuáles han sido los temas tratados por varios investigadores, el lector puede revisar el trabajo de Pelletier *et al.* (2011).

Para el caso chileno, Munizaga y Palma (2012) construyeron la matriz OD del STP Transantiago (que incluye metro y buses) utilizando la información de la SC *Bip!*. Los destinos de cada viaje fueron estimados en base a dónde se realizó la siguiente validación de la SC bajo supuestos que dependen de la hora del día en que viaje y el tiempo entre validaciones, entre otros. Su metodología fue capaz de asignar un destino al 80% de los viajes totales. Esta metodología fue validada más adelante por Munizaga *et al.* (2014). La validación se realizó con tres fuentes de información: los mismos datos para desarrollar el modelo, la encuesta OD del sistema de metro y con un grupo de voluntarios que entregaron

información detallada de sus viajes. Los resultados muestran que la metodología fue capaz de estimar correctamente el paradero de origen en un 98.9% de los casos y el paradero de destino en un 84.2% de los casos.

A pesar de los buenos resultados que se obtienen al trabajar con información de las SC, tiene dos limitaciones importantes intrínsecas en la tecnología. La primera es que no se puede saber qué ruta elige un usuario que tiene la posibilidad de elegir múltiples rutas entre su origen y su destino. Para una red pequeña y simple este problema no es significativo, pero a medida que aumenta el número de estaciones y la red se complejiza, aparecen naturalmente múltiples rutas entre pares de estaciones. Esta limitación no afecta la estimación de la matriz OD, pero sí la de la carga por arco ya que no se puede observar qué arcos recorre un usuario entre su origen y su destino.

La segunda limitación de las SC tiene que ver con el tiempo de respuesta de los datos y con la dependencia de información futura. Muchos de los viajes que se realizan en STP son viajes de *commuting* que se pueden estimar fácilmente con la información de SC ya que el usuario tiene un comportamiento estable. Pero, cuando ocurre alguna contingencia en la ciudad que afecte el sistema de transporte (cierres de una estación de metro, un paro de buses, un choque vehicular, una marcha, lluvia, un corte de luz, etcétera) los usuarios del sistema no tienen un comportamiento estable ni predecible, por lo que estimar los destinos mediante la SC pierde robustez.

Por último, al utilizar las SC de transporte público se debe considerar una fuente de error adicional: la evasión. Además, hay usuarios que podrían estar evadiendo solo en algunas etapas del viaje, haciendo que la estimación del destino no solo sea imposible de llevar a cabo sino que podría ser incorrecta al asumir que el destino del usuario fue donde validó nuevamente con su tarjeta.

### 3. METODOLOGÍA.

Uno de los objetivos de este trabajo consiste en el desarrollo de una metodología que permita generar información relativa a la demanda de un STP a partir de los datos recolectados por WS y las SC. A fin de evaluar su pertinencia y potencialidad, esta metodología será implementada mediante la realización de una prueba piloto. En este capítulo se presenta la metodología desarrollada, mientras que en el capítulo siguiente se describe la implementación de la prueba piloto.

La metodología busca cumplir cuatro objetivos simultáneamente, a saber:

- Estimar la matriz OD de un STP.
- Estimar la carga por arco de un STP.
- Obtener información sobre cómo cambia la penetración en el tiempo y el espacio para estudiar el sesgo que pudiera existir al trabajar con datos de WS.
- Obtener datos de forma ordenada para luego estimar información relevante para un STP: tiempos de espera, viajes diarios por persona y saber qué sucede en días de contingencias.

La metodología está enfocada en STP que tengan las siguientes características:

- Todos los usuarios validan su entrada al sistema mediante una SC (*tap in*), con lo cual se almacena el instante y el lugar de esa validación.
- Cada parada o estación cuenta con un WS, cuyo alcance es suficiente para cubrir todo el espacio donde los usuarios esperan antes de subirse al vehículo y el espacio donde el vehículo se detiene a dejar y recoger usuarios.

Dado que hay STP que operan cerca de otros sistemas de transporte como vehículos o peatones, un WS detectará dispositivos que no necesariamente corresponden a usuarios del sistema. La metodología asume que el STP está aislado de otros medios de transporte y por ende todas las señales que son captadas provienen de usuarios que están dentro del sistema.

Para aplicaciones reales en que esto no ocurra, se deben filtrar las detecciones que provengan de otros modos de transporte.

Como los WS solo están tomando una muestra de los usuarios, la metodología incluye un método de expansión de la muestra basado en ponderadores. Estos ponderadores buscan transformar los dispositivos en usuarios reales del STP. La hipótesis es que la muestra es insesgada y que representa una proporción del total de usuarios que varía en el tiempo y en el espacio. Por lo tanto, cada ponderador depende del tiempo y del espacio. Es decir, hay un ponderador para cada periodo de tiempo (se entiende por un periodo de tiempo a diferentes horarios durante el día o pequeños lapsos de cierta cantidad de minutos) y en cada una de las paradas o estaciones de la red, que puede o no ser diferente al de otros periodos de tiempo u otras estaciones.

La función del ponderador es transformar a un dispositivo emisor de señal Wifi que comenzó su viaje durante cierto periodo de tiempo en cierta estación de la red, en una cantidad de usuarios equivalente. En otras palabras, un dispositivo que se mueve por la red será igual a la cantidad de usuarios equivalentes moviéndose que indique su ponderador correspondiente. Así, los resultados que se entregan en la metodología están en unidad de cantidad de usuarios y no de cantidad de dispositivos Wifi.

Además, el ponderador, en conjunto con información geográfica de la ciudad e información socioeconómica de sus habitantes, ayuda a estudiar sesgos relacionados con el origen del viaje de los usuarios que existen al trabajar con tecnología Wifi. Según estudios de Pew Research Center (2016), existen diferencias en la posesión de *smartphones*<sup>3</sup> en al menos tres ámbitos diferentes en el mundo: edad, nivel de educación e ingreso. En el caso de Chile, la penetración total de *smartphones* es de 65%. Es decir, que 65 de cada 100 personas posee *smartphone*. Sin embargo, se tienen las siguientes observaciones:

- La penetración de *smartphone* en personas de entre 18 a 34 años es 86%, mientras que en personas mayores a 35 años es 50%.

---

<sup>3</sup> Como se mencionó anteriormente, este trabajo es apto para todos los dispositivos con capacidad Wifi presentes en la red de transporte, pero la penetración que tienen los *smartphones* en la población hace que naturalmente cualquier implementación de la metodología se enfoque en ellos.

- Solo el 16% de las personas con bajo nivel de educación tienen *smartphones*, en contraste con el 74% de las personas con alto nivel de educación.
- El 46% de las personas con bajo ingreso posee *smartphone*, pero el 79% de las personas con alto ingreso sí tiene *smartphone*.

La tabla de dónde provienen estos datos y los de otros 39 países se puede encontrar en el Anexo A.

Por lo tanto, entendiendo que la distribución espacial de personas de diferente edad, nivel de educación e ingreso no es homogénea en una ciudad, los ponderadores cambian dinámicamente y se ajustan a las necesidades de cada estación en cada momento del día. Por ejemplo, una persona con alto ingreso puede viajar a una zona de bajo ingreso en la mañana y volver a su casa en la tarde, por lo que el viaje de la tarde se origina en una zona de bajo ingreso y su ponderador debe ajustarse a esa situación.

Además, la tendencia a que mientras menos edad se tenga existe una mayor probabilidad de tener *smartphone* (hasta cierto nivel mínimo de edad) es un argumento para pensar que los estudiantes escolares y universitarios tienen mayor penetración que las personas adultas. Los estudiantes tienen un comportamiento de viaje particular y distinto al resto de la población. El ponderador ayuda a controlar este comportamiento en los lugares que se originan los viajes de estudiantes, pero no controlan la forma en que estos eligen sus destinos, por lo que hay un sesgo involucrado.

Para intentar reducir el sesgo causado por trabajar con tecnología Wifi el primer paso es visualizar cómo se comporta la penetración. Los ponderadores, al ser el inverso de la penetración, entregan esta información con un nivel de detalle suficiente para observar cambios tanto en el tiempo y en espacio, y con esa información en conjunto con datos externos a la metodología, se puede estudiar posibles causas del sesgo.

### **3.1. Presentación de las variables.**

El modelo se construye en base a dos fuentes de datos (WS y SC) y es capaz de estimar variables de transporte en base a esas dos fuentes. El análisis se realiza por periodos de

tiempo del mismo largo (Ejemplo: 8:00 a 8:15), denotados por el subíndice  $p$  y en instantes de tiempo (Ejemplo: 8:03:56) denotados por el subíndice  $t$ . A continuación, se presentan las variables utilizadas en la metodología para obtener la matriz OD y la carga por arco. La sigla con que cada variable fue nombrada proviene de su nombre en inglés en la mayoría de los casos.

$T_{pij}$ : (*Trips*) Cantidad de usuarios que comienzan su viaje en el periodo  $p$  y viajan entre la estación  $i$  y  $j$ .

$D_{pij}$ : (*Devices*) Cantidad de dispositivos que comienzan su viaje en el periodo  $p$  y viajan entre la estación  $i$  y  $j$ .

$W_{pi}$ : (*Weighting factor*) Ponderador de los dispositivos que comienzan su viaje en el periodo  $p$  en la estación  $i$ .

$ASC_{pi}$ : (*Arrival of users obtained using Smart Card data*) Cantidad de usuarios que entraron a la red durante el periodo  $p$  por la estación  $i$  obtenida a partir de las SC. Es decir, la afluencia a la estación  $i$  durante periodo  $p$ .

$AWS_{pi}$ : (*Arrival of devices obtained using Wifi Scanners data*) Cantidad de dispositivos que entraron a la red durante el periodo  $p$  por la estación  $i$  obtenida a partir de los WS.

$vi_{nti}$ : (*Viaje Iniciado*) Variable binaria que se genera y toma valor 1 cuando el dispositivo  $n$  comienza un viaje en el instante  $t$  en la estación  $i$ .

$vf_{nt1t2j}$ : (*Viaje Finalizado*) Variable binaria que se genera y toma valor 1 si el dispositivo  $n$  comienza un viaje en el instante  $t1$  en la estación  $i$  y lo termina en el instante  $t2$  en la estación  $j$ .

$d_{nti}$ : (*detection*) Variable binaria que toma valor 1 si hubo una detección del dispositivo  $n$  en el instante  $t$  en la estación  $i$ , y 0 en otro caso.

$sL_{npj}$ : (*subLoad*) Variable binaria que toma valor 1 si el dispositivo  $n$  cruza el arco que une la estación  $i$  con la estación  $j$  en el instante  $t$ . Las estaciones  $j$  son las estaciones que están conectadas a través de un solo tramo con la estación  $i$  (es decir,  $j=i-1$  o  $j=i+1$ ).

$L_{pij}$ : (*Load*) Cantidad de usuarios que cruzan el arco que une la estación  $i$  con la estación  $j$  durante el periodo  $p$ . Las estaciones  $j$  son las estaciones que están conectadas a través de un solo tramo con la estación  $i$  (es decir,  $j=i-1$  o  $j=i+1$ ).

La razón por la que algunas variables están definidas en función de periodos de tiempo  $p$  y otras en instantes de tiempo  $t$  es por el nivel de detalle con que ingresan los datos originales de los WS y los usos que se le puede dar a esa información. Por ejemplo, un dispositivo  $n$  podría ser detectado en varias estaciones durante el periodo  $p$ , por lo que análisis posteriores de los datos no sabrían en qué sentido se movió el dispositivo  $n$  o cuál de todas las estaciones en las que fue detectado fue su estación de origen. Por otro lado, cuando se quieren estimar resultados valiosos para planificadores y operadores STP, la información tiene que tener cierto nivel de agregación para que sea manejable y es por eso que los resultados finales están a nivel de periodo  $p$ .

### **3.2. Procesos que componen la metodología para estimar matriz OD y carga por arco.**

La metodología está compuesta de dos grandes etapas: primero, el manejo de los viajes de dispositivos Wifi a través de la red que se van iniciando, actualizando y finalizando con la información que entregan los WS. Luego, esta muestra de viajes es expandida a usuarios reales y transformada en información útil utilizando datos de la SC y de los WS.

#### **3.2.1. Inicialización de un viaje.**

En esta sección se describe cómo se inician los viajes de los dispositivos y cuándo considerar que una detección es efectivamente un viaje. El proceso se basa en almacenar y procesar cada detección  $d_{nti}$  como el identificador único del dispositivo Wifi o código MAC ( $n$ ), el instante en que ocurre (fecha y hora,  $t$ ) y la estación en que se realizó la detección ( $i$ ). Cada una de estas detecciones  $d_{nti}$  se procesa y genera uno, y solo uno, de los dos siguientes resultados según sea el caso:

- El dispositivo  $n$  ya ha sido identificado previamente por este u otro WS durante el periodo de análisis. En este caso, se actualiza el tiempo y la estación de destino, en

caso de que su última detección haya sido en otra estación. En la siguiente sección se detalla qué ocurre cuando se actualiza el viaje.

- El dispositivo  $n$  no ha sido identificado dentro de la red en el periodo de análisis. En este caso, este dispositivo es candidato a iniciar un nuevo viaje con instante de origen  $t$  y estación de origen  $i$ . La forma de considerar un viaje como efectivamente iniciado se describe a continuación.

Cuando se detecta por primera vez un código MAC no es directo asumir que es un usuario del STP que lleva un dispositivo con su Wifi activado. Podría ser un dispositivo fijo, un dispositivo de alguien que trabaja en la estación, un dispositivo de alguien que pasó por la estación pero sin entrar a la red, o incluso podrían ser señales de dispositivos que están fuera de la estación que se filtraron hacia dentro. El problema es que no se puede diferenciar fácilmente si alguna detección proviene de alguna de estas fuentes, pero sí se puede diferenciar a los dispositivos que efectivamente están viajando en la red.

El filtro consiste en no asumir que un nuevo dispositivo detectado es un nuevo dispositivo viajando hasta que efectivamente haya viajado. Es otras palabras, un dispositivo será cargado como un viaje que entró a cierta estación  $i$  en cierto instante de tiempo  $t$  solo cuando sea visto nuevamente en una estación  $j$  diferente a su estación de origen  $i$  en un instante de tiempo  $t'$  mayor a  $t$ . Esto elimina las fuentes de error mencionadas anteriormente excepto casos muy particulares. La variable que dicta este proceso se define de la siguiente forma:

$$vi_{nti} = \begin{cases} 1, & \text{si } d_{nti} = 1 \text{ y } d_{nt'j} = 1, \text{ tal que } i \neq j, t' > t \\ 0, & \text{e. o. c.} \end{cases}$$

Recordar que el dispositivo  $n$  no ha sido identificado anteriormente en la red, por lo que  $d_{nti}$  es su primera detección. Así, las afluencias de usuarios en cada periodo  $p$  por cada estación  $i$  estimado utilizando los WS se obtiene de la siguiente forma:

$$AWS_{pi} = \sum_n \sum_{t \in p} vi_{nti} \quad \forall i, p$$

Existe una simplificación de la formulación matemática para dispositivos que hacen más de un viaje durante el periodo de análisis. Cuando el mismo dispositivo  $n$  vuelva a entrar a la red su código MAC único será el mismo que en el viaje anterior. Para no generar un nuevo subíndice que indique el viaje diario de cada dispositivo se asume que, cuando un dispositivo  $n$  vuelve a ser detectado en la red luego de terminado su viaje, en cada nueva detección se agrega un indicador al final del código MAC para que pueda ser diferenciado consigo mismo entre viajes diferentes.

### 3.2.2. Actualización de un viaje.

Cuando ingresa una nueva detección  $d_{nt'j}$  y existe una detección  $d_{nti} = 1$  tal que las estaciones  $i$  y  $j$  estén conectadas directamente por un arco (es decir,  $j = i - 1$  o  $j = i + 1$ ) y  $t' > t$ , entonces significa que el dispositivo  $n$  cruzó el arco que une la estación  $i$  con la estación  $j$  en el instante  $t'$  (se asume como instante de cruce el instante en el que el dispositivo  $n$  es detectado en la estación final del arco). Para esta parte de la metodología, se asume que todo dispositivo que entra al área de detección de un WS envía al menos una detección.

A través de estos pares de detecciones se van identificando los arcos que cruza cada dispositivo y se almacenan en la siguiente variable  $sL_{npj}$ . Esta variable binaria toma valor 1 si el dispositivo  $n$  cruza el arco que une la estación  $i$  y la estación  $j$  en el instante  $t \in p$ . En esta parte de la metodología la información se comienza a agregar en periodos y no instantes de tiempo, ya que los resultados finales se trabajan a nivel de periodo. La variable se define de la siguiente forma:

$$sL_{npj} = \begin{cases} 1, & \text{si } d_{nti} = 1 \text{ y } d_{nt'j} = 1, \text{ tal que } j = i + 1 \text{ o } j = i - 1, t' > t, t' \in p \\ 0, & \text{e. o. c.} \end{cases}$$

Ahora, no es posible calcular de forma directa la carga completa de un arco porque los dispositivos que cruzan cierto arco que va de la estación  $i$  a la estación  $j$  en cierto periodo  $p$  pueden tener distinto ponderador ya que pueden haber entrado al STP en distintas estaciones y en distintos periodos de tiempo. La expansión de dispositivos a usuarios se debe hacer de manera desagregada. Este proceso se explica más adelante.

### 3.2.3. Finalización de viajes.

En esta sección se explica cómo y cuándo un viaje que está siendo realizado por algún dispositivo pasa a considerarse como que ha finalizado. A diferencia de la carga por arco, la matriz OD se construye con los viajes que ya han sido finalizados, por lo que la manera en que se va a considerar que un viaje finalizó afecta directamente sus resultados.

El criterio para considerar un viaje como finalizado es simple: si un dispositivo  $n$  no ha sido visto por ningún WS por más de cierta cantidad de tiempo  $L$ , entonces se asume que el dispositivo  $n$  salió de la red en el instante  $t$  por la estación  $j$  que viene dado por su última detección  $d_{ntj}$ . El valor del parámetro  $L$  debe ser suficientemente grande para no finalizar viajes que todavía se encuentran efectivamente viajando pero que no han podido ser detectados por algún WS. Por otro lado,  $L$  debe ser lo suficientemente pequeño para identificar que un dispositivo puede salir de la red, hacer una actividad y luego volver a la red para viajar nuevamente a otro lugar sin considerar que todo se hace en un único viaje.

Los viajes terminados se almacenan en la variable binaria  $vf_{nt1it2j}$  que toma valor 1 cuando el dispositivo  $n$  inicia un viaje en el instante  $t1$  en la estación  $i$  y lo finaliza en el instante  $t2$  en la estación  $j$ . La variable está definida según dos restricciones: la primera, para definir que un viaje finalizado que se inicia en la estación  $i$  en el instante  $t1$  tiene que haber efectivamente iniciado su viaje en esa estación en ese instante según la variable  $vi$ . La segunda, para definir que un viaje finalizado que termina en la estación  $j$  en el instante  $t2$  tenga la última detección de ese dispositivo en esa estación en ese instante.

$$vf_{nt1it2j} \leq vi_{nt1i}$$

$$vf_{nt1it2j} \leq d_{nt2j}, \text{ con } t2 = \max(t) \forall n, j$$

A medida que los viajes van finalizando se va construyendo la matriz OD de viajes  $D_{pij}$ , es decir, de los viajes que hicieron los dispositivos Wifi. Esta matriz OD puede ser útil para planificadores y operadores ya que es directo sacar proporciones de viajes de dispositivos Wifi por ejemplo desde un origen  $i$  hacia el resto de la red o hacia un destino  $j$  desde el resto de la red. La ecuación para estimar la matriz OD de dispositivos Wifi es la siguiente:

$$D_{pij} = \sum_n v f_{nt1it2j} \quad \text{tal que } t1 \in p \forall p, i, j$$

Cabe destacar que como un viaje se incluye en la matriz OD cuando termina, los viajes más cortos estarán incluidos antes en la matriz OD que los viajes largos ya que terminan antes. Para un planificador que le interesa información de largo plazo esto deja de tener importancia. Pero para la toma de decisiones operacionales en tiempo real podría ser relevante ya que se estarán sobrestimando los viajes cortos y subestimando los viajes largos si se trabaja con la matriz OD de cierto periodo  $p$  donde no todos los viajes que se iniciaron antes o durante ese periodo han terminado. Esta es una limitación de la metodología producto de trabajar con datos revelados y no declarados ni estimaciones futuras.

La finalización (e inicialización de un viaje) se puede ver afectada si es que un usuario apaga o enciende la capacidad Wifi de su dispositivo mientras está viajando. No es posible corroborar que un usuario se baja en cierta estación y por ende no vuelve a ser visto por un WS, o que simplemente apagó la capacidad de Wifi de su *smartphone* o apagó su *smartphone* completamente en dicha estación pero siguió viajando por la red. La matriz OD y la carga se distorsionarían si esto ocurre. Se asume que la cantidad de usuarios con este comportamiento es despreciable.

#### **3.2.4. Estimación de los ponderadores.**

En esta etapa se genera la fusión de datos entre las detecciones Wifi obtenidas por los WS y las afluencias de usuarios a las estaciones obtenidas a través de la SC. Los WS entregan información detallada de los flujos y movimientos de todos los dispositivos que fueron detectados y las SC entregan información detallada de la cantidad de usuarios reales (asumiendo evasión nula) que están entrando al sistema.

El objetivo de los ponderadores, como se menciona al principio de este capítulo, es expandir la muestra de dispositivos obtenidos con los WS y permitir estudiar algunos de los sesgos socioeconómicos que se producen al trabajar con tecnología Wifi que están asociados a la estación y periodo de origen del viaje. El ponderador busca convertir a un dispositivo Wifi en una cantidad de usuarios equivalente para que el resultado de la

metodología esté en unidades de usuarios y no de dispositivos, lo que es mucho más atractivo para un planificador u operador del STP.

El ponderador se estima usando la ecuación que se muestra a continuación:

$$W_{pi} = \frac{ASC_{pi}}{AWS_{pi}} \quad \forall p, i$$

A modo de ejemplo, si durante el periodo  $p$  en la estación  $i$  se iniciaron 50 viajes de dispositivos Wifi que fueron detectados por los WS, y en el mismo periodo  $p$  y estación  $i$  hubo 100 validaciones de SC, entonces el ponderador para el periodo  $p$  y la estación  $i$  será  $100/50 = 2$ . Cualquier viaje iniciado que haya sido iniciado en  $t \in p$  y en  $i$  (es decir, todo  $n$  tal que  $vi_{nti} = 1$ ) tendrá un ponderador correspondiente a 2. Eso quiere decir que cuando ese dispositivo  $n$  cruce un arco, la carga de ese arco aumentará en 2 y no en 1. Asimismo, los viajes del par origen destino del viaje que realizó el dispositivo  $n$  aumentarán en 2 y no en 1.

El valor de los distintos  $W_{pi}$  es un dato relevante para estudiar el sesgo que puede existir al planificar un STP usando Wifi. De la forma en que el ponderador está construido, es el inverso de la penetración observada en el lugar de origen del viaje (entendida como porcentaje de usuarios que viajan por la red con un dispositivo Wifi del total de usuarios que viajan por la red). El análisis de la magnitud de la penetración en diferentes horarios, días o lugares en conjunto con otras fuentes de datos tanto de los WS y SC como fuentes externas, puede entregar luces sobre el sesgo de utilizar tecnología Wifi.

Para la carga por arco, la cantidad total de usuarios que cruza durante el periodo  $p$  el arco que une directamente la estación  $i$  con la estación  $j$  viene dada por la siguiente ecuación. Dentro de la sumatoria, el término de la izquierda representa si el dispositivo  $n$  cruzó el arco que va de  $i$  a  $j$  en  $p$ . El de la derecha representa el ponderador del dispositivo  $n$ .

$$L_{pij} = \sum_n sL_{npij} \cdot W_{p'k} \quad \text{tal que } vi_{ntk} = 1, t \in p' \forall p, p' \leq p, i, j = i \pm 1$$

Para la matriz OD, la cantidad total de usuarios que inicia su viaje durante el periodo  $p$  en la estación  $i$  y finaliza su viaje en algún periodo en la estación  $j$  viene dada por la siguiente ecuación:

$$T_{pij} = D_{pij} * W_{pi}$$

Estos dos cálculos finales cierran el proceso de la metodología y logran fusionar los datos provenientes de los WS ( $sL_{npij}$  y  $D_{pij}$ ) y los datos que se derivan de la información de las SC ( $W_{pi}$ ) para la estimación de la matriz OD y la carga por arco.

El dato de las afluencias de las SC,  $ASC_{pi}$ , está disponible inmediatamente una vez que finaliza el periodo  $p$  (nadie puede entrar a la estación  $i$  en el periodo  $p$  después de que termine el periodo  $p$ ). Pero, el dato de las afluencias de los WS,  $AWS_{pi}$ , tiene un desfase temporal ya que al terminar el periodo  $p$ , todavía hay dispositivos que pueden ser considerados como que entraron en el periodo  $p$  a la estación  $i$ . Esto debido a que un dispositivo no se considera dentro de  $AWS_{pi}$  hasta que fue detectado en otra estación diferente a  $i$ , por lo que pasa un tiempo entre que el dispositivo es detectado por primera vez hasta que efectivamente se considera como que entró al sistema.

Se asume que ningún usuario circula con dos o más dispositivos con su capacidad Wifi activada. Existen usuarios que podrían tener un *tablet* con Wifi activado y además un *smartphone*, lo que sería un problema ya que se estaría considerando que su viaje es en realidad realizado por dos personas diferentes, sobreestimando ese viaje en particular y alterando la estimación del ponderador. El efecto de este supuesto es difícil de cuantificar si no es preguntado directamente a los usuarios.

### **3.3. Metodología para la estimación de otros resultados.**

En esta sección se muestra la metodología seguida para obtener tres resultados posteriores a la obtención de la matriz OD y carga por arco: tiempos de permanencia en estación, cantidad de viajes diarios por dispositivo y el comportamiento en días especiales.

### **3.3.1. Tiempos de permanencia en estación.**

Se puede estimar el tiempo de permanencia que experimentan los usuarios dentro de las estaciones al momento de entrar al STP. Se habla de tiempo de permanencia en vez de tiempo de espera ya que es difícil llegar a un acuerdo sobre qué se considera como tiempo de espera (desde que entra a la estación, desde que valida su entrada, hasta que se sube al vehículo, hasta que el vehículo sale de la estación, entre otras).

Para estimar el tiempo de permanencia se utiliza la primera y última detección de cada dispositivo Wifi en la estación de origen de su viaje para estimar el tiempo que estuvo este dispositivo dentro del área de detección del WS. Luego, se promedia el tiempo de espera para todos los usuarios que estaban en cierto periodo en cierta estación para obtener un tiempo de permanencia promedio por periodo por estación. Notar que con la colección de tiempos de permanencia individuales también es posible estimar una distribución esta variable y obtener así otros indicadores de interés (rango, percentiles, etc.)

Cualquier dispositivo que su primera detección fuese en el mismo instante que la última no fue considerado en el análisis. Aun así, esta forma de estimar el tiempo de permanencia subestima al real ya que las detecciones enviadas por los dispositivos siempre estarán completamente contenidas dentro del tiempo de permanencia.

También se deben filtrar los tiempos de permanencia muy altos ya que probablemente estos corresponden a dispositivos fijos o a usuarios que están usando el STP de forma irregular como por ejemplo esperando a alguien antes de viajar.

### **3.3.2. Viajes diarios por dispositivo.**

Al momento de planificar cuánta información sobre el STP se entrega a los usuarios, una pregunta clave es cuántos viajes hace cada usuario. No es necesario proveer el mismo nivel de información de la red del STP a un usuario que hace todos los días el mismo viaje de su casa al trabajo y luego regresa, que a un turista que solo estará en la ciudad por algunos días y por ende no tiene conocimiento de la red previo a realizar su viaje.

Con las SC es posible conocer, fácilmente, cuántas veces una misma SC es utilizada en un día o en la semana, ya que la SC tiene un ID único. Esta información no necesariamente

representa la cantidad de veces que un mismo usuario utiliza el STP, ya que hay personas que usan más de una SC y hay SC usadas por más de una persona.

Utilizando los WS se puede hacer el mismo análisis usando el código MAC, aunque también se tendrá un error por personas que viajan con más de un dispositivo. Se puede tomar los códigos MAC únicos detectados cada día y luego hacer una comparación entre días para identificar qué usuarios hacen viajes regulares o de *commuting* y cuáles hacen viajes esporádicos. Además, se puede saber dónde se están realizando estos viajes y así planificar el nivel de información a entregar en esas estaciones.

El mismo análisis se puede hacer a nivel de día asumiendo que usuarios que viajan solo una vez al día son usuarios que requieren más información que otros que hacen el mismo viaje de ida y vuelta el mismo día.

### **3.3.3. Días especiales.**

Las tecnologías utilizadas son capaces de entregar información sobre el sistema de transporte en días en que el comportamiento de los usuarios no fue habitual. Este comportamiento puede tener múltiples causas: accidentes, lluvias, evento deportivo o recreativo masivo, protestas masivas en la ciudad, días previos a feriados, entre otras.

Con las SC es posible determinar cómo se modifica la entrada de usuarios a estaciones, pero la estimación de destinos es poco confiable ya que la siguiente validación puede también estar influenciada por el efecto que transformó el comportamiento dentro de la red del STP en primer lugar. Con los WS, se puede determinar los destinos de los usuarios de la misma forma que si fuese un día convencional.

La metodología no se modifica, pero sí es importante entender que la penetración Wifi podría ser diferente en caso que ocurra cierto evento. Por ejemplo, si ocurre un accidente que haga que la red de retrase, muchos usuarios tomarán su *smartphone* para comunicar a quien corresponda que llegarán atrasados a su destino. Como se mencionó en el capítulo 2, usar activamente el *smartphone* hace que este envíe más señales y por lo tanto sea más probable que sea detectado.

#### 4. IMPLEMENTACIÓN Y PRUEBA PILOTO.

La prueba piloto consiste en la aplicación de la metodología a un caso de estudio real. Para esto, se instaló un WS en cada una de cuatro estaciones consecutivas de un STP. El WS funciona como el que se ha descrito hasta ahora en este trabajo: escucha las señales Wifi enviadas por los dispositivos que se encuentran en su área de detección (cuyo rango depende de la antena que se instale en el WS) y almacena el identificador único (código MAC) y el instante de la detección.

El WS tiene la capacidad de encriptar el código MAC real del dispositivo detectado para proteger la privacidad de los usuarios. Por lo tanto, en ningún momento se observó el código MAC de algún usuario. Más aún, dado que la encriptación de cada código MAC cambia día a día, no fue posible hacer un estudio longitudinal de cómo se movían ciertos individuos durante varios días.

Cada dato de detección incluye el campo de estampilla de tiempo (*timestamp*), WS que realiza la detección y el identificador único encriptado del dispositivo. En la Figura 4-1 se puede ver un ejemplo de los datos.

<b>Timestamp</b>	<b>WS</b>	<b>Mac encriptado</b>
Dia2 0:00:00	E3	d2e283b923fa4c6cc444b1cd448a6d608c4841c6d399b0bd5424b96ddfed147c
Dia2 0:00:00	E1	489c2ef9ec70b04dd9f053ae5c77aeb5b24cc337bcb69d3a030d2bde54486615
Dia2 0:00:01	E2	b5247a8b16dd6be9c34bd5dfbacfa5ff15f7964c15408d0c84a51105cb3ff42
Dia2 0:00:01	E1	489c2ef9ec70b04dd9f053ae5c77aeb5b24cc337bcb69d3a030d2bde54486615
Dia2 0:00:03	E4	8cc92c4fe819a035ae1ba2f9a1d63f6c1948aa60c71a0265807f68f85ccf03d7

Figura 4-1: Ejemplo de datos obtenidos con los WS.

Se cuenta con datos de afluencias de SC a las estaciones de la prueba piloto cada 15 minutos de 6 semanas del 2016. Cada observación incluye el día, el periodo del día, la estación y la cantidad de usuarios separando a los estudiantes (escolares y universitarios) y tercera edad del resto de los usuarios.

La implementación computacional se hizo con periodos de 15 minutos (aunque es un parámetro fácilmente ajustable) para ajustarse a los datos de las SC y porque para observar

la variabilidad en ciertos momentos del día se necesitan periodos de esa magnitud. El fin de un periodo y el inicio del siguiente están determinados por los mismos periodos que las afluencias de las SC (14:00, 14:15, 14:30, 14:45, 15:00, etc.). En la Figura 4-2 se puede ver un ejemplo de datos obtenidos con las SC. El periodo ‘14.2’ corresponde al segundo periodo de 15 minutos de las 14 horas, es decir entre las 14:15 y 14:30.

día	periodo	origen	usuarios_normales	usuarios_estudiantes	usuarios_3raedad
Dia1	23,2	E4	22	4	0
Dia1	23,3	E4	0	1	0
Dia2	14,1	E3	869	146	95
Dia2	14,2	E3	808	134	87
Dia2	14,3	E3	950	150	105

Figura 4-2: Ejemplo de datos obtenidos con las SC.

#### 4.1. Descripción de la red

La red utilizada consta de cuatro estaciones consecutivas ubicadas en un extremo de la red del STP. Las estaciones se denotan por E1, E2, E3 y E4, donde E1 corresponde a la estación terminal (o extrema) y E4 conecta a esta subred con el resto de la red del STP.

Estas estaciones están físicamente aisladas de otros medios de transporte. La importancia de esto es que no es necesario aplicar un filtro sofisticado a las detecciones para eliminar los viajes en otros modos de transporte. Además, la red de estaciones equipadas tiene una única ruta entre cada par de estaciones. Esto simplifica la metodología ya que no es necesario analizar la ruta que tomó cada usuario para determinar la carga por arco. Tampoco existen estaciones de combinación con otros recorridos del STP, por lo que no hay entradas ni salidas en estaciones intermedias a otras estaciones de la red.

El ponderador de la estación que conecta a esta subred con el resto del STP (E4) no tiene sentido utilizarlo ya que la afluencia de usuarios a la estación estimada con WS es mucho más grande que los que entraron con SC, ya que los WS detectan a todos los usuarios que entraron por otras estaciones fuera de la red de la prueba piloto pero viajaron hacia ella.

Para resolver este problema, se le asigna a esta estación el ponderador de la E3 en el mismo periodo de tiempo por ser la estación más cercana.

La red de la prueba piloto tiene una estación terminal. Como el ponderador de la metodología se calcula en base a la estación de origen, hay usuarios que entran a la red del STP por una estación que no tiene un WS (porque no es parte de la red de la prueba piloto) y luego se trasladan a las estaciones que sí tienen WS. Esto causa una limitación a la prueba piloto, ya que la metodología asumiría que este usuario entró a la red por la estación en que fue detectado por primera vez. Este error se tendrá en todas las estaciones con WS que estén conectadas directamente con una estación que no tenga WS.

Cabe destacar que este problema no se tendría en una implementación masiva de la metodología con WS en toda la red del STP. En ese caso, todos los dispositivos tendrían un ponderador correspondiente al momento y lugar real de su entrada a la red. Para esta prueba piloto se tenía un número limitado de WS, por lo que se buscó una sección de la red que buscara reducir la fuente de error mencionada.

#### **4.2. Instalación en terreno**

El WS se instaló en un lugar elegido estratégicamente para captar la mayor cantidad de señales de los dispositivos de los usuarios. En términos simples, el lugar elegido debía estar lo más cerca del centro de la estación, en altura y sin grandes objetos que lo obstruyan (incluyendo flujo de personas).

Otro factor importante a la hora de decidir la ubicación fue que estuviese cerca de la salida de la estación ya que en algunas estaciones la entrada a la estación es diferente a la salida. La razón es que los usuarios que entran a la estación están un tiempo esperando el vehículo por lo que hay más chances que su dispositivo envíe una señal, pero un usuario que sale de la estación solo está unos segundos dentro del área de detección así que se prioriza ese movimiento de usuarios al ubicar el WS más cerca de la salida.

### **4.3. Implementación computacional de la metodología.**

La implementación computacional de la metodología se realizó en el lenguaje de base de datos (BD) postgresQL 9.5 para ser ejecutado en el *software* pgAdmin III. La programación se basó en el uso de *Triggers*, que es una función de base de datos que se ejecuta automáticamente cuando ocurre un evento específico. Las únicas dos fuentes de datos externa son las detecciones de los WS y la información de afluencias a las estaciones de la SC. Al ingresar cada una de esas piezas de información en la base de datos, se van generando todos los resultados intermedios y finales de la metodología automáticamente gracias a los *Triggers*.

El sistema está preparado para recibir información en tiempo real, es decir que cuando ocurra una detección en cualquier WS, ésta podría ingresar al sistema y se ejecutarán los eventos correspondientes. Por otro lado, las afluencias de la SC se ingresan en los periodos temporales descritos antes por la metodología, inmediatamente una vez que termina el periodo. El programa asume que la información se le está entregando en tiempo real y en el orden descrito (detecciones a medida que ocurren y afluencias al final de cada periodo).

Se creó un programa controlador del flujo de información para simular que los datos están llegando en tiempo real. El programa controlador fue programado en el lenguaje C# y ejecutado en el *software* Visual Studio. En el Anexo B se describe brevemente este programa.

#### **4.3.1. Descripción del programa en el lenguaje de BD PostgreSQL**

Se programó la metodología descrita en el capítulo anterior. Las variables teóricas fueron representadas por diferentes tablas de bases de datos. Algunas tablas tienen asociados *Triggers* que se gatillan cuando ocurre un evento (se inserta, actualiza o elimina una fila de dicha tabla) y ejecutan una acción en otra tabla.

En la Figura 4-3 se puede ver la relación entre las tablas del programa en PostgreSQL, la variable de la metodología a la que están asociadas y los *Triggers* que se ejecutan cuando ocurre un evento en dicha tabla. Dentro de cada rectángulo está primero el nombre de la

tabla del programa y luego la variable teórica que tiene asociada. Junto a cada flecha punteada está el nombre del *Trigger* asociado a esa flecha.

Además, se muestran las relaciones que existen entre tablas. Cada flecha representa una relación del siguiente tipo:

- Dos tablas unidas por una flecha continua (a) significa que la primera tabla aporta los datos necesarios para construir la segunda tabla.
- Dos tablas unidas por una flecha punteada (b) significa que una acción ejecutada (ya sea insertar, actualizar o eliminar un dato, dependiendo de cada caso) en la primera tabla gatilla un *Trigger* que tiene un efecto en la segunda tabla.
- Dos tablas unidas por una flecha discontinua (c) significa que el programa externo en C# se encarga de manejar el flujo de información de la primera tabla hacia la segunda.

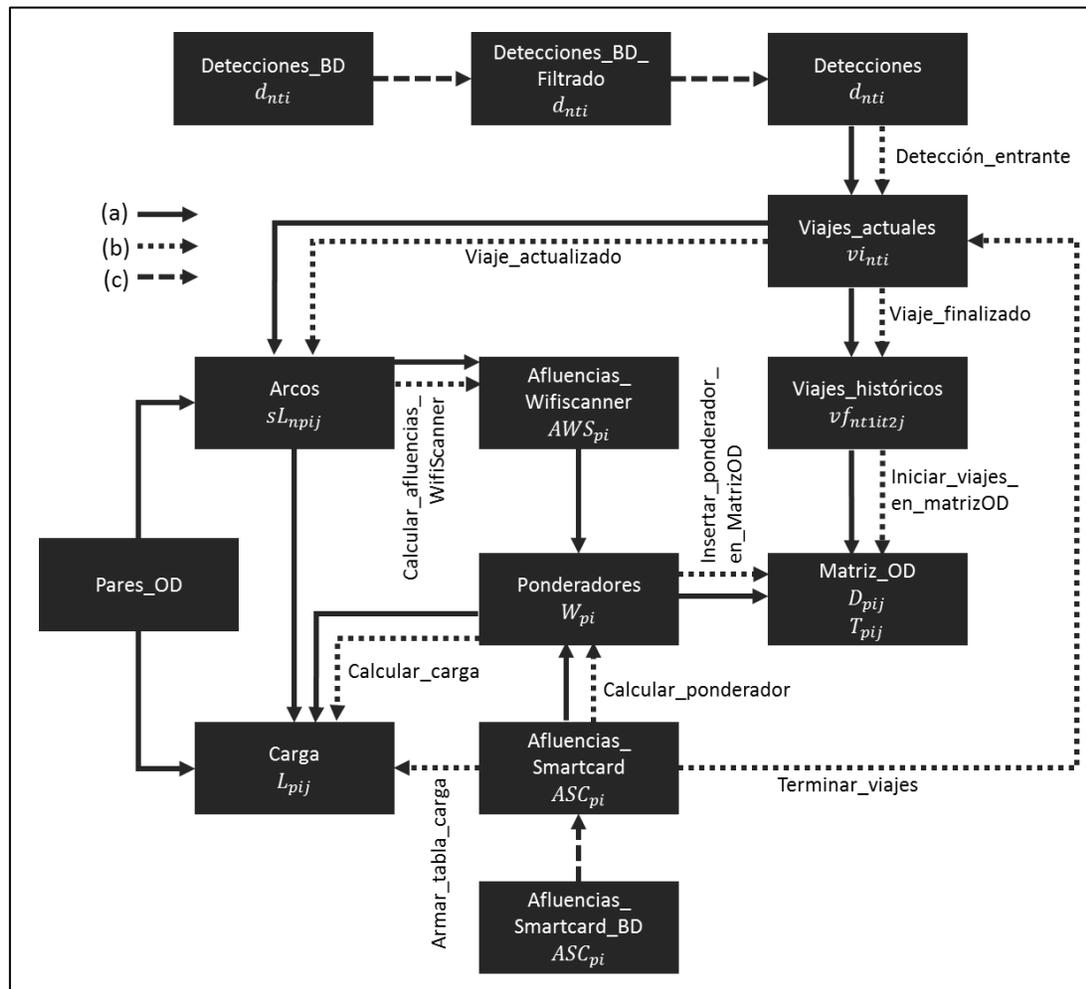


Figura 4-3: Diagrama con flujo de información a través de tablas de base de datos.

A continuación, se menciona qué gatilla cada *Trigger* y se explica su función con foco en lo que se diferencia la implementación con la metodología descrita anteriormente. Las tablas que contiene las letras ‘BD’ son tablas que sirven como base de datos para que el programa en PostgreSQL trabaje como si la información estuviese entrando en tiempo real. La información pasa de las tablas BD al programa en PostgreSQL a través del programa externo escrito en C#.

### Trigger Deteccion\_entrante

Se gatilla luego de insertar una detección  $d_{nti}$  en la tabla Detecciones.

Función: actualiza el viaje en curso del dispositivo  $n$  que fue detectado o inicia un nuevo viaje si es que el dispositivo no estaba viajando.

Actualizar el viaje de un dispositivo implica que se modifique una variable  $sL_{npj}$  si es que el dispositivo  $n$  cruza un arco. En la implementación, la tabla Arcos almacena de forma agrupada las distintas variables  $sL_{npj}$  de distintos dispositivos que comparten el mismo periodo de origen  $p$ , estación  $i$  de origen del arco y estación  $j$  de destino del arco. Los resultados no se ven afectados, solo es una manera de almacenar la información que reduce la cantidad de código ejecutado.

### **Trigger Terminar\_viajes**

Se gatilla luego de insertar un dato de afluencia  $ASC_{pi}$ , entregado por las SC para la estación  $i$  durante el periodo  $p$ , en la tabla Afluencias\_Smartcard. La llegada del dato de afluencia marca el fin del periodo  $p$  y comienzo del periodo  $p+1$ .

Función: elimina de la tabla Viajes\_actuales a todos los viajes de dispositivos que no hayan sido detectados hace más de  $L$  minutos. En esta implementación,  $L = 10$  minutos.

Observación: el proceso de considerar viajes finalizados descrito en el capítulo anterior requiere una revisión constante de toda la lista de viajes actuales para ver cuáles es necesario eliminar es esa lista. Esto es bastante costoso computacionalmente, por lo que se optó por revisar solamente cuando se ingrese información de entrada de afluencias reales a las estaciones. Como se mencionó anteriormente, en esta implementación el largo de  $p$  fue de 15 minutos.

Por lo tanto, un viaje no necesariamente será eliminado exactamente  $L$  minutos después de haber sido detectado la última vez. Sin embargo, sí es posible entregar cotas al tiempo que puede pasar entre la última detección de un dispositivo y el instante en que el viaje de ese dispositivo es considerado como finalizado.

Un dispositivo que su última detección es justo antes de  $L$  minutos antes del final del periodo  $p$ , será considerado como que su viaje ha terminado en ese mismo periodo  $p$ , ya que pasó más de  $L$  minutos desde la última detección hasta el momento de la revisión al final del periodo  $p$ . Por lo tanto, el tiempo mínimo que tiene que pasar para eliminar un viaje es  $L$  minutos.

Por otro lado, si la última detección es justo después de  $L$  unidades de tiempo antes del final del periodo  $p$ , el viaje no será considerado como finalizado al final del periodo  $p$  ya que todavía no han pasado  $L$  minutos desde su última detección. Luego, ese dispositivo será eliminado al final del periodo  $p+1$ . Por lo tanto, el máximo tiempo que puede transcurrir para que un dispositivo sea eliminado es  $L + p$  minutos.

### **Trigger Armar\_tabla\_carga**

Se gatilla luego de insertar un dato de afluencia  $ASC_{pi}$ , entregado por las SC para la estación  $i$  durante el periodo  $p$ , en la tabla Afluencias\_Smartcard.

Función: prepara la tabla donde se almacena la variable  $L_{pij}$ . Se llegó a la conclusión que crear la tabla con un *Trigger* diferente al que inserta datos en la misma tabla era la mejor opción computacionalmente.

### **Trigger Calcular\_ponderador**

Se gatilla luego de insertar un dato de afluencia  $ASC_{pi}$ , entregado por las SC para la estación  $i$  durante el periodo  $p$ , en la tabla Afluencias\_Smartcard.

Función: calcula el ponderador  $W_{pi}$  utilizando la información de la tabla Afluencias\_Smartcard y Afluencias\_Wifiscanner.

Observación: como se mencionó en el capítulo 3, la variable  $AWS_{pi}$  tiene un leve desfase temporal en su cálculo. Dado que puede haber dispositivos que hayan entrado a la red por la estación  $i$  en el periodo  $p$  y que al término del periodo  $p$  aún no han sido detectados en una de las estaciones aledañas a  $i$ , no pueden ser considerados como dispositivos que están

viajando. Por esta razón, el ponderador  $W_{pi}$  se calcula al final del periodo  $p$  con la información que se tiene hasta el momento, pero se actualiza su valor cuando se calculan los ponderadores de los periodos siguientes. La cantidad de periodos que se vuelven a calcular varía dependiendo del largo del periodo con el que se está trabajando, el tiempo que podría transcurrir desde que alguien entra a la red hasta que es detectado en la siguiente estación y de la capacidad computacional que se tenga disponible. En esta implementación con periodos de 15 minutos, la primera estimación del ponderador era bastante estable y solo cambiaba marginalmente cuando se actualizaba.

### **Trigger Viaje\_actualizado**

Se gatilla luego de actualizar un viaje en la tabla Viajes\_actuales, es decir, una vez que un viaje que ya se inició se movió a otra estación.

Función: inserta la información del viaje de una estación a otra (si es que la actualización del viaje incluye un cambio de estación) en la tabla Arcos.

Observación: puede ocurrir que un dispositivo realizó cierto viaje de una estación a otra que no está directamente conectada (es decir, no son estaciones vecinas), ya que no fue detectado por el WS de la(s) estación(es) intermedia(s). En ese caso, se recurre a la tabla Pares\_OD que contiene la información de la ruta entre cada par de estaciones. El viaje se carga entonces a todas las estaciones por las que debiese haber pasado, en el instante de la detección que actualizó el viaje.

Se podría haber estimado tiempos de viaje entre estaciones para asignar un mejor valor al instante en el que se asigna el cruce de estas detecciones ficticias, pero cuando el dispositivo se salta una o dos estaciones el tiempo que se le asigna es poco significativo ya que solo afectaría si pasa de un periodo a otro.

### **Trigger Viaje\_finalizado**

Se gatilla luego de eliminar un viaje en la tabla Viajes\_actuales ya que el dispositivo que realiza el viaje no ha sido detectado por más de  $L$  minutos.

Función: inserta el viaje eliminado en la tabla de Viajes\_historicos, siempre y cuando el viaje tenga como origen y destino una estación diferente. Esto para no considerar los dispositivos fijos en las estaciones ni dispositivos que pueden haber pasado por la estación pero sin realizar un viaje a otra estación.

### **Trigger Insertar\_viajes\_en\_matrizOD**

Se gatilla luego de insertar un viaje  $v_{f_{nt1}it2j}$  en la tabla Viajes\_historicos.

Función: incluye el viaje insertado en el par OD y periodo correspondiente en la tabla Matriz\_OD. También se multiplica el nuevo valor de viajes de dispositivos por el ponderador correspondiente para obtener la matriz OD de personas. Puede que el ponderador no esté disponible todavía cuando termina algún viaje corto, en cuyo caso simplemente se va aumentando la cantidad de viajes de dispositivos y se espera hasta que el periodo termine para que el ponderador se calcule.

### **Trigger Calcular\_carga**

Se gatilla luego de insertar o actualizar un ponderador  $W_{pi}$  en la tabla Ponderadores.

Función: calcula la carga por arco a partir de la tabla Arcos y la inserta en la tabla Carga.

### **Trigger Insertar\_ponderador\_en\_matrizOD**

Se gatilla luego de insertar o actualizar un ponderador  $W_{pi}$  en la tabla Ponderadores.

Función: inserta el ponderador  $W_{pi}$  en la tabla Matriz\_OD en todas las filas que tengan como origen la estación  $i$  durante el periodo  $p$ . Multiplica los viajes de WS que tiene hasta ahora en esos pares por el ponderador insertado. Notar que más adelante se podrían agregar más viajes ya que cuando el ponderador se calcula, no necesariamente los viajes entre la estación  $i$  y todo el resto de las estaciones han terminado. En caso de que la matriz OD esté siendo revisada en tiempo real, se debe tener en cuenta que se estará subestimando la

cantidad de viajes largos ya que como todavía no han terminado, no aparecen en la matriz todavía cuando muchos viajes cortos ya han aparecido.

### **Trigger Calcular\_afluencias\_WifiScanner**

Se gatilla luego de insertar o actualizar la variable  $sL_{npj}$  en la tabla Arcos.

Función: aumenta la variable  $AWS_{pi}$  de la tabla Afluencias\_Wifiscanner. Cuando se inserta o actualiza la variable  $sL_{npj}$  significa que el dispositivo  $n$  cruzó de una estación a otra. Cuando ese cruce es el primer cruce que hace un dispositivo, entonces este pasa a considerarse como que está viajando y por lo tanto la afluencia de dispositivos aumenta.

## **5. RESULTADOS Y DISCUSIÓN.**

En esta sección se muestran los resultados obtenidos en la prueba piloto y se realiza un análisis exhaustivo de los mismos. Se analizó una semana laboral tipo (de lunes a viernes) donde no hubiesen ocurrido incidentes que afecten la red ni tampoco eventos o situaciones externas que hubiesen podido afectar el comportamiento de los usuarios. Se obtuvo y estudió la penetración de la tecnología WS en el tiempo y el espacio, la matriz OD y la carga por arco.

Además, se muestran los resultados de la estimación de tiempos de espera y de la cantidad de viajes diarios por dispositivo. Finalmente, se analizó un día que ocurría una contingencia en la ciudad que generó un adelantamiento de los viajes en la tarde.

### **5.1. Ponderadores y penetración.**

Este trabajo desarrolló una forma de unir dos fuentes de datos de manera inteligente, y la expansión de señales de los WS a partir de la SC es la instancia en que estas dos fuentes de datos se unen. En primer lugar, se muestran en la figura 5-1 las afluencias de la SC sumadas para las estaciones de la prueba piloto para la semana de estudio, sin agregar la estación E4.

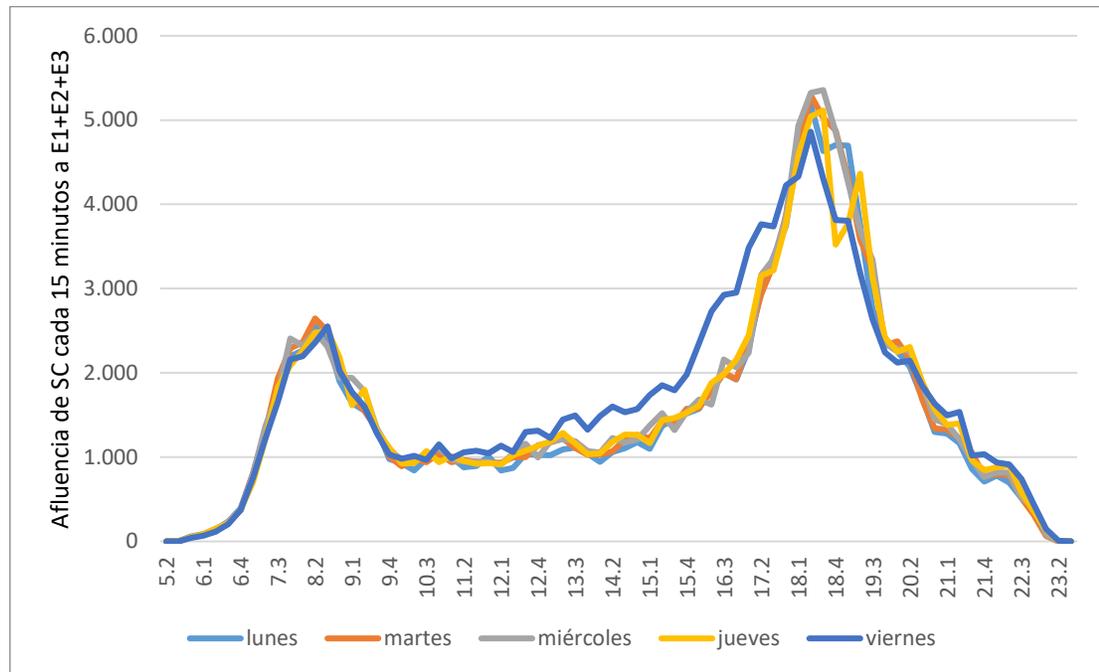


Figura 5-1: Afluencias sumadas de las estaciones E1, E2 y E3 cada 15 minutos para una semana laboral obtenidas a través de la SC.

Se puede ver que el día lunes, martes, miércoles y jueves son muy similares. El día viernes es un día particular ya que en general las personas salen antes de su trabajo, por lo que la hora punta de la tarde ocurre más temprano. Por otro lado, el día viernes en la hora punta mañana no tiene diferencias significativas con el resto de los días de la semana. El día jueves tiene una baja pequeña en la hora punta tarde a la que no se le pudo dar explicación.

Ahora, se puede ver en la figura 5-2 las afluencias de dispositivos Wifi obtenidas con los WS según la metodología descrita anteriormente. La afluencia de la estación de borde (E4) no se considera ya que un viaje que se inició en esta estación en realidad podría haberse iniciado fuera de las estaciones de la prueba piloto, distorsionando el valor de la penetración ya que se estarían contando viajes de usuarios que validaron su entrada con la SC en una estación diferente.

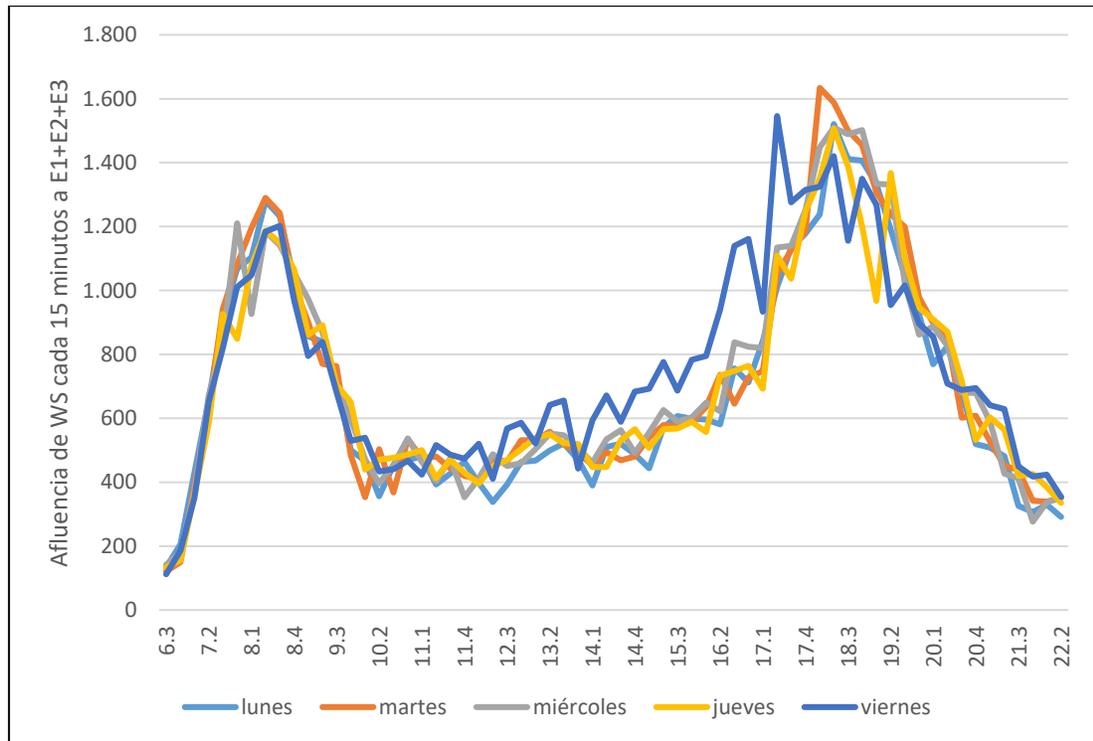


Figura 5-2: Afluencias sumadas de las estaciones E1, E2 y E3 para una semana laboral obtenidas a través de los WS.

Las curvas de afluencia registradas por los WS tienen una forma muy similar a aquellas obtenidas mediante la SC. La mayor diferencia es que durante la punta tarde de los WS se alcanzan niveles aproximadamente un 25% más altos que la punta mañana, pero en el caso de las curvas de la SC la punta tarde es aproximadamente el doble (100%) que la punta mañana. Además, se nota que los WS tienen mayor dispersión entre un periodo y el siguiente.

### 5.1.1. Penetración a través del tiempo

Con las afluencias obtenidas a partir de las SC y afluencias obtenidas con los WS, se puede estimar de manera sencilla la penetración de la tecnología Wifi en cada estación para los

inicios de los viajes. No es lo mismo que estimar cuántas de las personas que hay en la estación tienen dispositivo Wifi, sino que cuántas de las personas que inician su viaje en la estación tienen dispositivo Wifi. La segunda información es más importante para la metodología ya que el ponderador que se le asigna a cada dispositivo depende de la estación y momento de origen del viaje. En la figura 5-3 se puede ver la penetración por periodo del día para las estaciones E1, E2 y E3 agregadas.

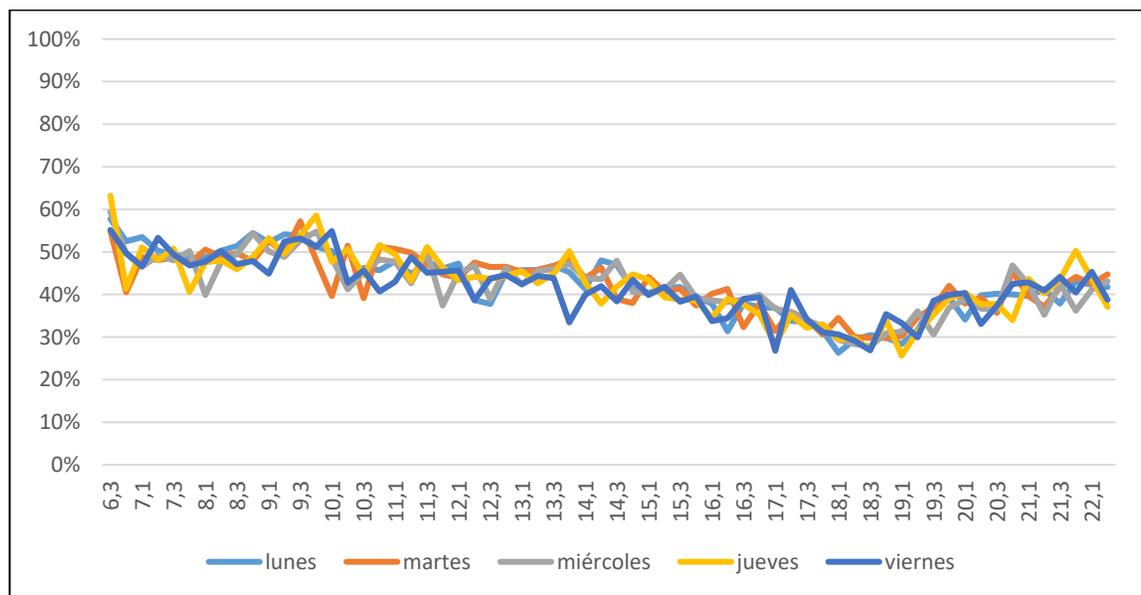


Figura 5-3: Penetración conjunta de E1, E2 y E3 estimada cada 15 minutos.

La disminución en la penetración desde aproximadamente un 50% en la mañana a un 30% en la tarde no es fácil de explicar *a priori*. Se tienen cuatro hipótesis no excluyentes para explicar esta baja en la penetración de la tarde. A continuación, se analiza una a una las cuatro teorías:

**Hipótesis 1: Los usuarios que ingresan a estas estaciones en la mañana tienen mayor ingreso económico que los de la tarde, por lo que hay mayor probabilidad que cuenten con un smartphone.**

Como se mencionó en el capítulo 3 de este trabajo, una persona con bajo ingreso tiene menor probabilidad de poseer un *smartphone* que una persona de alto ingreso (Pew Research Center, 2016). Por lo tanto, en sectores de alto ingreso la penetración observada debiese ser más alta que en sectores de bajo ingreso, cuando las personas que viven en ese sector inician su viaje.

Los hogares cercanos a donde se realizó la prueba piloto se caracterizan por ser: 14,1% de ingreso bajo, 45,9% de ingreso medio y 40,1% de ingreso alto. Por otro lado, esos valores para la ciudad completa son 35,2%, 55,1% y 9,7%, respectivamente. Por lo tanto, las personas que durante la mañana usan como estación de origen alguna de las estaciones de la prueba piloto ya que viven cerca de ahí, tienen mayor probabilidad de tener *smartphone* que el resto de la población dado que tienen más alto ingreso.

**Hipótesis 2: dada la alta afluencia que se observa en la tarde, los WS alcanzan su capacidad y no son capaces de identificar todos los dispositivos que ingresan a las estaciones.**

Las afluencias de SC a las tres estaciones agrupadas no tienen la misma magnitud en su punto más alto en la mañana que en la tarde. De hecho, ese punto es aproximadamente el doble en la tarde que en la mañana. Se sospecha que la baja de penetración en la tarde entonces esté dada porque el WS no es capaz de detectar las señales de la gran cantidad de personas que está entrando a la estación en la tarde, y por lo tanto la proporción de personas que se considera que tienen un dispositivo es menor que los de la mañana. Además, la penetración más baja se da cuando las afluencias de SC son más altas.

Para comprobar esto, se tomaron todas las detecciones (sin considerar detecciones repetidas en el mismo segundo) enviadas por los dispositivos que iniciaron su viaje en las tres estaciones analizadas (E1, E2 y E3). Se pueden ver estas detecciones para el día lunes de la semana laboral analizada en la figura 5-4 en conjunto con las afluencias a las estaciones obtenidas de las SC. Si en algún momento la curva de afluencias de SC sube pero la curva de detecciones totales no la acompaña con la misma pendiente sino que se mantiene plana, podríamos sospechar que el WS ha alcanzado su capacidad máxima de detección. Como se

ve en la figura 5-4, esto no ocurre. Ambas curvas se acompañan y alcanzan sus máximos en los mismos momentos del día, por lo que se puede asegurar que los WS no colapsan.

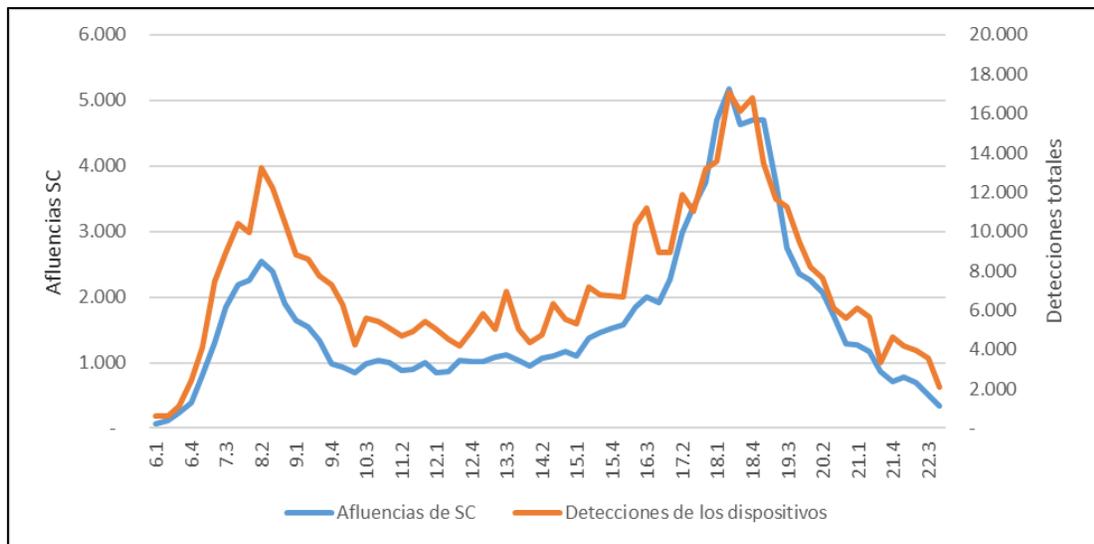


Figura 5-4: Detecciones totales de dispositivos y afluencias de SC para las estaciones E1, E2 y E3 agrupadas por periodo del día lunes de la semana analizada.

Se hizo este análisis para otros días laborales y se obtuvo resultados similares.

**Hipótesis 3: los usuarios usan más su *smartphone* en la mañana que en la tarde, por lo que es más probable que sean identificados por un WS.**

Otro factor que puede influir en la penetración es el uso que le den los usuarios a su *smartphone* mientras circulan por el área de detección de los WS. A mayor uso, más señales envía el dispositivo y por lo tanto mayor es la probabilidad de ser detectado. A priori, se podría pensar que las personas usan más el dispositivo en la mañana ya que van preparándose para el día laboral, contestando mensajes que llegaron durante la noche, revisando el correo y porque probablemente tienen más batería que en la tarde.

Esto provocaría no solo que la cantidad de detecciones por dispositivo aumente sino que también que más dispositivos sean detectados y por lo tanto aumente la penetración. Para comprobarlo se graficaron las detecciones promedio por dispositivo (sin considerar detecciones repetidas en el mismo segundo) enviadas por los dispositivos que iniciaron su viaje en las tres estaciones analizadas en conjunto con la penetración. Los datos son del día lunes de la semana laboral analizada, pero para otros días laborales se ven resultados que conducen a las mismas conclusiones.

En la figura 5-5 se puede ver los resultados, demostrando que a medida que la penetración baja la cantidad promedio de detecciones por dispositivo se mantiene prácticamente constante (aunque con bastante variabilidad). La curva Linear, que es la regresión lineal de la curva Detecciones por dispositivo tiene una pendiente prácticamente nula de 0.02. Lo anterior sugiere que la diferencia entre el uso de *smartphone* entre la mañana y la tarde no es significativa como para reducir la penetración de un 50% a un 30%.

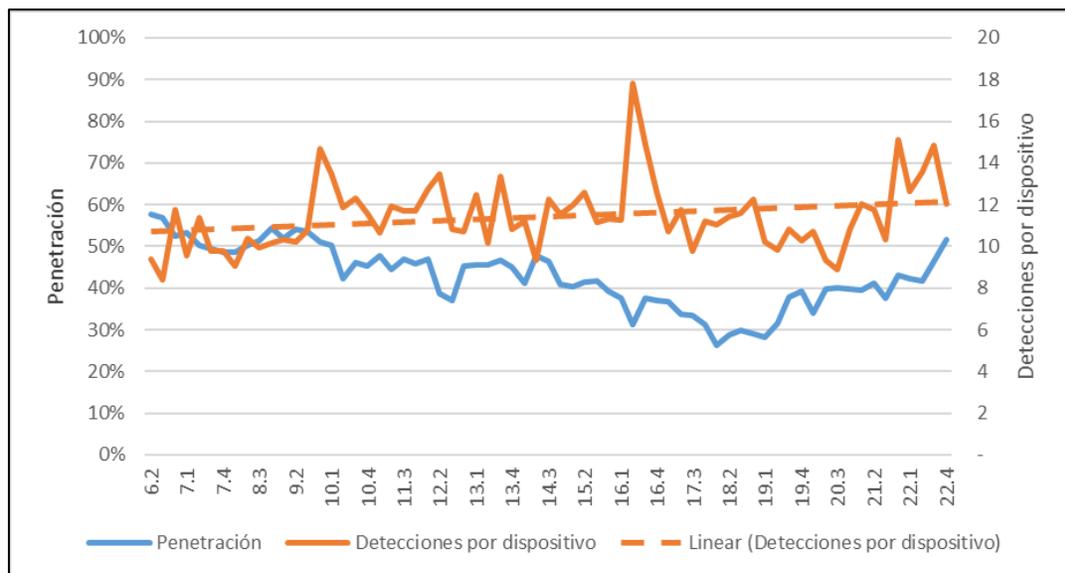


Figura 5-5: Detecciones por dispositivo y penetración del día lunes.

**Hipótesis 4: la edad promedio de quienes ingresan a las estaciones es menor en la mañana comparada con la tarde, por lo que es más probable detectar un dispositivo en la mañana.**

En el capítulo 3 se menciona que las personas más jóvenes (entre 18 y 34 años) tienen mayor probabilidad de tener *smartphone*. Los estudiantes en general son personas más jóvenes y además la intuición indica que un estudiante tiene mayor probabilidad de tener *smartphone* y de usarlo. Los viajes de los estudiantes se podrían considerar un aproximado a los viajes de personas jóvenes. El cambio de penetración entonces podría estar afectado por el momento del día en que los estudiantes están viajando. Afortunadamente, la información de la SC está desagregada entre las personas que pagan pasaje completo, estudiantes y adulto mayor representando el 75%, 21% y 4% de los viajes de lunes a viernes de todos los datos de SC que se tienen en las tres estaciones analizadas.

Se separaron los datos de SC por día y para el periodo AM (entre 7:00 y 10:00) y PM (entre 17:00 y 20:00). Estos periodos fueron elegidos porque es entre esos periodos cuando el cambio de penetración es más brusco. En la Tabla 5-1, se presenta para cada día la proporción de los viajes totales del periodo que son realizados por estudiantes en las tres estaciones analizadas agrupadas.

Tabla 5-1: Proporción de viajes realizados por estudiantes para E1, E2, y E3 agrupadas por periodo del día.

	MES 1												MES 2								
	Día	AM	PM	Día	AM	PM	Día	AM	PM	Día	AM	PM	Día	AM	PM	Día	AM	PM			
Lu				05	26%	18%	12	25%	18%	19	12%	15%	26	26%	18%	03	26%	18%	10	15%	18%
Ma				06	26%	18%	13	25%	19%	20	20%	16%	27	27%	19%	04	26%	19%	11	25%	18%
Mi				07	26%	18%	14	25%	18%	21	23%	17%	28	26%	19%	05	26%	19%	12	26%	18%
Ju	01	25%	18%	08	26%	18%	15	23%	18%	22	23%	18%	29	26%	19%	06	26%	19%			
Vi	02	24%	17%	09	24%	17%	16	19%	14%	23	22%	16%	30	25%	17%	07	24%	17%			

Se puede ver una clara diferencia entre ambos periodos, con un promedio de 25% para AM y 18% para PM. La disminución de penetración de la tecnología Wifi de un 50% en la mañana a un 30% en la tarde puede estar en parte explicada por la disminución de

proporción de estudiantes en la tarde, y por lo tanto la disminución de proporción de individuos con alta probabilidad de tener *smartphone*.

En conclusión, a pesar que no se puede afirmar a ciencia cierta que la baja en la penetración depende exclusivamente del nivel de ingreso de las personas y de la cantidad de estudiantes que ingresen a la red, es altamente probable que la disminución de penetración que se ve en la tarde en la prueba piloto se deba en parte a estos efectos.

Se demostró que la penetración no es constante en el tiempo y por lo tanto el método para expandir la muestra de patrones de viaje obtenidos con los WS debe depender del tiempo.

### **5.1.2. Penetración a través del espacio.**

El mismo efecto que se ve en la red en las estaciones E1, E2 y E3 de alto ingreso se podría ver en otras estaciones pero de forma opuesta. Es decir, que la penetración sea más alta en la tarde ya que cerca de esas estaciones trabajan o estudian personas de más alto ingreso que las que viven cerca de ahí.

Con eso en mente, también es importante entender qué pasa entre estaciones que están cerca y por lo tanto no debiesen tener diferencias tan significativas en cuanto al nivel de ingreso de los usuarios que las visitan.

La realidad es que esta situación no ocurre para la prueba piloto realizada. En la Figura 5-6 se puede apreciar la penetración para el día laboral lunes (los otros días se ven resultados similares) separada para las tres estaciones analizadas.

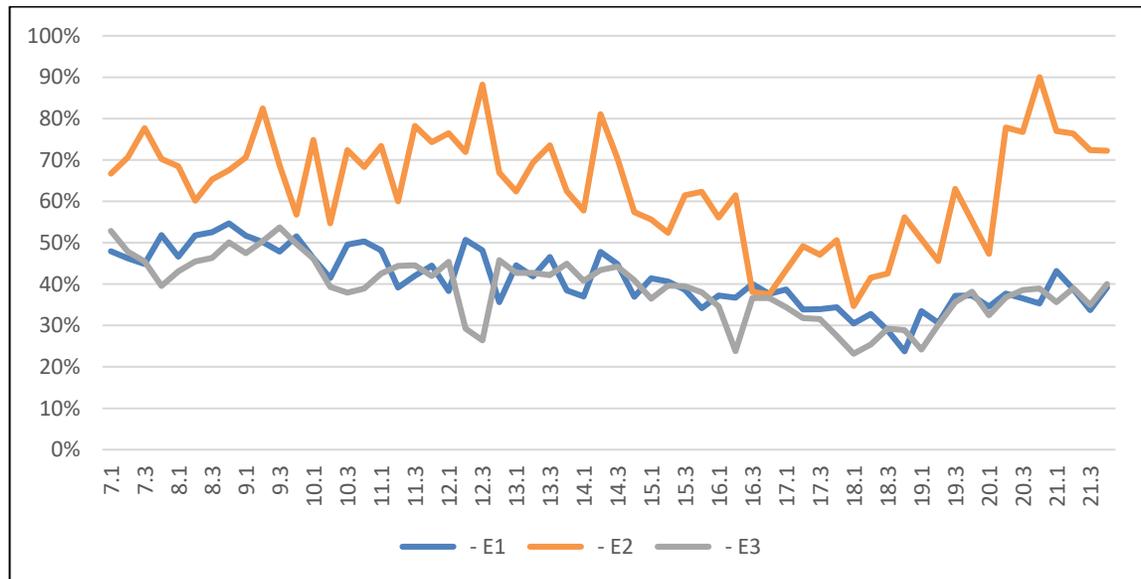


Figura 5-6: Penetración para el lunes de cada estación analizada.

Se puede ver que E3 y E1 tienen penetraciones muy parecidas, pero E2 es más alta que las otras dos. Las tres estaciones tienen el efecto de disminución de penetración en la tarde. Se tienen dos hipótesis que podrían explicar por qué E2 tiene mayor penetración durante todo el día que las otras dos estaciones. A continuación, se analiza cada una de las teorías:

#### **Hipótesis 1: la proporción de estudiantes es más alta en E2 que E1 y E3.**

Tal como se vio para el cambio temporal de la penetración, la cantidad de estudiantes es un factor importante al momento de calcularla ya que en general tienen mayor probabilidad de tener *smartphone*. La proporción promedio de estudiantes, obtenida con las SC, de cada una de estas tres estaciones para todos los días de lunes a viernes de todos los datos de SC que se tienen son:

- E1: 24%
- E2: 29%
- E3: 18%

Estos valores son muy estables entre todos los días analizados. La proporción de E2 es más alta que en E1 y E3, por lo que la proporción de estudiantes podría explicar la diferencia

entre E2 y el resto. Sin embargo, llama la atención que E1 sea mayor a E3 y entre ellas no tengan diferencia significativa en sus penetraciones.

**Hipótesis 2: El WS en E2 está mejor ubicado dentro de la estación que E1 y E3, por lo que es más probable que pueda detectar dispositivos.**

El WS de E2 tiene una ubicación favorecida en comparación con las otras dos. En teoría la antena es capaz de captar todas las señales de dispositivos que ingresan al área de detección, pero la tecnología Wifi tiene un factor importante de aleatoriedad en la captación de la señal, sobre todo porque los dispositivos son los que envían las señales en los instantes que estimen conveniente, y el WS solo escucha.

La estación E2 tiene solo un punto de acceso y el WS está ubicada justo en ese acceso, por lo que todo dispositivo que ingresa pasa cerca del WS. Por otro lado, las estaciones E1 y E3 tienen más de un punto de acceso por lo que muchos dispositivos que ingresan a la estación nunca pasan cerca del WS y por ende baja la probabilidad de ser detectado,

Además, la estación E2 tiene pocos objetos grandes que bloqueen las señales. E3 y E1 tienen varios puentes cada una, sobre todo E3 que tiene varias estructuras de concreto dentro de la estación. Estos grandes bloques de concreto y metal bloquean las señales Wifi y hacen que sea más difícil la detección.

Por último, muchos más usuarios eligen como origen las estaciones E3 y E1 que E2. En la Figura 5-7 se puede ver las afluencias a las tres estaciones para el día lunes (la diferencia se mantiene para otros días laborales). Esto indica que el WS de la estación E2 tiene menor interferencia provocada por usuarios que las otras dos estaciones al momento de detectar los viajes que se originan ahí.

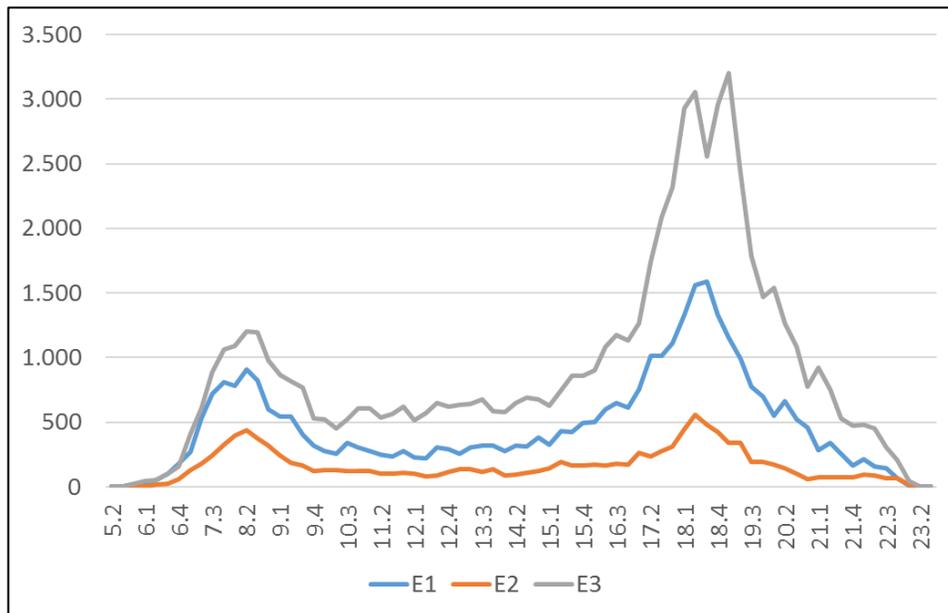


Figura 5-7: Afluencias de SC por estación por periodo para el lunes.

## 5.2. Matriz OD.

Se obtuvo la matriz OD de la red de la prueba piloto cada 15 minutos para toda la semana laboral analizada, la que denotaremos por  $MOD_{ws}$ . En esta sección se muestran los resultados de la  $MOD_{ws}$  del martes entre las 17:00 y las 20:00. Para construir la matriz de todo el periodo en base a las matrices de 15 minutos simplemente se sumó la cantidad de personas durante todo el periodo de análisis.

La  $MOD_{ws}$  fue comparada con la matriz OD del STP obtenida con una metodología basada en el uso de la SC para estimar el origen y destino de los viajes de la red, la que denotaremos por  $MOD_{sc}$ . Se tiene la  $MOD_{sc}$  del día del martes para el mismo periodo horario. Las matrices están separadas por 17 meses. Sin embargo, la red del STP no cambió de manera significativa en ese periodo. Por lo tanto, si bien es esperable que las matrices no sean iguales, el cambio en los patrones de viaje se espera que sea menor.

Para comparar estadísticamente las matrices, se utilizó el test chi cuadrado con  $MOD_{sc}$  como la matriz esperada. Usando  $MOD_{sc}$  se calcularon las probabilidades ( $P_{ij}$ ) de ir de

cada estación  $i$  a cada estación  $j$ , y luego se multiplicó cada una de esas probabilidades por el total de viajes de la  $MOD_{WS}$ . Así, se obtiene la matriz OD esperada que luego se compara con la  $MOD_{WS}$ . El resultado fue que se rechazó la hipótesis nula, es decir, las matrices son estadísticamente diferentes.

A continuación, en la Figura 5-8, se muestran ambas matrices OD de forma porcentual, es decir, cada celda representa el  $P_{ij}$  para cada matriz.

	E1	E2	E3	Red
E1		0.2%	0.5%	20.8%
E2	0.1%		0.1%	5.5%
E3	0.5%	0.2%		45.5%
Red	8.6%	4.2%	14.0%	

(a)

	E1	E2	E3	Red
E1		3.8%	2.8%	13.5%
E2	0.9%		1.0%	4.5%
E3	2.0%	2.1%		38.2%
Red	7.8%	6.7%	16.6%	

(b)

Figura 5-8: (a)  $MOD_{SC}$ . (b)  $MOD_{WS}$ .

Cabe destacar que  $MOD_{SC}$  no es el resultado que se esperaba llegar ya que fue estimada hace 17 meses y además es otra metodología para estimar la matriz OD. Por esto, no necesariamente representa la realidad. Más aun, se espera que las matrices tengan diferencias ya que ambas metodologías tienen limitaciones que hacen que su estimación se aleje de la realidad. Dicho esto, se cree que  $MOD_{SC}$  se acerca más a la realidad según la experiencia de uso en esas estaciones y conteos simples realizados en las estaciones de la prueba piloto.

La razón más importante de las diferencias se debe a que el WS no siempre detecta a todos los dispositivos en su estación de origen o de destino. Si no lo detecta en una estación intermedia el resultado de la matriz OD no se ve afectado. Pero, si el dispositivo no es detectado por ejemplo en su estación de origen, luego viaja a una estación aledaña y sí es detectado por el WS en esa estación, entonces la metodología asume que entró a la red en la segunda estación. Dado que los códigos MAC de los datos estaban encriptados, no fue posible cuantificar este error usando dispositivos para hacer pruebas. Además, como el

código único de cada dispositivo cambiaba todos los días, tampoco fue posible hacer un análisis del mismo dispositivo en diferentes días.

Al estudiar las detecciones faltantes se debe considerar lo siguiente:

- Un dispositivo en su estación de origen entra al área de detección del WS, se puede desplazar por la estación, espera a que llegue el vehículo y luego sale en el vehículo del área de detección.
- Un dispositivo que pasa por una estación intermedia en su viaje dentro del vehículo entra y sale del área de detección del WS estando dentro del vehículo, el cual a su vez ayuda a bloquear las señales.
- Un dispositivo en su estación de destino entra en el vehículo al área de detección del WS, se baja del vehículo y luego sale caminando del área de detección.

Según este análisis cualitativo, la probabilidad de ser detectado en la estación de destino es mayor que una estación intermedia, y a su vez la probabilidad de ser detectado en alguna de las últimas dos es menor que la de ser detectado en la estación de origen. Una prueba interesante sería cuantificar estas probabilidades por estación, teniendo una gran masa de códigos MAC conocidos que hagan cierto viaje conocido.

Se sospecha que los resultados de la matriz OD serán mejores con redes más grandes ya que el dispositivo puede ser detectado en más estaciones y así detectar mejor su patrón de viajes. Además, tener grupos de estaciones en vez del detalle de cada estación puede ayudar a detectar los macro flujos de personas en desmedro de saber exactamente qué estación usa como origen o destino cada usuario. El nivel de agrupación depende de la necesidad por la que se está obteniendo la matriz OD, pero no es necesario hacer cambios en la metodología sino que solo agrupar la información de los WS de forma distinta.

### **5.3. Carga por arco.**

Al igual que para la matriz OD, se obtuvo la carga para cada arco de la red de la prueba piloto cada 15 minutos para toda la semana laboral analizada. En esta sección se muestran los resultados de la carga por arco del martes entre las 17:00 y las 20:00. Para construir la

carga de todo el periodo en base a las cargas de 15 minutos simplemente se sumó la cantidad de personas durante todo el periodo de análisis.

Se obtuvieron las cargas para el mismo periodo horario de un día martes de 17 meses antes que el martes de la metodología de esta tesis, obtenidas con la metodología que usa el operador del STP basada en la SC para comparar órdenes de magnitud. A continuación, en la figura 5-9 se muestran los resultados de ambas estimaciones de las cargas por arco en ambos sentidos de operación:

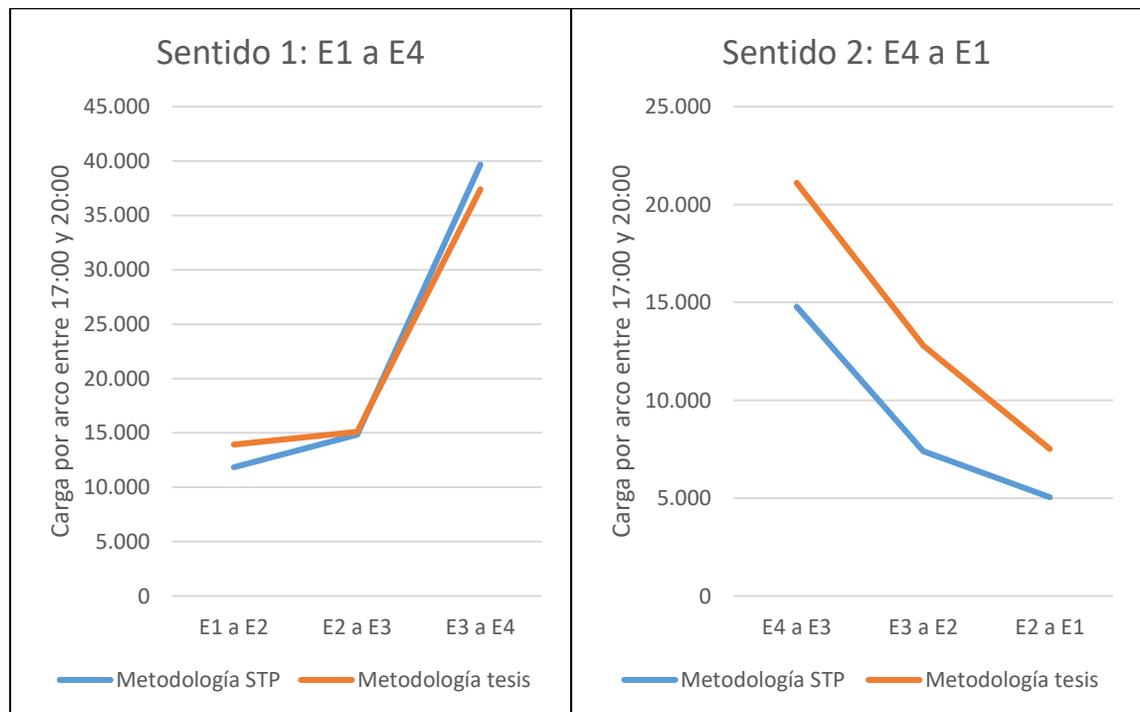


Figura 5-9: Carga por arco estimada con la metodología de esta tesis y con la del operador del STP para el periodo un día martes entre 17:00 y 20:00.

Existen diferencias pero no tan notorias como en la matriz OD. Esto se debe a que la estimación de la carga por arco es más estable que la matriz OD porque independiente de que el dispositivo no envíe una señal en su estación de origen o destino, aun así podría enviar señales en otras etapas de su viaje que irán completando el resto de las etapas donde

no se envió señales. Así, se cargan los arcos por los que tuvo que cruzar el dispositivo entre la primera y última detección de su viaje.

Las estimaciones en el sentido desde E1 a E4 (sentido 1) son más similares entre ellas que las del sentido opuesto (sentido 2). Todos los usuarios que circulan en el sentido 1 tienen un ponderador correspondiente a su estación real de origen (o al menos de una de las aledañas). Por otro lado, los usuarios que circulan en el sentido 2 entraron a la red del STP por cualquier otra estación fuera de la prueba piloto, pero se asume que entraron por E4. Como se mencionó, el ponderador de E4 no se calcula ya que no tiene sentido, y a esos usuarios se les asigna el ponderador de E3 por ser la estación más cercana.

Por lo tanto, en una implementación completa de esta metodología donde todas las estaciones tengan un WS, el cálculo de cada ponderador tendrá sentido y se espera una mejor estimación de los resultados.

#### **5.4. Tiempos de permanencia en estación.**

Se estimó el tiempo de permanencia en la estación de origen de cada viaje usando la metodología descrita en la sección 3.3.1. En la figura 5-10 se pueden ver los valores para las tres estaciones por separado (E1, E2 y E3) para el lunes de la semana analizada (se obtuvo los resultados para otros días y fueron similares). Para este análisis no se consideraron los tiempos de permanencia mayores a 10 minutos ya que, dada la alta frecuencia con que opera el STP, probablemente no corresponde a un usuario del sistema.

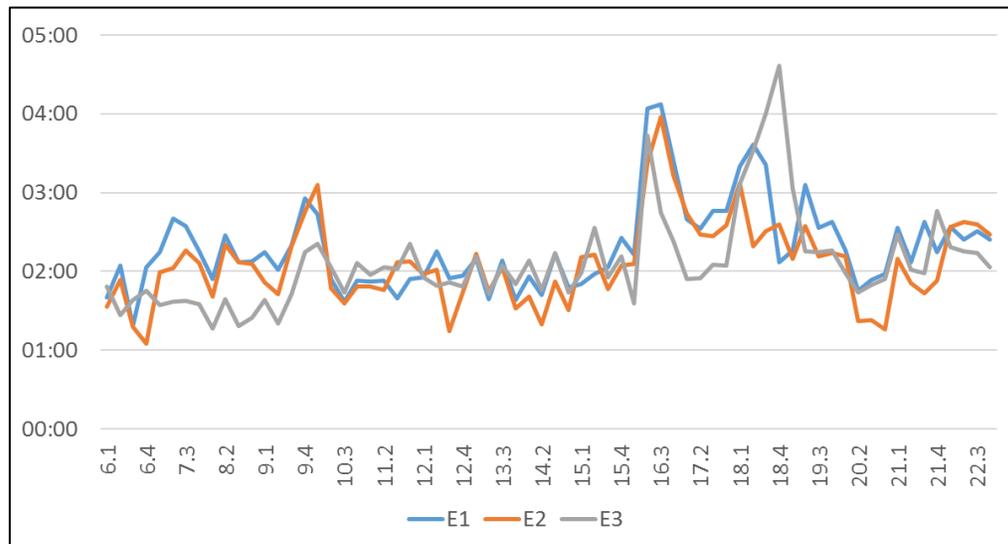


Figura 5-10: Tiempo de permanencia en minutos estimado con los WS para el lunes.

Se puede ver en las curvas que en los periodos de alta afluencia de usuarios a las estaciones (revisar curvas de capítulo 5.1.) el tiempo de permanencia aumenta, aunque los valores son muy variables. El aumento probablemente se produce por tres razones: usuarios que no alcanzan a abordar el primer vehículo ya que viene muy lleno; usuarios que no abordan el primer vehículo para lograr ser los primeros en abordar el siguiente vehículo, y así conseguir un asiento; usuarios que se demoran en entrar a la estación debido al exceso de personas.

Un resultado interesante que se ve aquí es que la estación E3, para los periodos punta mañana y tarde (excepto entre 18.1 y 20.1), tiene un tiempo de permanencia más bajo que las otras dos estaciones. Esto es extraño ya que E3 es la estación con mayores afluencias dentro de estas tres, por lo que el colapso y congestión de usuarios debiese ser mayor. Además, los usuarios de E1 se suben cuando el vehículo está vacío ya que es estación terminal y por lo tanto es más fácil abordar.

Sin embargo, durante dos periodos del día (entre 7:00 y 10:00 y luego entre 16:00 y 20:00) la estación E3 se convierte en terminal para algunos vehículos. Es decir, algunos vehículos no operan en la red completa, sino que llegan hasta E3 y se devuelven. Esta operación tipo

bucle opera con el doble de frecuencia que las estaciones fuera del bucle, por lo que las estaciones E1 y E2 reciben la mitad de frecuencia de vehículos que E3 en los horarios punta y por lo tanto tienen tiempos de permanencia más altos que E3.

La razón de por qué el argumento anterior no se cumple para el periodo entre 18.1 y 20.1 que sí está dentro del bucle puede ser que E3 colapsa en esos horarios. La cantidad de usuarios que entra al STP por E3 en esos horarios es tan grande que tienen que esperar más de un vehículo para poder usar el STP. Este efecto hace que su tiempo de permanencia aumente.

El tiempo de permanencia estimado a través de los WS puede subestimar los tiempos reales de espera, dependiendo de la cantidad de señales que envíen los dispositivos. El mejor caso es que un dispositivo entre al área de detección del WS e inmediatamente envíe una señal, y luego de subir al vehículo y justo antes de abandonar el área de detección envíe otra señal. En el otro extremo, está el dispositivo que envía una única señal y por lo tanto su tiempo de permanencia es cero. El problema es que no es posible reconocer si un dispositivo envió pocas señales porque el dispositivo está programado así o porque estuvo poco tiempo dentro del área de detección.

### **5.5. Viajes diarios por dispositivo.**

Como se mencionó antes, los WS usados en este proyecto tienen un sistema de encriptación para que los investigadores no vean el código MAC de cada usuario, para no atentar contra la privacidad de los usuarios. El código encriptado es único para cada dispositivo durante las 24 horas del día, pero a medianoche el mismo dispositivo adquiere un nuevo código encriptado, logrando que sea imposible hacer un seguimiento a los dispositivos durante varios días, pero sí a nivel de día.

Este último análisis reveló que, a lo largo de la semana laboral analizada la proporción de las frecuencias de viajes diarios de los dispositivos es muy constante. En la tabla 5-2 se muestran los resultados. La primera columna indica la cantidad de veces que cierto dispositivo viajó durante el día, mientras que el resto de las columnas muestra la proporción de dispositivos de ese día que viajaron cierta cantidad de veces al día en la red estudiada.

Tabla 5-2: Proporción de frecuencia de viajes para una semana laboral.

VIAJES DIARIOS	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES
<b>1</b>	78.5%	77.8%	77.8%	78.4%	79.1%
<b>2</b>	20.5%	21.1%	21.1%	20.5%	19.8%
<b>3</b>	0.7%	0.8%	0.8%	0.8%	0.9%
<b>4+</b>	0.3%	0.3%	0.3%	0.3%	0.3%

Se puede ver que la mayoría de los dispositivos viaja una única vez en el día, lo que es sorprendente porque en el conocimiento colectivo se habla de que las personas viajan dos veces al día: para ir desde su hogar a hacer una actividad (estudiar, trabajar, comprar) y luego para volver a su hogar, aunque podrían hacerlo en otro modo de transporte o con su *smartphone* apagado. Estos viajes representan alrededor del 21% del total de dispositivos. Los viajes únicos son la gran mayoría y se podría pensar que los usuarios de este STP son usuarios con gran variabilidad en su forma de transportarse y en sus actividades diarias.

### 5.6. Días especiales.

Un día jueves a las 18:00 horas ocurrió un evento deportivo masivo que se esperaba que fuese visto por muchas personas a través de la televisión, ya sea en casa o en bares. En la figura 5-11 se muestran las afluencias agrupadas a las tres estaciones analizadas (E1, E2 y E3) para 6 días jueves consecutivos. La curva del día del evento deportivo es la curva del jueves 6.

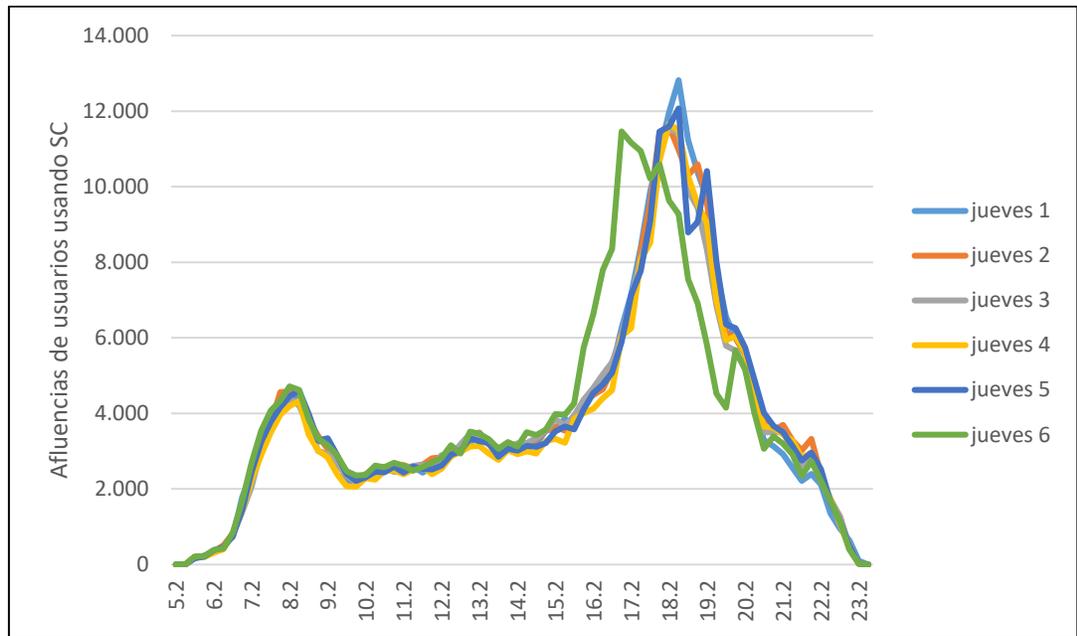


Figura 5-11: Afluencias de SC agrupadas de E1, E2 y E3 de 6 días jueves consecutivos.

Es evidente que el día del evento los usuarios comenzaron a usar la red más temprano para alcanzar a llegar a ver el evento a las 18:00. La punta tarde se adelanta y además se alarga, disminuyendo su magnitud máxima. Para este tipo de días, la frecuencia de los vehículos se debe ajustar a la nueva hora punta para que la red no colapse y para que los usuarios perciban que las autoridades se preocupan activamente de mejorar el sistema. Esta información ayuda a predecir qué podría ocurrir en el futuro en eventos similares.

Se sospecha que los pares OD en este día también son diferentes ya que muchas personas en vez de ir a sus casas van a ir a algún bar o donde un amigo, pero la red de la prueba piloto es muy pequeña para demostrarlo. Sí se observan variaciones, pero éstas son menores.

## **6. CONCLUSIONES Y FUTURAS APLICACIONES.**

En este el último capítulo primero se muestran las principales conclusiones encontradas. Luego, se describen futuras aplicaciones que tiene la metodología desarrollada en esta investigación.

### **6.1. Conclusiones**

Esta investigación ha obtenido interesantes resultados en cuanto al uso de nuevas tecnologías para el transporte para monitorear la operación de un STP y comprobar cómo afectan factores socioeconómicos a la penetración de *smartphone*. Estos resultados se obtuvieron gracias a la fusión de dos fuentes de datos tecnológicas. Ambas tienen muy alta penetración de uso en la población, son baratas de operar y se pueden obtener en tiempo real sin interferir con el servicio que el STP entrega a los usuarios.

En este trabajo se desarrolló una forma sencilla de fusionar dos fuentes de datos para obtener resultados muy valiosos para un operador y planificador de un STP. Los principales supuestos que exige la metodología están presentes o podrían estarlo con relativa facilidad en varios STP del mundo. Para implementar masivamente esta metodología no se requieren grandes cambios por lo que se cree que tiene gran potencial para ser implementada en otros STP.

La metodología desarrollada fue capaz de obtener la matriz OD, cargas por arco y otros resultados valiosos. Esta metodología está diseñada para obtener dichos resultados en tiempo real si es que las fuentes de datos los entregan en tiempo real. Las operaciones de los STP son variables y dinámicas, por lo que urge la necesidad de tomar decisiones y hacer gestión en tiempo real utilizando información cuantitativa de calidad.

La metodología fue probada en una prueba piloto en un STP. La prueba entregó resultados importantes para la investigación ya que se pudo comprobar que la metodología obtiene los resultados que se buscaban y además lo hace en un tiempo computacional razonable. Dado que pocas estaciones fueron equipadas con WS, los resultados obtenidos se deben tratar con cuidado y entendiendo los supuestos de esta prueba específica. La tecnología Wifi, a pesar

de ser variable y a veces difícil de entender su comportamiento, puede entregar resultados muy valiosos de la operación de un STP.

Una limitación de la metodología es que la expansión de patrones de viajes Wifi desarrollada no considera el destino de los usuarios. El ponderador se calcula en base al periodo y estación de origen de cada usuario, pero asume que la probabilidad de elegir un destino no depende del tipo de usuario. Si todos los usuarios eligieran sus destinos de la misma forma, entonces no hay problema. Pero, existen grupos de usuarios como estudiantes viajando a una universidad que pueden tener mayor penetración Wifi que el resto de la población, por lo tanto se estarán sobreestimando los viajes a esa estación de destino.

Otra limitación de trabajar con tecnología Wifi es que hay estaciones en las que un dispositivo puede pasar sin ser detectado. Si es una estación intermedia del viaje, entonces este puede ser completado sin problemas con la excepción de que haya múltiples rutas entre la estación anterior y posterior a la no detección, ya que no se sabrá qué ruta tomó el usuario. La situación cobra más relevancia cuando la no detección ocurre en la estación de origen o destino del dispositivo ya que no hay forma de completar el viaje de manera confiable. La ubicación óptima del WS en la estación o la cantidad de WS necesarios para cada estación es una variable que puede impactar la cantidad de datos recibidos.

Trabajar con tecnología Wifi tiene sesgos que fueron estudiados en esta investigación. Los sesgos pueden ser significativos y no deben menospreciarse, sobre todo si lo que se busca es ayudar a las personas más vulnerables de la sociedad (sector de bajo ingreso, adultos mayores o bajo nivel de educación). La penetración de *smartphone* varía entre el 30% y 70%, dependiendo del nivel de ingreso y edad de los usuarios, así como también de la ubicación del WS dentro de la estación.

El hecho de que las penetraciones encontradas sean tan variables justifican aún más el uso de una metodología dinámica de expansión como la desarrollada en este trabajo. Por otro lado, la curva de penetración cada 15 minutos tiene una forma definida, pero con cierta componente que hace que se vea ruidosa. Por ello se debe buscar, para cada caso de implementación, un largo de periodo  $p$  tal que las curvas tengan una forma lo

suficientemente definida pero que no agrupe demasiados datos como para perder visualización de ciertos efectos, sobre todo en las horas punta.

Utilizar dos fuentes de datos tecnológicas de forma conjunta logra entregar información clave de la operación de un STP. Al utilizar dos fuentes de datos tecnológicas de forma conjunta se pueden potenciar sus beneficios y, en casos como esta investigación, resolver algunas de sus limitaciones como la dificultad de la SC para estimar el destino de cada usuario o el hecho que los WS solo tomen una muestra de los usuarios.

## **6.2. Futuras Aplicaciones.**

Esta investigación abre puertas en distintos frentes del mundo académico. En esta sección se mencionarán las que se consideró más relevante para trabajos futuros.

El ponderador se estima según las afluencias de usuarios obtenido a través de la SC y las afluencias de dispositivos obtenidas con los WS. Esto quiere decir que el viaje de un usuario será ponderado según la estación y momento del día en el que se inició el viaje. Más adelante se podrían armar ponderadores que además dependan del destino escogido por el dispositivo. Como no se tiene información de las salidas de las SC del STP, se tendrían que estimar conteos de salidas de usuarios de cada estación.

Sería interesante realizar un estudio más acabado del sesgo que puede existir al trabajar con tecnología Wifi. Esta está aumentando su penetración en el mundo rápidamente por lo que un estudio sobre sus usos en la población puede quedar obsoleto con la misma rapidez. Aun así, se podrían hacer estudios locales con muestras de usuarios para determinar cuál es la penetración real y además entender cómo se comporta según el nivel de ingreso, educación, edad u otra variable socioeconómica.

Una prueba piloto en una red más extensa es necesaria para corroborar resultados mostrados en este trabajo. La prueba piloto en solo cuatro estaciones hace que los resultados de los dispositivos que viajan desde la estación que conecta con el resto de la red (E4) hacia la red de la prueba piloto tengan poca validez. Estos dispositivos entraron al STP

por cualquier lugar de la red fuera de la prueba piloto y por ende el ponderador de este grupo de usuarios tiene mucha variabilidad.

Además, se podría validar alguno de los resultados obtenidos en este trabajo con mediciones paralelas en terreno. Los resultados de la matriz OD y la carga por arco fueron validados con resultados obtenidos con otras metodologías teóricas (basada solo en la SC) obtenidos un año y medio antes de la prueba piloto de esta tesis. Más adelante, se podrían validar los resultados con información detallada obtenida con conteos en el mismo periodo del que se obtengan los datos de los WS y las SC.

La metodología funciona sin tener que hacer cambios relevantes en varios STP presentes hoy en el mundo como BRT, metro, tren, entre otros. Estos STP generalmente están aislados del resto de los modos de transporte y se debe validar la entrada al sistema en una estación. Para otros STP donde la validación se hace en el vehículo, se puede implementar esta metodología siempre y cuando se tenga la ubicación exacta de los vehículos en cada momento. Además, se debería tener información de los recorridos de los vehículos a través de la red si es que no todos siguen la misma ruta fija.

Como se mencionó en los capítulos anteriores, la metodología desarrollada asume que no se detectan las señales de dispositivos viajando en otros medios de transporte en una red paralela a la del STP. En la prueba piloto, esto se cumplía. Si se quisiera implementar esta metodología en un STP donde existan detecciones de dispositivos que viajen en vehículos particulares o peatones se necesita un algoritmo para filtrar estos viajes y no considerarlos dentro del STP.

Un enfoque a este filtro podría ser la velocidad de traslado entre estaciones. Asumiendo que el STP está aislado de la congestión que podría existir en la red paralela de vehículos, se podría calibrar un modelo que determine la velocidad de viaje entre estaciones del STP y la de los vehículos externos para así discriminar si un dispositivo viaja dentro del STP según su velocidad entre estaciones.

Otra forma más sofisticada de filtrar los datos sería tener la información del instante de llegada de vehículos del STP a cada estación. Ningún dispositivo que la metodología asuma

que está viajando dentro del STP puede llegar a una estación si ningún vehículo ha llegado a esa misma estación en un instante cercano. Esto implica una tercera fuente de información (aparte de las detecciones de los WS y los conteos de las SC) pero se podría estimar un modelo probabilístico que incluya la diferencia entre el tiempo de llegada del vehículo a la estación y el instante de la detección en esa estación de los dispositivos que supuestamente viajan en ese vehículo.

Tener la posición de cada vehículo ayudaría a tener una mejor estimación del tiempo de permanencia. Si todos los usuarios suben al primer vehículo que pasa, entonces se podría saber con exactitud en qué momento cada dispositivo deja la estación, en vez de tener que estimar ese instante usando su última detección. Así, el tiempo de permanencia sería el tiempo entre la primera detección en su estación de origen y la salida del vehículo de esa estación. Incluso si no sube al primer vehículo, se puede detectar a qué vehículo se subió cada dispositivo (usando información de detecciones en estaciones posteriores) y así estimar el tiempo de permanencia en su estación de origen.

Por último, se podría pensar en un enfoque distinto a esta metodología donde el WS no esté ubicado en la estación, sino que en el vehículo. Así, se puede hacer un seguimiento a cada dispositivo durante todo el tiempo que permanezca dentro del vehículo, y luego identificar su ruta dependiendo de la ubicación del vehículo (otra fuente de información). Esta forma de manejar los datos es más compleja y más susceptible a fallas, pero tiene la ventaja de estimar bien múltiples rutas en redes muy complejas. Por otro lado, se pierde la capacidad de estimar tiempos de permanencia en estación.

## BIBLIOGRAFIA

- Alexander, L., Jiang, S., Murga, M., & González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation research part c: emerging technologies*, 58, 240-250.
- Abbott-Jard, M., Shah, H., & Bhaskar, A. (2013). Empirical evaluation of Bluetooth and Wifi scanning for road transport. En *36th Australasian Transport Research Forum (ATRF)*. Brisbane, Queensland, Australia.
- Abedi, N., Bhaskar, A., & Chung, E. (2013). Bluetooth and Wi-Fi MAC address based crowd data collection and monitoring: benefits, challenges and enhancement. En *36th Australasian Transport Research Forum (ATRF)*. Brisbane, Queensland, Australia.
- Abedi, N., Bhaskar, A., & Chung, E. (2014). Tracking spatio-temporal movement of human in terms of space utilization using Media-Access-Control address data. *Applied Geography*, 51, 72-81.
- Abedi, N., Bhaskar, A., Chung, E., & Miska, M. (2015). Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses. *Transportation Research Part C: Emerging Technologies*, 60, 124-141.
- Barceló, J., Montero, L., Bullejos, M., Serch, O., & Carmona, C. (2012). A kalman filter approach for the estimation of time dependent OD matrices exploiting bluetooth traffic data collection. *TRB 91st Annual Meeting Compendium of Papers DVD* (pp. 1-16).
- Caceres, N., Wideberg, J. P., & Benitez, F. G. (2007). Deriving origin destination data from a mobile phone network. *IET Intelligent Transport Systems*, 1(1), 15-26.
- Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., & Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation research part C: emerging technologies*, 26, 301-313.
- Danalet, A., Bierlaire, M., & Farooq, B. (2012). Estimating Pedestrian Destinations using Traces from WiFi Infrastructures. *Pedestrian and Evacuation Dynamics 2012* (pp. 1341-1352). Springer International Publishing.
- Danalet, A., Farooq, B., & Bierlaire, M. (2014). A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures. *Transportation Research Part C: Emerging Technologies*, 44, 146-170.
- Harwood, M. (2009). *CompTIA network+ N10-004 exam cram*. Que Publishing Company.
- Kostakos, V. (2008). Towards sustainable transport: wireless detection of passenger trips on public transport buses. *arXiv preprint arXiv:0806.0874*.
- Michau, G., Nantes, A., & Chung, E. (2013). Towards the retrieval of accurate OD matrices from Bluetooth data: lessons learned from 2 years of data. En *36th Australasian Transport Research Forum (ATRF)*. Brisbane, Queensland, Australia.

- Munizaga, M. A., & Palma, C. (2012). Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*, 24, 9-18.
- Munizaga, M., Devillaine, F., Navarrete, C., & Silva, D. (2014). Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies*, 44, 70-79.
- Musa, A. B. M., & Eriksson, J. (2012, November). Tracking unmodified *smartphones* using wi-fi monitors. En *Proceedings of the 10th ACM conference on embedded network sensor systems* (pp. 281-294). ACM.
- Pelletier, M. P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4), 557-568.
- Pew Research Center. (2016, February 22). *Smartphone* ownership and internet usage continues to climb in emerging economies. Recuperado de: <http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/>
- Pew Research Center. (2014, February 14). Emerging nations embrace internet mobile technology. <http://www.pewglobal.org/2014/02/13/emerging-nations-embrace-internet-mobile-technology/>

**ANEXOS**

**ANEXO A: DIFERENCIAS DE PENETRACIÓN DE SMARTPHONE POR PAÍS  
Y CARACTERÍSTICAS SOCIOECONÓMICAS.**

Country	Total %	Age			Education			Income		
		18-34 %	35+ %	Diff	Less %	More %	Diff	Lower %	Higher %	Diff
South Korea	88	100	83	17	80	95	15	79	95	16
Australia	77	95	70	25	67	85	18	62	88	26
Israel	74	87	67	20	68	80	12	63	83	20
United States	72	92	65	27	59	81	22	64	84	20
Spain	71	91	64	27	63	85	22	63	81	18
United Kingdom	68	91	60	31	60	80	20	59	84	25
Canada	67	94	58	36	58	73	15	61	80	19
Chile	65	86	50	36	16	74	58	46	79	33
Malaysia	65	88	46	42	26	79	53	44	75	31
Germany	60	92	50	42	48	68	20	49	68	19
Italy	60	88	52	36	56	83	27	46	74	28
Turkey	59	81	39	42	34	86	52	**	**	**
China	58	85	43	42	40	86	46	49	75	26
Palestinian territories	57	73	39	34	38	74	36	47	66	19
Lebanon	52	74	37	37	17	79	62	20	85	65
Jordan	51	60	41	19	32	72	40	34	64	30
France	49	85	35	50	38	69	31	37	59	22
Argentina	48	71	35	36	38	72	34	28	53	25
Russia	45	76	29	47	**	**	**	25	53	28
Venezuela	45	60	35	25	33	65	32	**	**	**
Brazil	41	61	26	35	21	64	43	25	54	29
Poland	41	75	25	50	12	47	35	27	52	25
Japan	39	77	31	46	31	52	21	21	54	33
South Africa	37	46	30	16	21	54	33	18	50	32
Mexico	35	54	22	32	17	67	50	23	50	27
Vietnam	35	56	17	39	23	53	30	28	51	23
Nigeria	28	39	13	26	6	38	32	20	36	16
Ukraine	27	56	13	43	6	28	22	13	38	25
Kenya	26	34	14	20	10	49	39	13	37	24
Peru	25	41	15	26	5	37	32	9	31	22
Philippines	22	31	14	17	9	31	22	11	31	20
Ghana	21	27	15	12	11	52	41	7	26	19
Indonesia	21	39	7	32	7	42	35	10	31	21
Senegal	19	26	12	14	12	57	45	12	27	15
India	17	27	9	18	7	29	22	7	22	15
Burkina Faso	14	17	10	7	9	52	43	12	29	17
Pakistan	11	13	7	6	5	22	17	5	14	9
Tanzania	11	14	6	8	**	**	**	7	14	7
Ethiopia	4	**	**	**	3	17	14	2	13	11
Uganda	4	6	2	4	**	**	**	3	6	3

## **ANEXO B: PROGRAMA EN C# PARA CONTROLAR EL FLUJO DE INFORMACIÓN.**

Para manejar el ingreso de datos a pgAdmin III, se escribió un programa en el lenguaje C# para ser ejecutado en Microsoft Visual Studio 2010. Este programa se encarga de ir ingresando las detecciones conjuntas de todos los WS de la red en orden temporal, y luego ingresar la información de las afluencias de SC. Este programa también informa sobre el progreso de la información dentro de pgAdmin III, por lo que ayuda a ver el nivel de saturación del programa y cuándo se producen problemas.

Dado que cada detección Wifi tiene su estampilla de tiempo real, y las afluencias agrupadas de SC están agrupadas también en el mismo tiempo coordinado de las detecciones, se puede simular que la información es ingresada en tiempo real con facilidad. Basta ordenar las detecciones e ingresarlas en el orden que corresponde, y al mismo tiempo hacer que la información de las afluencias de SC se ingrese usando las detecciones Wifi como controlador de un reloj virtual.

El programa controlador del flujo de información también tiene dos funciones menores. La primera, es realizar filtros a la base de datos cruda de detecciones antes de que sean insertadas al programa de la metodología. Esto con el único fin de mejorar el tiempo de ejecución de dicho programa. El filtro elimina todas las detecciones repetidas para cada código MAC en cada WS en cada segundo. Este filtro no pierde información ya que un dispositivo que tenga más de una detección en un segundo es información redundante para los efectos de este trabajo. En términos generales, se logra reducir el número de detecciones totales a un 30% del valor original, mejorando considerablemente el tiempo de ejecución del programa.

La segunda función menor es crear fácilmente la tabla Pares\_OD que define todos los movimientos necesarios para ir de una estación a otra. Esta tabla se usa para cargar los arcos que corresponda cuando un dispositivo viaja de una estación a otra. La tabla tiene todos los arcos que conectan cada par de estaciones, por lo que puede llegar a ser una tabla bastante grande de construir si es que son muchas estaciones. Por eso, se programó un

método para hacer esta tarea de forma automática entregando solamente una lista de las estaciones ordenadas.