



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA CIVIL**

**GENERACIÓN DE DATOS DE PATRONES DE VIAJE A PARTIR DE
TRANSACCIONES BIP**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL

PAMELA ANTONIETA MORA PAREDES

**PROFESOR GUÍA:
MARCELA MUNIZAGA MUÑOZ**

**MIEMBROS DE LA COMISIÓN:
CAROLINA PALMA ALVARADO
ROBERTO COMINETTI COTII-COMETTI**

**SANTIAGO DE CHILE
ENERO 2010**

Resumen de la memoria para optar al título
de Ingeniero Civil

Por: Pamela Antonieta Mora Paredes

Fecha: 28 de Diciembre de 2009

Profesor Guía: Sra. Marcela Munizaga M.

“GENERACIÓN DE DATOS DE PATRONES DE VIAJE A PARTIR DE TRANSACCIONES BIP”

La información que se puede extraer, deducir e interpretar a partir de las transacciones realizadas con las tarjetas BIP (*smart card*) en conjunto con los datos entregados por el sistema de posicionamiento global (GPS) existente en los buses del sistema de transporte de la ciudad de Santiago son una nueva herramienta muy potente que provee gran cantidad de información que antes era impensado tener.

El objetivo principal del presente trabajo de título es establecer una metodología sólida que sea capaz de realizar la extracción y el procesamiento de la información obtenida a partir de las transacciones BIP del sistema de transporte público de Santiago, proporcionando exámenes detallados de cada objeto dentro del sistema y determinando la correcta estimación del destino del viaje que comienza con cada transacción.

Se revisaron experiencias mundiales que utilizan *smart cards* como medio de pago además de algunas metodologías de extracción de información y tratamientos de datos, identificando y adaptando (de acuerdo a las características del caso de Santiago) diversos supuestos que apuntan a la estimación de bajada, detección de errores, recuperación de información y generación de matrices origen-destino.

Para la información disponible se determinó que las principales fortalezas radican en los altos porcentajes de información correcta y útil que presentan las bases de datos, mostrando un 97% de información completa para la ubicación de transacciones (lo que implica la posibilidad de tener dicho porcentaje de estimaciones de punto de bajada), y un 87% de registros de posición útiles que permiten el seguimiento de los usuarios.

A partir de la sensibilidad con el problema adquirida al realizar un exhaustivo análisis exploratorio sobre una muestra pequeña de tarjetas, se llega al resultado final donde se plantea una metodología de 5 procesos principales que consideran la interacción entre los elementos presente en el sistema y que permite conocer los destinos de las etapas así como la composición de los viajes y características de éste. Destaca el proceso de estimación de bajada, el cual según el sitio donde se realiza la transacción (bus, metro ó zona paga) establece el procedimiento a seguir considerando la dinámica del sistema, e incorporando las características del modo utilizado, además en este proceso se enriquece los métodos encontrados en la literatura, incorporando en la elección del usuario la variable tiempo de viaje además de la mínima distancia, para las etapas de viaje realizadas en bus. Los procesos establecidos enriquecen la información disponible generando herramientas que facilitan el posterior análisis de ella.

Índice de contenidos

1	Introducción	4
1.1	Aspectos Generales	4
1.2	Objetivos del Estudio	5
1.3	Estructura del Documento	5
2	Revisión Bibliográfica	7
2.1	Experiencias a nivel mundial que utilizan tarjetas inteligentes	7
2.2	Metodologías para extracción de información	10
2.3	Síntesis	21
3	Descripción y Análisis de Datos	24
3.1	Descripción de la Información	24
3.2	Análisis de Datos	29
3.2.1	Tabla Transacciones BIP	29
3.2.2	Tabla Posicionamiento	50
3.2.3	Otras	53
3.3	Síntesis	53
4	Metodología para la Extracción y el Procesamiento de la Información: Exploración, procedimiento y criterios propuestos	55
4.1	Análisis Exploratorio: Estudio de una Muestra Pequeña de Tarjetas	55
4.1.1	Extracción y procesamiento de la información	56
4.1.2	Resultados obtenidos	63
4.2	Metodología Propuesta	72
4.2.1	Estimación de bajada	74
4.2.2	Recuperación y deducción de información	87
4.2.3	Cálculos de tiempos	90
4.2.4	Identificación de etapa de viaje y viaje	92
4.2.5	Información relevante	93
5	Conclusiones y líneas futuras de investigación	95
5.1	Comentarios y Conclusiones	95
5.2	Líneas Futuras de investigación	98

ANEXOS

Anexo A: Información Complementaria Capítulo 3

Anexo B: Detalle Muestra Pequeña de Usuarios.

Anexo C: Glosario de términos utilizados en Diagramas

Índice de figuras

Figura 2-1: Modelo de identificación de paradero de bajada.....	17
Figura 3-1: Interacción entre sistema de recolección automática y sistema de posicionamiento global, para Transantiago.....	24
Figura 3-2: Bases de datos y campos relevantes para el estudio.....	26
Figura 3-3: Distribución de transacciones según tipo de servicio.....	29
Figura 3-4: Distribución de transacciones según lugar donde se realizó validación.....	29
Figura 3-5: Transacciones por día.....	30
Figura 3-6: Distribución horaria de las transacciones en día laboral.....	34
Figura 3-7: Distribución horaria de las transacciones en día laboral según pago y tipo de tarjeta.....	35
Figura 3-8: Distribución horaria de las transacciones en día laboral por modo.....	35
Figura 3-9: Distribución horaria de las transacciones en día sábado.....	37
Figura 3-10: Distribución horaria de las transacciones en día sábado según pago y tipo de tarjeta.....	38
Figura 3-11: Distribución horaria de las transacciones en día sábado por modo.....	38
Figura 3-12: Distribución horaria de las transacciones en día domingo.....	40
Figura 3-13: Distribución horaria de las transacciones en día domingo según pago y tipo de tarjeta.....	40
Figura 3-14: Distribución horaria de las transacciones en día domingo por modo.....	41
Figura 3-15: Transacciones sin posición por día.....	45
Figura 3-16: Transacciones sin posición a lo largo del día, para todos los días.....	46
Figura 4-1: Transacciones según lugar donde se realizaron.....	64
Figura 4-2: Rango de valores distancia Δd	65
Figura 4-3: Distribución según motivo, de etapas sin estimación de punto de bajada ...	66
Figura 4-4: Distribución según lugar donde se realizó la transacción, de clasificación de punto de bajadas de las etapas.....	68
Figura 4-5: Distribución de tiempos de viaje para muestra usuarios.....	69
Figura 4-6: Distribución de tiempos entre etapas Δt para muestra usuarios.....	70
Figura 4-7: Distribución de tiempos de entre transacciones para muestra usuarios.....	71
Figura 4-8: Metodología propuesta.....	73
Figura 4-9: Modelo de estimación de bajada.....	75

Figura 4-10: Diagrama general de tratamiento de la información para la estimación del punto de bajada.....	76
Figura 4-11: Esquema búsqueda de punto gps que minimiza el tiempo generalizado...	78
Figura 4-12: Diagrama de la estimación del punto de bajada, para transacciones realizadas en bus.	80
Figura 4-13: Diagrama de la estimación del punto de bajada, para transacciones realizadas en metro.	83
Figura 4-14: Diagrama de la estimación del punto de bajada, para transacciones realizadas en zona paga.	86
Figura 4-15: Diagrama global para deducción y recuperación de información.....	88
Figura 4-16: Proceso de recuperación y deducción de información.....	89
Figura 4-17: Diagrama para determinar la existencia de transbordo.....	91
Figura 4-18: Diagrama para identificar etapa de viaje de viaje.	92

1 Introducción

1.1 Aspectos Generales

La información que se puede extraer, deducir e interpretar a partir de las transacciones realizadas con las tarjetas BIP, en conjunto con los datos entregados por el sistema de posicionamiento global (GPS) existente en los buses y el conocimiento de la operación del sistema de transporte de la ciudad de Santiago, son una herramienta nueva muy potente que provee gran cantidad de información que antes no era posible tener, resultando ser el mejor candidato para obtener información continua de los viajes.

La tarjeta BIP es una tarjeta inteligente (*smart card*) que recauda y administra las transacciones realizadas en el sistema de transporte público de la ciudad, permitiendo que el acceso a éste sea de una manera más rápida, eficiente y confiable. El sistema de transporte público de la ciudad de Santiago se ha caracterizado por incorporar satisfactoriamente el uso de la tarjeta BIP como medio oficial de pago, esta tarjeta es la única forma de pago que garantiza tarifa integrada entre buses y metro, lo cual lo distingue de otros sistemas a nivel mundial.

Existen estudios previos que desarrollan metodologías de extracción y procesamiento de información a partir de bases de datos obtenidas de sistemas que operan con tarjetas inteligentes, y que además cuentan con un sistema de información geográfica, en ellos se obtienen destinos de viajes, perfiles de carga, y se realizan análisis por segmentos de usuarios. Sin embargo, estos estudios son aplicados a sistemas de transporte de menor escala y complejidad en comparación con el sistema de transporte de la ciudad de Santiago, que presentan distintos sistemas de recolección de tarifas (más de una forma de pago), y diferentes características de operación del sistema.

Desarrollar una metodología que se aplique a la ciudad de Santiago, teniendo en cuenta las características propias del sistema, la magnitud de éste, la cobertura que posee, y la cantidad de viajes que se generan en él, resulta ser una herramienta valiosa para conocer cómo se están comportando los usuarios.

La información que se tiene es muy rica en cuanto a los productos que se puede conseguir a partir de ella. Con una serie de procesos que limpian y enriquecen la información obtenida del sistema, en conjunto con un análisis extenso sobre el comportamiento dinámico de las transacciones, es factible estimar o deducir destinos, tiempos de viaje, estructuras de viaje (combinación de modos y número de transbordos), patrones de viaje, horarios punta, perfiles de carga, comportamiento de tarjetas, identificación lugares de interés, etc.

Los resultados a obtener ayudan a modelar, reconocer y estimar patrones de viaje de los usuarios según períodos de tiempo dentro del día, de la semana, o según tipo de usuario (estudiante, adulto).

1.2 Objetivos del Estudio

El objetivo principal de este trabajo es establecer una metodología sólida que sea capaz de realizar la extracción y el procesamiento de la información obtenida a partir de las transacciones BIP del sistema de transporte público de la ciudad de Santiago, proporcionando exámenes detallados de cada objeto dentro del sistema.

Dentro de la metodología a plantear, lo fundamental es establecer la correcta estimación del destino del viaje que comienza con cada transacción, ya que a partir del conocimiento del origen y destino de cada viaje (posición y tiempo) se genera el posterior desarrollo de la metodología, permitiendo reconstruir los viajes y generar patrones de éstos, de manera de conocer el comportamiento de las tarjetas en el sistema, y ser una herramienta de apoyo para la planificación del sistema de transporte público.

La metodología debe reconocer e identificar los errores presentes en la información, estableciendo condiciones y criterios que permitan la recuperación o deducción de ella. Además debe diferenciar entre etapa de viaje y viaje propiamente tal, estableciendo criterios y parámetros que permitan identificar transbordo de actividades en el destino, obteniendo tiempos de viaje/etapa y transbordo.

Otro de los objetivos es identificar claramente las fortalezas y limitaciones que la información presenta, determinando los sesgos existentes. De manera de tener una idea real de los alcances de la herramienta que se está implementando.

1.3 Estructura del Documento

Para lograr los objetivos descritos anteriormente, esta memoria fue organizada en cinco capítulos.

En el capítulo 2 se realiza una revisión de experiencias mundiales que utilizan *smart cards* como medio de pago, identificando características propias de cada sistema. También se examina algunas metodologías de extracción de información y tratamientos de datos, que presentan diversos supuestos que apuntan a la estimación de bajada, detección de errores, recuperación de información y generación de matrices origen-destino. El objetivo de este capítulo es identificar aquellos supuestos que sean un aporte para el caso de Santiago, y reconocer las características particulares del sistema de transporte a estudiar.

En el capítulo 3 se describe y analiza cualitativa y cuantitativamente la información disponible. Como resultado de este análisis se identifican las fortalezas y debilidades de la información a utilizar.

En el capítulo 4 se presenta la metodología para la extracción y el procesamiento de la información así como la exploración, procedimiento y criterios propuestos establecidos en este estudio.

En este capítulo se realiza un estudio exhaustivo de una muestra pequeña de usuarios, donde se adquiere sensibilidad con el problema, se determina la forma en que debe ser procesada la información y los posibles errores de información que se pueden reconocer, además se identifican los pasos a seguir para un correcto análisis.

A partir del diagnóstico y análisis de la pequeña muestra, se establece la metodología para extraer y procesar la totalidad de la información. La cual establece un método de cinco etapas.

En el capítulo 5 se presentan las conclusiones finales con respecto a la metodología planteada y al trabajo realizado. A partir de la metodología propuesta se generan recomendaciones sobre futuras líneas de investigación, que ayudan a mejorar y hacer más robusta la metodología propuesta.

2 Revisión Bibliográfica

En este capítulo se sintetiza la revisión de documentos existentes en la literatura que plantean metodologías para la extracción de datos que se generan al utilizar las tarjetas inteligentes en un sistema u operador de transporte. También se examina el uso de estas tarjetas en el mundo, de manera de conocer similitudes y diferencias que se presentan con el caso estudiado. El objetivo de este capítulo es identificar aquellos aportes documentados en la literatura que pueden colaborar con la metodología que se desea determinar, y reconocer las características que singularizan el sistema de transporte de la ciudad de Santiago, y que deben estar presentes en dicha metodología.

El capítulo se divide en tres secciones. La primera resume las principales experiencias mundiales que cuentan con la implementación de la tarjeta inteligente, la segunda detalla las metodologías planteadas en la literatura para extraer información valiosa de las bases de datos que generan las transacciones realizadas con las *smart cards*, y la tercera es un resumen de los principales puntos encontrados que contribuyen en esta investigación.

2.1 Experiencias a nivel mundial que utilizan tarjetas inteligentes

La aparición en el mundo de las *smart cards* data de los años setenta, y desde que se patentaron han tenido un constante avance en relación a la tecnología que usan así como en el ámbito donde se aplican. En el área de transporte comenzaron a utilizarse a fines de los años noventa como medio de pago en algunas ciudades del mundo como Washington¹ y Tokio².

En los últimos años el uso de *smart cards* ha comenzado a masificarse sumándose Londres³, algunos operadores de transporte público de San Francisco⁴ (Buneman, 1984), Portland (Furth et al., 2006), New York (Barry et al., 2002), Chicago (Zhao et al., 2007), Netherlands (Muller y Furth, 2001; Furth et al., 2006) Changchun (China) (Lianfu et al., 2007) y *Société de Transport de l'Outaouais* (Gatineau, Quebec, Canadá) (Trépanier et al., 2007; Chapleau y Chu, 2007; Chapleau y Chu, 2007; Chapleau et al., 2007; Chapleau et al., 2008) y Santiago (Chile)⁵, entre otros.

En estas ciudades (excepto Santiago) la tarjeta inteligente es usada como un medio de pago parcial, con tarifa rebajada en muchos casos, tarifa normal en otros o

1 Información obtenida de <<https://www.smartrip.com/rcsc.html>> y <<http://www.wmata.com/>>

2 Información obtenida de <<http://www.jreast.co.jp/e/development/story/>>, <http://www.jreast.co.jp/e/info_counters>, <<http://www.pasmo.co.jp/en/index.html>>,

3 Documento Transport for London: Oyster Factsheet, Julio 2009. Obtenido de <<https://www.tfl.gov.uk/assets/downloads/corporate>>

4 Información obtenida < <http://www.bart.gov/tickets/index.aspx> >

5 Información obtenida < <https://www.transantiago.cl> >

abonos de transporte. Ésta característica es la principal diferencia con lo implementado en Santiago, donde la tarjeta BIP es el único medio de pago para acceder al sistema de transporte público de la ciudad.

A continuación se presenta un resumen general de las experiencias en que se conoce algún tipo de información del funcionamiento de las *smart cards*.

En Washington las tarjetas inteligentes empezaron a operar como medio de pago para trenes y buses a fines de los años noventa. Esta tarjeta es de uso personal e intransferible y existe la posibilidad de recargarla a bordo de los buses y en las estaciones de metro además de lugares establecidos para ello. La implementación de la *SmarTrip* se enmarca en un proyecto de transporte multipropósito que integra transporte público y estacionamientos enfocado en la integración regional¹.

En Japón el año 2001 las líneas de trenes y metro que operaba la compañía *JR East* introdujeron el uso de la tarjeta prepago *SUICA (smart card)* en las ciudades de Tokio, Sendai y Nigata. Ya en el año 2007 entró en vigencia la tarjeta *PASMO* la cual sirve para la mayoría de las líneas de trenes y algunas líneas de buses en Tokio, aumentando la participación de la tarjeta en el sistema de transporte como medio integrador².

En Londres a partir del año 2003 se incorporó el uso de las tarjetas inteligentes como sistema de pago de pasajes para el sistema de transporte público. La *Oyster Card* presenta rebajas en la tarifa al ser usada en vez de los boletos físicos, y ya para el 2007 era el medio de pago más usado en los viajes realizados en el sistema de transporte público. La *Oyster Card* puede utilizarse en el metro de Londres, autobuses públicos, tranvías, y algunas estaciones de la red férrea del país³.

En San Francisco el sistema de metro operado por la agencia *Bay Area Rapid Transit (BART)* (Buneman, 1984), permite como medios de pago un sistema de boletos y la *EZ Rider card*. En este sistema se cobra según distancia recorrida, por lo que se marca con la tarjeta el ingreso a la estación como la salida de ella. En caso de utilizar boletos se procede de igual forma, ingresando el boleto al comienzo y final del viaje. Esta agencia posee cinco líneas de trenes con 43 estaciones para el total del servicio⁶.

La ciudad china Changchun es la capital de la provincia de Jilin y está ubicada en el centro de la zona nororiental de China. En el 2004 en dicha ciudad, habían sido emitidas 250.000 *IC card* las cuales correspondían al 10% del total de la población y con las cuales se realizaban el 25% de los viajes diarios.

6 Información obtenida de <<https://ezrider.bart.gov/ezrider/>>

La *IC card* es aceptada como medio de pago en los servicios de la *Changchun Transit Company*. Los buses de esta compañía no cuentan con un sistema de localización automática (AVL) ni con un sistema de información geográfica (SIG) que permita establecer la ubicación espacial y temporal de los buses (Lianfu et al., 2007), a diferencia de lo que ocurre en Santiago donde se cuenta con la información entregada por los GPS cada 30 segundos.

En Chicago opera *Chicago Transit Authority* (CTA), agencia que cubre el servicio de transporte público en la ciudad de Chicago y en 380 zonas residenciales aledañas, realizando aproximadamente 1,5 millones de viajes en un día de semana promedio, lo cual representa un 80% de todos los viajes en transporte público realizados en los seis condados del área metropolitana de Chicago.

Esta agencia es la segunda más grande de Estados Unidos, operando el servicio de autobuses y metro de la zona. Además cuenta con los sistemas de recolección automática de tarifa, sistema de localización automática de vehículos y sistemas de conteo de pasajeros (Zhao et al., 2007). Para usar los servicios de la CTA existen distintos medios de pago tales como efectivo, pases, y la *Chicago card* (tarjeta inteligente).

En el año 2004 los usuarios que utilizaban la *Chicago card* representaron un 88% del total de usuarios de la CTA. Además, la CTA cuenta con 153 rutas de buses que sirven cerca de 12.000 paradas en toda región, y un sistema de metro que consta de aproximadamente 242,2 kilómetros de extensión y que cuenta con 144 estaciones⁷.

En Canadá, se utiliza la *smart card* como uno de los medios de pago de tarifa de la agencia de transporte STO (*Société de Transport de l'Outaouais*), cuyas rutas conectan Gatineau (Québec), con el distrito central de negocios (CBD) de Ottawa (Ontario), ciudades que forman una conurbación en la región de la capital nacional. La agencia STO para el año 2005 tenía 16.500.000 usuarios anuales, contaba con una flota de 240 buses, 56 rutas y 1.700 paradas para cubrir la ciudad y aproximadamente 22.000 tarjetas inteligentes en circulación.

En este sistema la tarjeta es personal, y posee sistemas de posicionamiento de buses (Trépanier et al., 2007; Chapleau y Chu, 2007; Chapleau y Chu, 2008; Chapleau et al., 2008). Además de pagar con la *smart card* se puede usar efectivo, ticket, pases mensuales y pases diarios. En la ciudad existe otra agencia de transporte la cual posee una estructura tarifaria distinta, por lo que los boletos no son intercambiables entre agencias y no existe integración de medios de transporte⁸.

7 Información obtenida de <www.transitchicago.com>

8 Información obtenida de < www.sto.ca>

En Santiago, Chile, el medio de pago para ingresar al sistema de transporte público de la ciudad, Transantiago, es la tarjeta BIP!, tarjeta inteligente que posibilita una tarifa integrada que actualmente funciona con una tarifa plana entre modos, que permite viajar de manera gratuita (en la misma dirección) durante dos horas desde el inicio del primer viaje (primera validación) y realizar un máximo de tres transbordos donde sólo uno de ellos puede ser Metro, debiendo validar en cada etapa del viaje para tener el beneficio de tarifa integrada⁹. En septiembre del 2008, había alrededor de 11.163.900 tarjetas emitidas¹⁰.

Transantiago está compuesto por el servicio de buses y metro. Para lo operación de los servicios de buses se dividió Santiago en 10 zonas, donde cada una está constituida por un grupo de comunas de la ciudad, en cada zona existen dos tipos de servicios, troncales y alimentadores. Los primeros conectan distintas zonas cruzando los ejes importantes de la ciudad, mientras que los segundos son recorridos específicos dentro de cada zona de servicio y que alimentan o acercan a los servicios troncales. Para la semana de estudio el sistema contaba con más de 6200 buses, aproximadamente 300 recorridos (ida y vuelta) y sobre 10.000 paraderos¹¹.

El metro es considerado, por Transantiago, como un servicio troncal más¹², el cual para el año 2008 tenía 92 estaciones y 85 kilómetros de red¹³.

Otra característica del sistema es la implementación de aproximadamente 150 zonas pagas, paraderos cerrados dotados de validadores en los cuales se paga la tarifa al momento de ingresar a dichos recintos, esperando en ellos los buses de los servicios que pasan por dicha parada⁹.

2.2 Metodologías para extracción de información

En esta sección se revisa metodologías planteadas en la literatura para la extracción y el tratamiento de los datos que se obtienen a partir de transacciones realizadas con tarjetas inteligentes, aprovechando esta herramienta en pro del sistema de manera de ayudar a la planificación y manejo de los sistemas de transporte. Estas metodologías están basadas en algunas experiencias que fueron examinadas en la sección anterior de este capítulo.

La experiencia reportada por Lianfu et al. (2007), basada en la ciudad de Changchun, China, propone un método para construir la matriz origen destino de paradas de los buses a partir de la información obtenida del sistema de recolección

9 Información obtenida de <www.tarjetabip.cl/como-funciona.php>

10 Información entregada por Transantiago

11 Elaboración propia, capítulo 3: Descripción y Análisis de Datos

12 Información obtenida de <www.transantiagoinforma.cl/descripcion.do>

13 Memoria Anual 2008 <www.metro.cl/gobierno_corporativo.php?s=memoria>

automática de tarifa (AFC). Este método se basa en validar los datos obtenidos, estimar las paradas de subida y bajada con la información disponible y generar una matriz OD de paradas utilizando el método de Furness.

Para dar validez a la información, se compara la distribución temporal de los viajes obtenidos con la *IC card* y la distribución temporal de los viajes estimados en la encuesta OD del año 2003 de la ciudad. Ésta forma de dar validez a los datos se rescata de este paper para ser aplicado y adaptado para el desarrollo de esta memoria.

Una vez validada la información que se obtiene del uso de la *IC card*, ellos implementan un método para estimar las paradas de subida y bajada de los viajes. Como ya se mencionó los buses de la *Changchun Transit Company*, no cuentan con sistema alguno que permita conocer información sobre la ubicación espacio temporal de los buses. Esta falencia del sistema se suple mediante un registro realizado por el conductor del bus, en el cual se anota la hora en que el bus llega a cada parada establecida de la ruta.

Por otro lado la información recolectada por la tarjeta permite su identificación, conocer la hora en que se abordó el bus, la ruta utilizada y la identificación del conductor que operaba el bus. Relacionando ambas informaciones mediante la ruta, el tiempo en que el bus llegó a la parada y el tiempo en que se abordó el bus, ellos consiguen completar la información del viaje del pasajero, con la parada de subida al bus.

Para estimar la parada de bajada, como no tienen información sobre donde se baja el pasajero, ellos asumen que dado que los viajes se concentran en horas punta de la mañana y de la tarde, cada pasajero tiene una parada de subida y bajada relativamente fija, con lo cual se puede asumir que la primera parada donde el pasajero sube en la tarde corresponde a la última parada donde el pasajero terminó su viaje en la mañana.

Posteriormente, concluida la estimación de paradas de subida y bajada, se crea una matriz origen destino (OD) de paradas de buses utilizando el Método de Furness, para poder aplicarlo realizan los siguientes pasos.

- Dividir la ciudad en varias zonas de acuerdo al diseño urbano y a la distribución de los usuarios de la *IC card*.
- Calcular factores de expansión para la generación (R_L^O) y atracción (R_L^D) de viajes para cada zona como:

$$R_L^O = \frac{O_L}{\sum_{i=1}^n O_{ICP_i}} \quad y \quad R_L^D = \frac{D_L}{\sum_{i=1}^n D_{ICP_i}} \quad \text{Ecuación 2-1}$$

- O_L : Total de viajes generados en la zona L.
 D_L : Total de viajes atraídos en la zona L.
 O_{ICP_i} : Total de viajes generados por los usuarios de la *IC card* en la parada de buses i.
 D_{ICP_i} : Total de viajes atraídos por los usuarios de la *IC card* en la parada de buses i.
 n : Total de paradas de buses en la zona L.

- Calcular los vectores de generación (O_{P_i}) y atracción (D_{P_i}) (expandidos) de viajes por parada, para cada zona.

$$O_{P_i} = O_{ICP_i} \cdot R_L^O \quad y \quad D_{P_i} = D_{ICP_i} \cdot R_L^D \quad \text{Ecuación 2-2}$$

Corroborar que el total de viajes generados en todas las zonas sea igual al total de viajes atraídos en ellas, de no ser así se debe volver a calcular los vectores de generación y atracción (expandidos) utilizando factores de balance.

- Deducidos los vectores de generación y atracción expandidos, aplican el método de Furness doblemente acotado. Obteniendo la matriz OD de paradas de bus OD_P .

$$T_{ij} = t_{ij} \times a_i \times b_j \quad \text{Ecuación 2-3}$$

$$s. a \quad \sum_j T_{ij} = O_{P_i}$$

$$\sum_i T_{ij} = D_{P_j}$$

- T_{ij} : Viaje generado desde la parada de bus i hacia la parada del bus j en la matriz OD_P .
 t_{ij} : Viaje generado desde la parada de bus i hacia la parada del bus j en la matriz OD_{ICP}
 OD_{ICP} : Matriz de viajes generados y atraídos por los usuarios de la *IC card* en las paradas de buses.
 a_i : Tasa de crecimiento para generación de viajes.
 b_j : Tasa de crecimiento para atracción de viajes.

En Chicago el sistema que opera la CTA, cuenta con los sistemas de recolección automática de tarifa (AFC), sistema de localización automática de vehículos (AVL) y sistemas de conteo de pasajeros (APC), la información que se obtiene de ellos y como

se manejan las herramientas tecnológicas para lograr estimar la matriz OD de los viajes realizados sólo en metro son descritos por Zhao et al. (2007) y se presentan a continuación.

El sistema AFC de la CTA es un sistema solo de entrada, es decir el pasajero usa la *Chicago card* solo para ingresar a la estación de metro ó al abordar el bus. Los datos registrados al realizar una transacción con la tarjeta son:

- Hora de transacción
- Ubicación (En caso de viajes en metro registra el código de la estación donde ingresó el usuario, si corresponde a un viaje en bus guarda el número de ruta del bus).
- Evento (Este puede ser una transacción de viaje o una transacción de no-viaje, dentro de la primera se considera el ingreso al metro y el abordar el bus, mientras que para la segunda son acciones de carga de dinero a la tarjeta ó adquisición de ella)
- N° de identificación de la tarjeta (único)

La información recolectada por el sistema AVL, registra un número de eventos relacionados con el progreso del vehículo a lo largo de una ruta, los eventos corresponden a paradas de subida y paradas de bajada. Cada evento tiene una hora asociada (registros al segundo), y la información de la posición es derivada desde el sistema GPS.

En resumen los datos obtenidos desde este sistema son la ruta, la identificación del bus y de la ruta patrón, el número de la secuencia de la parada y la hora de ella.

Los atributos anteriormente descritos, por si solos, no entregan información completa acerca del viaje de un usuario de la CTA. Para ello los autores primero procesan la información de cada sistema (AVL y AFC) de manera individual, y luego realizan una integración entre dichos sistemas.

Para procesar los datos del sistema AVL, se enfocan en identificar la parada de subida del bus (obtener sus coordenadas), ya que no se tiene de manera directa de la información arrojada por el sistema.

Para ello crean dos tablas auxiliares estáticas, la primera contiene el número de ruta que realiza el bus, la identificación (número) del bus y la ubicación de este (estimaciones de posición y tiempo obtenidas del sistema GPS), por otro lado, la segunda incluye el número de ruta que realiza el bus, la identificación del bus y el

número de evento que corresponde a la parada de subida con su hora respectiva dentro de la secuencias de eventos registrados. Al relacionar ambas tablas mediante el número de ruta, número de bus y coincidencia temporal entre el evento y la ubicación del bus, se logra identificar la parada de subida y obtener sus coordenadas.

En el caso de la información obtenida a partir de la recolección automática de tarifa (AFC), el procesamiento de datos se centra en identificar el bus, ya que este atributo no se registra en las bases de datos de los AFC de la CTA. Para ello utilizan la información que se tiene del equipo validador donde se lleva a cabo la transacción, este equipo tiene un código que lo identifica y está asociado a un bus (por ejemplo patente ó número dentro de la flota) con esto se obtiene, de manera inmediata, la identificación del bus donde se realiza la transacción.

La integración de la información entre los sistemas AVL y AFC se realiza mediante la hora de la transacción y el bus donde se realizó (atributos compartidos en las bases de datos de ambos sistemas) obteniendo la parada donde abordó el pasajero. En este trabajo interesa inferir los destinos de los viajes de los pasajeros de metro, por lo que se incorporan los posibles transbordos entre bus y metro.

Para ello se detectan las posibles transacciones en que por coincidencia temporal y espacial sean posibles. Para el análisis espacial las paradas de los buses y las estaciones de metro están geocodificadas en el sistema GIS, luego se calculan las distancias entre paradas y estaciones cercanas, si éstas son menores a 400 metros ó 5 minutos de caminata son posibles transbordos.

Luego se enfocan en inferir la matriz origen destino de los viajes realizados en metro, para ello suponen que los usuarios que realizan dos segmentos de viaje consecutivos no utilizan otro sistema que no sea el transporte público, que los viajes en metro comienzan su retorno donde terminaron su primer viaje, y los pasajeros terminan su último viaje del día en la estación donde iniciaron su primer viaje del día.

Para inferir los viajes definen los posibles escenarios en que pueden realizar los viajes de segmentos consecutivos entre metro y bus. Estos corresponden a identificar si están disponibles los sistemas AVL y GIS. Si no están disponibles ambos sistemas la información proporcionada por el bus no se toma en cuenta, si sólo está disponible el sistema GIS y no el AVL, entonces se puede utilizar información parcial del viaje en bus la cual está determinada dada la factibilidad del transbordo (infactible, no se conoce destino; único transbordo, se identifica destino; diferentes posibilidades de transferencia, múltiples destinos factibles), finalmente si se dispone de ambos sistemas se obtiene información completa acerca del viaje en bus.

Las investigaciones basadas en la experiencia de la empresa de Transporte STO (*Société de Transport de l'Outaouais*), la cual conecta Gatineau (Québec) con el distrito central de negocios (CBD) de Ottawa (Ontario) en Canadá, son múltiples. Dichos estudios son desarrollados por el Grupo MADITUC de la escuela politécnica de Montreal liderados por el profesor Robert Chapleau.

El trabajo de este grupo, consiste en comparar las ventajas y desventajas de los métodos tradicionales de encuestas versus los datos que se puede obtener de la *smart card*. Para ello desarrollan una metodología de análisis, la cual permite establecer el manejo de la información que se debe realizar, reconocer errores en los datos para su posterior eliminación ó corrección, estimar paradas de bajada, entre otros. A continuación se presenta un resumen de cada publicación revisada de este grupo.

El trabajo de Chapleau y Chu (2007) plantea un procedimiento íntegro, que permite identificar y reemplazar los datos incorrectos ó sospechosos que se puede encontrar en la información entregada por los sistemas de recolección automática de tarifa.

Para ello comienzan identificando las posibles fuentes de error, las cuales principalmente se deben a la incorrecta, incompleta o falsa información introducida por el conductor en el momento de iniciar un nuevo recorrido, estimando que las operaciones erróneas y que contienen valores sospechosos alcanzan un 15% de los datos que ellos disponen. Debido a lo anterior, concluyen sobre la necesidad de mejorar la calidad de los datos (exactitud de ellos), de manera de impedir la propagación posterior de los errores.

Para el procesamiento de la información, proponen un procesamiento eficiente que requiere el modelamiento de los objetos primarios, en el cual los datos son tratados de manera individual (es decir, desagregadamente) y son sometidos a lógicas espacio-temporales.

La identificación de datos erróneos y sospechosos, se establece en base a la regularidad de las operaciones del sistema de transporte y a los patrones de viaje históricos de los usuarios. Cabe señalar que muchos de los valores encontrados pertenecen a una secuencia.

La corrección tiene por objeto identificar y atribuir la carrera y la parada de subida más probable. Para ello utilizan el número del bus y el instante de la transacción lo cual permite rehacer el horario y trazado del recorrido del bus. Luego, empleando el enfoque bayesiano y las transacciones validadas en la historia de viaje del usuario, se identifica la carrera que este usuario toma de manera más frecuente en el mismo horario otros días. Así, las paradas de subida más probables son derivadas de acuerdo con las

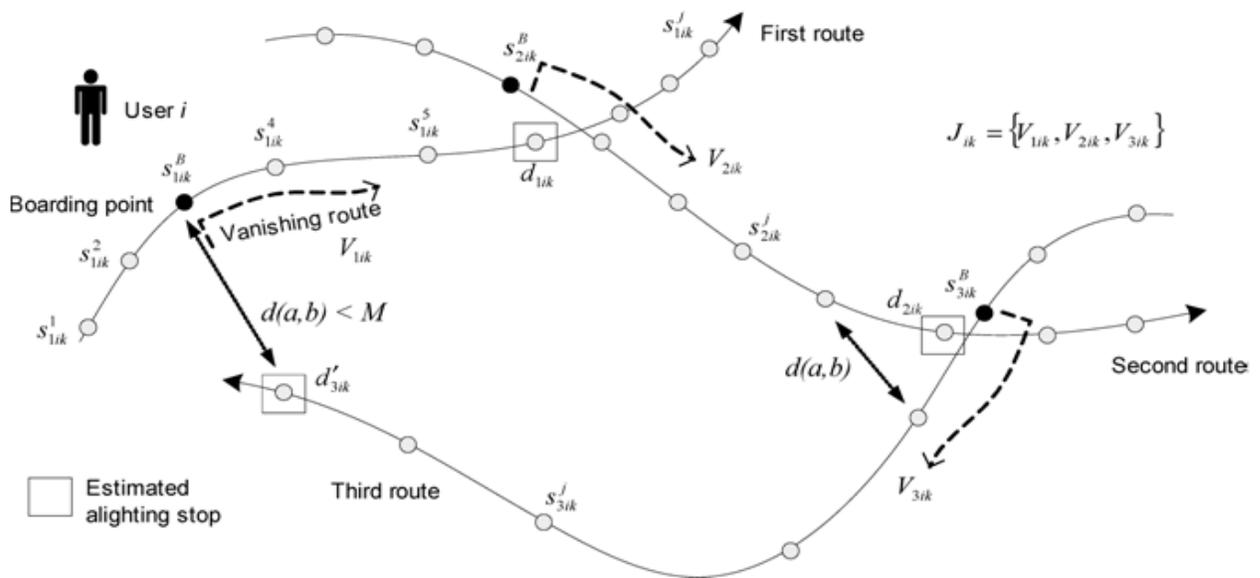
pautas de los viajes históricos de los usuarios o por interpolación o extrapolación lineal, utilizando la distancia y el tiempo de viaje registrado en las transacciones.

Trépanier et al. (2007), presentan un modelo para estimar la localización del destino para cada pasajero que utilizó la *smart card* como medio de pago al subir al bus. Para obtener el algoritmo los autores basaron su metodología en tres tareas: examinar la arquitectura del sistema de recaudación de la agencia STO, identificar los objetos a ser analizados para plantear el modelo final y, establecer las consideraciones que se debe tener con el modelo para poder aplicarlo.

Al examinar la arquitectura del sistema de tarifas, Trépanier et al. plantean un esquema de relación para los componentes de la información, para ello ésta se clasifica según la naturaleza de los objetos (campos de información) de acuerdo al método propuesto por Trépanier y Chapleau (2001), el cual distingue cuatro grandes clases de objetos: estáticos (relacionados con el mantenimiento de transporte), cinéticos (describen movimientos), dinámicos (actúan en el transporte) y sistémicos (redes, sistemas). Dicho método es una extensión natural del enfoque totalmente desagregado, ya que se basa en un análisis individual de los elementos del sistema.

Por otro lado, para simplificar el esquema, se agrupan los objetos en cuatro grupos: objetos de red (elementos centrales de la red de transporte de la STO, rutas, paradas, buses), objetos de operación (elementos relacionados con la operación de los buses, conductores, terminales), objetos administrativos (elementos propios del sistema de recolección automático de tarifas, tarjetas, transacciones, puntos de venta de tarjeta, punto de carga de tarjeta) y objetos de demanda (etapas de viaje, viajes, usuario *smart card*, demanda de transporte). Obteniendo los campos valiosos del set de datos que entrega el sistema.

Una vez examinada la arquitectura del sistema e identificada la información valiosa, Trépanier et al. define el modelo final, el cual se ilustra en la Figura 2-1.



Fuente: Trépanier et al., 2007

Figura 2-1: Modelo de identificación de paradero de bajada

El método contempla la definición de una secuencia de paradas (s_{rik}^j) para la ruta r del usuario i realizada el día k , donde j corresponde al número de parada dentro de la secuencia.

En la Figura 2-1, se observan para el usuario i , los puntos de subida y las rutas que aborda a lo largo del día, definidas por las secuencia de paradas correspondientes. El punto donde se aborda el servicio en cada ruta, da paso a lo que Trépanier et al. llaman *vanishing route*, ya que no se sabe qué pasa con ese usuario hasta que se tiene una transacción posterior realizada por él en una ruta distinta.

En el ejemplo de la figura se conoce los puntos de subida a las rutas uno, dos y tres. Dado que se desconoce donde se baja el pasajero en la primera ruta pero sí se sabe dónde se sube a la siguiente ruta, se estima la parada buscando aquella que sea la mejor candidata, en la ruta uno, a ser la más cercana al punto donde se aborda la ruta dos, lo que se obtiene de calcular la distancia mínima entre las paradas de la ruta uno y la parada donde aborda la ruta dos.

Dicha distancia se compara con un valor de tolerancia asociado a una distancia máxima caminable, la cual para este caso se estableció en dos kilómetros. Si la estimación de parada de bajada cumple con dicho criterio, es aceptada como válida. Este proceso se puede repetir con el número de rutas que se quiera, mientras se tenga una transacción posterior en el día. Éste tipo de viaje lo denominan “viaje normal”.

Para el caso de la última transacción del día, los autores asumen que la última parada de bajada del día del usuario i es el primer punto donde abordó dicho día, llamando a este tipo de viaje, “último viaje”.

Los autores llaman “otros viajes” a todos aquellos viajes donde el destino no pudo ser estimado mediante el modelo planteado. Estos viajes son comparados con todos los viajes realizados por el mismo usuario durante el mes, de manera de encontrar uno similar que tenga estimación de destino, la semejanza implica que usen la misma ruta, que la parada de subida y la diferencia entre los tiempos de las transacciones (horas de subida de cada viaje) sean lo suficientemente pequeñas. De esta forma se infieren destinos que no eran factibles de obtener.

Los viajes cuyos destinos no pueden ser estimados o inferidos mediante el modelo, pueden ser clasificados en tres categorías: “No válidos” (datos que tienen error y no pueden ser corregidos), “Demasiado lejos” (la distancia entre la parada de bajada estimada y la siguiente parada de subida es demasiado grande y no existe otra posibilidad de estimarle destino) e “individuales” (viajes solos a los cuales no se les puede realizar un análisis de regularidad).

Para implementar el modelo, se necesita realizar previamente una serie de pasos experimentales a la base de datos disponible, los cuales consisten en:

- Pre-procesamiento de errores: Corrección de algunos errores, previo al primer análisis.
- Análisis global a la red de transporte.
- Análisis de comportamiento de usuarios.
- Análisis de la oferta (reconstrucción del funcionamiento del sistema).
- Aplicación del modelo de estimación de destinos de los viajes.

Los resultados al aplicar este modelo a los datos obtenidos de la STO, resultan esperanzadores con un 66% de destinos estimados. Este porcentaje está constituido por un 33% de “viajes normales”, un 20% correspondiente a “últimos viajes” y un 13% a “viajes individuales” (no presentan regularidad, pero es factible estimar destino). Al restante 34% no se le encontró destino, correspondiendo un 21% a viajes “no válidos”, un 7% a “demasiado lejos” y un 6% a “individuales” (a los cuales no se les podía estimar destino y al no presentar regularidad tampoco es factible inferirlo). Cabe señalar que en horas punta el porcentaje de estimaciones de destino aumenta llegando a un 80%.

Chapleau y Chu (2008), describen una metodología que permite identificar viajes que están conectados entre sí, usando el concepto de coincidencia espacio-temporal para detectar transbordo en los viajes, con el fin de enriquecer la información y hacerla más completa para fines de planificación. Para ello, primero reconstruyen a partir de las transacciones de la tarjeta los itinerarios de cada usuario, luego estiman el tiempo de

bajada para cada segmento de viaje y plantean un algoritmo para detectar transferencia, finalmente realizan un análisis de patrones de transferencia.

Para la reconstrucción de itinerarios desglosan los datos de las transacciones de cada tarjeta, de manera de tener cada segmento de viaje por separado, almacenando el origen y destino estimado (posición (x, y) de parada) de cada segmento (Trépanier et al., 2007), el tiempo de subida, la ruta escogida, y los atributos del usuario de la tarjeta. Cada itinerario constituye un viaje-persona que se define como un movimiento de una persona entre dos puntos, paradas de origen y destino, con un propósito específico.

Para reconstruir los itinerarios e identificar la parada de origen y destino finales, se debe determinar correctamente los viajes enlazados entre sí, lo que se logra estableciendo que la parada de bajada de la primera ruta y la parada de embarque de la siguiente ruta estén físicamente cerca y accesible a pie. Además debe existir una relación temporal entre el momento de bajada (no se conoce a priori) y la hora de la siguiente transacción. Se considera que los usuarios de la tarjeta actúan de manera racional y que tienen conocimiento completo de la red, adquiridos por la experiencia de viaje.

Como los datos de la tarjeta inteligente sólo proporcionan información parcial (tiempo de transacción y ruta abordada), es necesario identificar con precisión la trayectoria espacio-temporal de la ruta, la que se reproduce a partir de las transacciones de embarque y la secuencia de paradas. Lo anterior se realiza con el fin de considerar todas las limitaciones temporales existentes para el cálculo de la velocidad del bus, de manera de estimar la hora más probable de bajada.

Para ello toman en cuenta la hora prevista de salida del bus del terminal, la transacción posterior del usuario (límite superior para el tiempo estimado de bajada del bus), en aquellas paradas donde no existe registro de transacciones se interpola linealmente para estimar el tiempo de llegada del bus a la parada, considerando la primera y última transacción en cada parada y la distancia entre estas, finalmente se asume que el bus llega a la hora prevista a la terminal. Si la velocidad resultante de realizar el recorrido es demasiado alta o baja se utiliza la velocidad media para el cálculo de la hora estimada de bajada.

Cabe señalar que los autores mencionan que la complejidad de este problema se reduce bastante si los registros GPS detallan el seguimiento del bus de manera de poder ser combinados con los registros de transacciones (caso de Santiago).

Una vez estimado el tiempo de bajada del usuario del bus (obtenido a partir de la información temporal de los buses en la red), se compara con la hora de la siguiente transacción. Por otro lado, se calcula el tiempo de acceso a la siguiente ruta, a partir de

la distancia entre la parada de bajada estimada y la parada de subida (distancia acceso), considerando velocidad de caminata 1,2 m/s y distancia en línea recta. El algoritmo comprueba que el siguiente bus que aborda el usuario al llegar a la parada corresponda a la primera pasada de esa ruta desde que el usuario llegó a dicha parada.

Para compensar el hecho de que algunos usuarios de tarjetas no tienen conocimiento perfecto de la red y de los horarios, la variación de la velocidad de caminata, y el riesgo de la congestión de las rutas con alta frecuencia, agregan un pequeño margen temporal (5 minutos) a la hora de llegada del usuario a la siguiente parada. Esto permite que el algoritmo incorpore la incertidumbre asociada a estos eventos. Cabe señalar que se considera que los buses no tienen restricción de capacidad, es decir el usuario siempre aborda el bus correspondiente, además de incorporar variantes o desviaciones de la ruta que son importantes para la red.

El algoritmo también considera aquellos casos donde a pesar de cumplir las condiciones de coincidencia espacio-temporales, no corresponde a transbordo, como lo son dos transacciones seguidas en la misma ruta y dirección (escala dentro del viaje para otro propósito), dos transacciones consecutivas en la misma ruta con distinta dirección acusa viaje ida y vuelta.

Para terminar, realizan un análisis de los patrones de transferencia encontrados, de manera de obtener una comprensión completa del modelo de transferencia de la red, para ello determinan tiempos de espera, de transferencia y distribuciones de estos tanto temporal como espacialmente.

El principal resultado se obtiene al comparar el número de transacciones que el sistema de recolección automática de tarifas etiqueta como operación de transferencia (utilizando sólo un umbral temporal) y las operaciones que el algoritmo definido identifica como viajes de transferencia, se obtiene experimentalmente que los datos originales sobreestiman el número de viajes de transferencia en cerca de un 40%, es decir al no considerar coincidencia espacio-temporal se encubren viajes con actividad corta en el destino, y se consideran viajes de regreso como viajes de transferencia.

Chapleau et al. (2008), presentan una recopilación de trabajos anteriores para proponer una metodología de depuración y validación de datos con el fin de asegurar la precisión de ellos, dado que los datos a utilizar son un subproducto de un sistema de recolección de tarifa y no necesariamente tiene todos los ingredientes que son necesarios o que se quisieran tener. La metodología propuesta consta de 5 etapas:

- Validación de la información.
- Corrección de valores erróneos o sospechosos.
- Estimación de bajada.
- Detección de transbordo.
- Análisis post

La validación de la información se realiza a través de la detección de datos erróneos o sospechosos, mediante un análisis de consistencia interna de los datos que incluye los movimientos espacio-temporales de los buses y las reglas de operación del sistema. Los errores detectados se deben principalmente a la manipulación de los equipos (información ingresada por el conductor del vehículo), y al error asociado al funcionamiento y precisión del GPS.

Una vez detectados los valores incorrectos, se procede a corregirlos sustituyéndolos por datos correctos obtenidos ó deducidos de la operación del sistema, en esta etapa es importante mencionar que no se eliminan los valores equivocados ya que todos los datos están relacionados con otros datos operativos que tienen una lógica espacio-temporal y que pueden ser de ayuda para otros fines en el sistema, dándole validez a la base.

Ya depurada la base de datos, se procede a implementar el modelo propuesto por Trépanier et al. (2007), el cual estima la parada de bajada mediante el seguimiento del usuario (puntos de subida a distintas rutas), y las secuencias de paradas dentro de cada ruta abordada.

Al implementar el modelo de estimación de bajada, ellos obtienen los destinos de los viajes para aquellos en que fue factible estimar o deducir. Conociendo los destinos de los viajes y la hora en que se estima se llegó a éste se procede a identificar transbordo, utilizando el algoritmo propuesto por Chapleau y Chu (2008), basado en la coincidencia espacio-temporal de los viajes.

Finalmente, a partir de la información deducida e interpretada del viaje, se realiza un post-análisis que ayuda a contextualizar la información, permitiendo construir itinerarios, y deducir algunos detalles acerca del viaje tales como perfiles de carga, distribución espacial de las paradas, posición espacio-temporal del vehículo, carga entre paradas, identificar puntos de interés, etc.

2.3 Síntesis

Este es un tema nuevo dado la creciente incorporación masiva de tarjetas inteligentes en los sistemas de transporte de distintos lugares del mundo, en conjunto

con los sistemas de posicionamiento incorporados a las flotas de buses. Si bien, el fin de los sistemas de recaudación automática corresponde a un objetivo de administración, la información que ellos entregan en combinación con sistemas de posicionamiento constituye una nueva y poderosa herramienta de apoyo a la planificación de la operación de los sistemas, que está comenzando a ser investigada.

Debido a la variedad de formas de implementar el pago a través de *smart cards* y las diversas características que identifican cada uno de los sistemas que la implementan (tarifas integradas, no integradas, parciales, etc), sumada a la constante integración de nuevas tecnologías, resulta difícil encontrar una metodología que sea aplicable a todos los casos.

Los esfuerzos en investigación se han concentrado principalmente en las siguientes áreas:

- Integrar y enriquecer la información entregada por los distintos sistemas de información (recaudación, posicionamiento, conteos de pasajeros, entre otros).
- Detectar y corregir errores en la información generada por la operación del sistema.
- Estimación de bajada ó destino del viaje.
- Identificación de transbordo.
- Generación de matrices origen destino a partir de la información.

Los principales desafíos resueltos son:

- Identificación del punto de bajada: basado en el supuesto de que el usuario termina su viaje anterior cerca de la siguiente transacción realizada en el sistema.
- Mejoramiento de la precisión de los datos: reconociendo errores mediante el estudio del comportamiento del usuario a lo largo de un período de tiempo.
- Detección de transbordo: incorporando el concepto de coincidencia espacio-temporal, el cual permite mejorar la definición de transbordo proveniente del sistema de recaudación (que considera un umbral temporal).

- Integración y enriquecimiento de la información: relacionando transacciones con posicionamiento de buses de manera de enriquecer y hacer más completa la información disponible permitiendo un entendimiento más amplio del sistema y del comportamiento de los usuarios.
- Generación de matrices origen destino: resuelto para aquellos casos donde se valida con la tarjeta a la entrada y salida del sistema.

Se debe destacar que se debe seguir trabajando en el desarrollo de herramientas que permitan profundizar de una manera más precisa y sofisticada en áreas de investigación como lo son identificación y corrección de errores (ya que estos provienen de múltiples fuentes como lo son emisiones gps erradas, relación entre tarjeta y usuario ya que no en todos los sistemas la tarjeta es personal e intransferible (caso de Santiago), post análisis una vez obtenidos origen-destino y tiempos de viaje, identificación de transbordo (considerando restricciones de capacidad de pasajeros) o actividad corta en el destino, generación de matrices de origen destino para casos en que el destino sea inferido.

3 Descripción y Análisis de Datos

Diariamente en el sistema de transporte público de Santiago se generan más de cinco millones de transacciones de uso (viaje) y diez millones de observaciones de posición (emisiones gps) de buses. La información que se recolecta proviene de la interacción entre el sistema de recolección automático de tarifas y el sistema global de posicionamiento gps (Ver Figura 3-1).

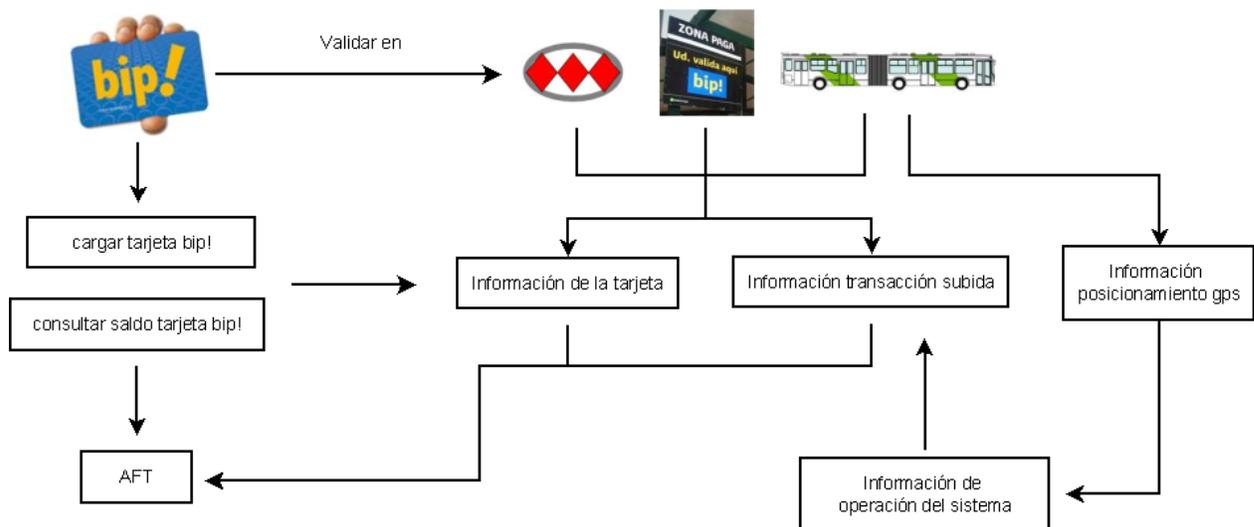


Figura 3-1: Interacción entre sistema de recolección automática y sistema de posicionamiento global, para Transantiago.

El administrador financiero de Transantiago (AFT) es quien se encarga, a través de la empresa Sonda, de gestionar el sistema de validación y pagos mediante la tarjeta BIP!, además de la gestión de flota y equipos en los buses.

En el marco del convenio de colaboración establecido entre Transantiago y la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, fueron entregados datos generados por la operación del sistema de transporte público de Santiago a lo largo de todo el día durante la semana comprendida entre los días 1 y 7 de Septiembre de 2008.

3.1 Descripción de la Información

La información original disponible se encuentra organizada por día, horario, operador y tipo de servicio (alimentador o troncal), en las siguientes 5 carpetas.

➤ BIPS: Registro de transacciones BIP realizadas en un período de tiempo dado, incluyendo campos de información como el número de identificación de la tarjeta (ID), fecha y hora en que se realiza la transacción (tiempo), sitio de la transacción (patente bus, validador zona paga, estación de metro), monto de la transacción, número de

secuencia de transacción, tipo de transacción (uso, carga o consulta. Solo fueron entregadas las de uso), tipo de contrato (escolar básico, media, superior, normal).

➤ Posicionamiento: Información GPS por placa patente única del bus (PPU), incluye los campos latitud, longitud, fecha y hora, PPU, concesionario, ignición (motor de bus prendido (si/no)). Cabe mencionar que los registros GPS son emitidos usualmente cada 30 segundos, sin embargo también existen emisiones en otros intervalos de tiempo, principalmente cada 10 y 300 segundos (Ver Tabla 3-22).

➤ 7° programa de operación: Programa de operación para el sistema de transporte de la ciudad, vigente para el período en estudio, en el cual se encuentran unidad de negocio, frecuencias programadas, velocidad media por servicio-sentido en cada período de asignación de servicio.

➤ Asignación de servicio a buses: Estimación realizada por Transantiago del servicio-sentido que realizó cada bus durante el período de estudio. Incluye la distancia recorrida, el tiempo en servicio, y velocidad media resultante para cada bus en cada servicio-sentido.

➤ Paradas y zonas pagas: Información sobre la ubicación y tipo de paradas, para estaciones de transbordo y paradas sencillas. Las zonas pagas presentan: código de parada, horario de operación, servicios que la utilizan, estimación del porcentaje de transacciones correspondientes a cada servicio, asignación de validadores a zonas pagas.

Para integrar toda la información en grandes bases de datos, es necesario procesar la información original y crear nuevas tablas que combinan distintos campos de información provenientes de las carpetas anteriormente citadas¹⁴.

En la Figura 3-2 se presenta los principales campos de información contenidos en las nuevas bases de datos y que se emplean en el desarrollo de la investigación.

¹⁴ Proceso desarrollado por el analista de sistemas Mauricio Zuñiga en el marco del desarrollo del Proyecto Bicentenario de Ciencias y Tecnología (PBCT). El detalle del procedimiento se encuentra reportado en el "Informe proceso de modificaciones base de datos Transantiago" documento interno del PBCT.



Figura 3-2: Bases de datos y campos relevantes para el estudio.

Como se puede apreciar, la tabla Transacciones BIP tiene dos nuevos atributos (posición y servicio), los cuales se obtienen de acuerdo al lugar donde se realizó la transacción.

La asociación de posición se realiza de acuerdo a:

- Bus: Información deriva de la tabla Posicionamiento (originada a partir de la carpeta Posicionamiento). Asociando una posición según la hora de la transacción.
- Metro: Información proviene de la tabla Rutas Metro. Se asocia posición a la estación de metro de subida.
- Zonas pagas: Información se genera a partir de la tabla Posicion_zp (creada desde la carpeta Paradas y zonas pagas), asociando posición a la zona paga donde se realiza la transacción.

La asociación de servicio a cada transacción se realiza de acuerdo a:

- Bus: Información se conoce directamente de la tabla Asignación_bus (originada a partir de la carpeta Asignación de servicio a buses). Asociando servicio correspondiente según hora de la transacción. Si no existe servicio asignado a PPU, campo queda en blanco.
- Metro: Campo de información queda en blanco.
- Zonas pagas: Información se obtiene de la tabla Zonas Pagas (creada desde la carpeta Paradas y zonas pagas), asociando servicios asignados para la zona paga donde se realiza la transacción.

Las tablas restantes son el resultado del proceso de integración de información realizado, exceptuando las tablas Estaciones de Metro y Rutas Metro que fueron generadas a partir de información entregada por Metro S.A.

A continuación se describe la información contenida en la semana de estudio y que se utilizó para experimentar (primera semana de Septiembre 2008).

El sistema opera integrando buses y metro, gracias al pago de la tarifa realizado mediante la tarjeta de prepago BIP!. En el sistema de buses existen dos tipos de servicios: troncales y alimentadores. En total hay cinco troncales y nueve alimentadores. El metro es considerado por Transantiago como un servicio troncal más del sistema (se suma a los cinco troncales de buses).

En el sistema estaban operando más de 6200 buses, los cuales ofrecían alrededor de 300 recorridos, estos buses tenían incorporado el sistema de información geográfica GPS y emitían su posición aproximadamente cada 30 segundos.

Las principales características de la base de datos de transacciones se detallan en la Tabla 3-1.

Tabla 3-1: Principales características Tabla Transacciones BIP.

Transacciones:	34.513.933
Tarjetas BIP! activas:	3.279.803
Transacciones con posición:	34.003.738
Buses con transacciones	5.940

Para fines de agosto del 2008 el sistema había emitido un total de 11.163.893 tarjetas. Es decir que para la semana en estudio solo un 29,3% de las tarjetas estaba activa, realizando en promedio cada una de ellas 1,77 transacciones diarias en un día laboral.

Este valor no necesariamente se traduce en viajes por día debido a que un viaje puede estar compuesto por más de una etapa. Una correcta interpretación de este número, es etapas de viaje por día con lo cual no se puede establecer a priori una comparación con los valores obtenidos en la EOD del año 2001, donde se obtuvo que en promedio cada habitante realizaba 1,7 viajes motorizados diarios en día laboral los cuales incluyen viajes en automóvil y otro medios de transporte (taxis, colectivos, etc).

Del total de transacciones, un 98,5% poseen posición asociada. Este porcentaje se traduce directamente en información sobre el origen del viaje. De las transacciones que poseen posición, 19.905.943 (58,5%) transacciones se realizan en bus, 12.377.129 (36,4%) en metro y las restantes 1.720.666 (5,1%) se realizan en alguna de las 137 zonas pagas.

De las transacciones realizadas en bus y que tienen posición, 16.389.809 (82,3%) presentan asignación de servicio, lo que significa que el bus donde se realizó la transacción cumplió con las condiciones que permitían a Transantiago establecer el servicio-sentido que ese bus estaba realizando en el momento de la transacción.

Del total de buses que presentan transacciones (Ver Tabla 3-1) 5933 tienen transacciones sin asignación de servicio, lo que permite establecer que un 99,8% de los buses con transacciones en algún recorrido de la semana no cumplieron con las condiciones para determinar el servicio-sentido que estaban realizando.

Por otra parte, las principales características de la base de datos de posicionamiento se detallan en la Tabla 3-2.

Tabla 3-2: Principales características Tabla Posicionamiento.

Total observaciones:	75.882.334
Nº buses:	5737
Nº recorridos:	310

Del total de observaciones el 91,7% de los registros (69.605.858) son emitidos con el motor del bus encendido. Dicho porcentaje se convierte en registros disponibles para conocer su posición.

Al comparar el número de buses que presentan transacciones (Ver Tabla 3-1) con el número de buses que presentan registros de posición (Ver Tabla 3-2) se observa una diferencia de 203 buses, los cuales presentan transacciones pero no se encuentran en la tabla de posicionamiento lo que causa algunas de las transacciones sin posición.

3.2 Análisis de Datos

Para el desarrollo de este trabajo la información más importante es la de transacciones BIP y la de posicionamiento de buses. Como ya se mencionó, la primera se fue complementando con información de otras bases de datos (inclusive la de posicionamiento) por lo que al analizar esta base se está mirando también las tablas que están vinculadas a estas (zonas paga, rutas metro).

3.2.1 Tabla Transacciones BIP

3.2.1.1 Análisis Global

Cada transacción indica el operador o unidad de negocio donde ésta fue realizada (ver Tabla A-1, Anexo A). Al agrupar según la categoría del operador se obtiene qué tipo de servicio abordó la transacción, logrando establecer que, del total de transacciones 15.187.468 son realizadas en servicios troncales, 12.377.129 en Metro y 6.949.336 en servicios alimentadores (Ver Figura 3-3).

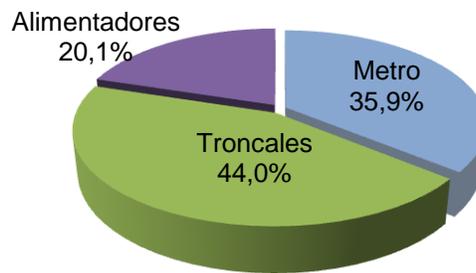


Figura 3-3: Distribución de transacciones según tipo de servicio.

Se puede ver que, en los servicios troncales se realizan un poco más del doble de transacciones que en los servicios alimentadores, y que metro presenta un alto porcentaje de participación.

En la Figura 3-4 se muestra la distribución de transacciones según el lugar donde se realiza la transacción. Realizándose 20.343.686 transacciones en bus, 12.377.129 en metro y 1.793.118 en zonas paga.

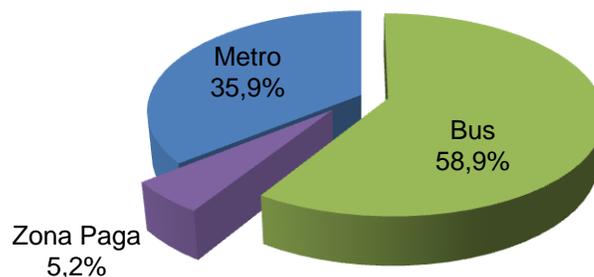


Figura 3-4: Distribución de transacciones según lugar donde se realizó validación.

De las transacciones realizadas a bordo del bus un 97,8% presenta posición, en tanto para las realizadas en zonas pagas este porcentaje corresponde a un 96%.

La existencia de transacciones sin posición, se debe a diferentes causas dependiendo del lugar donde se realizó ésta. Si se realizó a bordo del bus, la falta de posición asociada a la transacción se debe al no registro de posición gps, o bien falla en la emisión de ellos. En cambio, para el caso de zonas pagas, esto puede ser debido a, la falta de código de la parada, falta de posición del paradero o estación de transbordo vinculado a dicha zona paga, como también, a la asociación de más de una posición a estos.

También se puede conocer la distribución de transacciones por día de la semana, lo cual se muestra en la Figura 3-5. En ella, se puede ver que cada día hábil recibe aproximadamente un 17% del total de transacciones, mientras que el día sábado recibe 10% de ellas y el día domingo un 5% (ver Tabla A-3, Anexo A). En consecuencia, se puede afirmar que las transacciones se concentran en los días hábiles, comportamiento que concuerda con el observado en la EOD 2001.

Para esta semana en particular el máximo de transacciones se alcanza el día lunes y el mínimo (dentro de los días laborales) el día viernes, esto último es una singularidad de esta semana ya que por lo general los días viernes son días donde se generan mayor cantidad de viajes¹⁵.

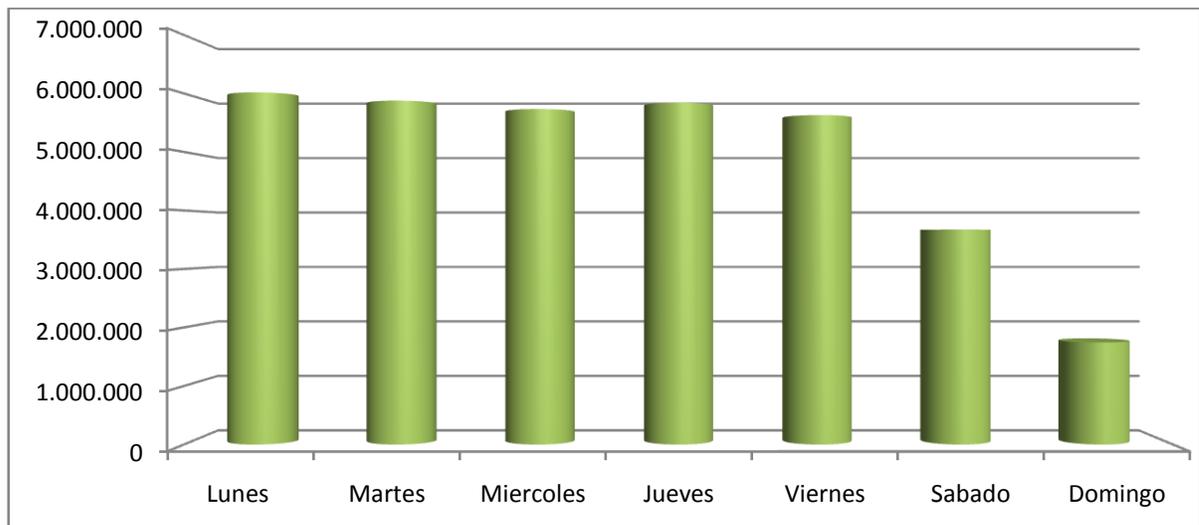


Figura 3-5: Transacciones por día

Al agrupar la información por tipo de día: día laboral (promedio de transacciones de lunes a viernes), sábado ó domingo se obtiene las transacciones promedio que se realizan de acuerdo a la clasificación, las cuales se aprecian en la Tabla 3-3. Se puede

¹⁵ Información obtenida a partir de las EOD.

observar que en un día laboral promedio se realizan 1,5 veces más transacciones que para un día sábado, mientras que en relación al día domingo estas se triplican.

Tabla 3-3: Transacciones según tipo día

Tipo día	N° transacciones
Laboral promedio	5.819.454
Sábado	3.673.177
Domingo	1.743.487

Si se asocia la información según tipo de contrato de la tarjeta (relacionado con el tipo de usuario que la utiliza) se puede identificar dos grupos de tarjetas: adulto y estudiante. En la Tabla 3-4 se puede observar la partición de las transacciones según tipo de tarjeta y día.

Tabla 3-4: Transacciones según tipo día y tarjeta

Tipo día	Estudiante	Adulto	Total
Laboral promedio	1.433.977	4.385.477	5.819.454
Sábado	909.774	2.763.403	3.673.177
Domingo	408.315	1.335.172	1.743.487
Total	2.752.066	8.484.052	11.236.118

Como se puede observar en la Tabla 3-4 la participación de las tarjetas tipo estudiante es un 25% del total de transacciones en día laboral y sábado, mientras que disminuye levemente (a 23%) en día domingo.

Además se puede ver que en un día laboral promedio las tarjetas tipo estudiante realizan 1,6 veces más transacciones que en día sábado y 3,3 veces más que un día domingo. En tanto, para las tarjetas tipo adulto la relación entre día laboral promedio y día sábado se mantiene, mientras que aumenta levemente (3,5 veces) en relación a las transacciones de un día domingo.

También se analizó la distribución de transacciones según tipo de tarjeta y pago realizado al sistema. La estructura tarifaria vigente establece una tarifa adulto en buses de 380 pesos. En el metro en tanto, esta es diferenciada por períodos, en horario valle tenía un valor de 380 pesos y en horario punta 420 pesos, por lo que al realizar un transbordo de bus a metro en horario punta, se cobra la diferencia (40 pesos).

La tarifa escolar es en todo período 130 pesos. El pago cero corresponde a aquellas transacciones posteriores a la primera transacción que se realizan en un

período de dos horas a partir de esta. El detalle acerca de cómo se reparten los pagos según tipo de tarjeta se presenta en la Tabla 3-5.

Tabla 3-5: Transacciones según tipo día, tarjeta y pago

Tipo día	Estudiante			Adulto				Total
	Pago 0	Pago 130	Pago 380	Pago 0	Pago 40	Pago 380	Pago 420	
Laboral promedio	658.300	775.500	177	1.582.797	181.757	2.238.036	382.887	5.819.454
Sábado	447.601	462.102	71	1.082.940	0	1.680.463	0	3.673.177
Domingo	205.122	203.181	12	501.031	0	834.141	0	1.743.487
Total	1.311.023	1.440.783	260	3.166.768	181.757	4.752.640	382.887	11.236.118

Del total de transacciones un 24,5% es realizada por tarjetas tipo estudiante y un 75,5% por tarjetas tipo adulto.

En relación a la distribución del total de transacciones según pago realizado, un 58,5% paga el total de la tarifa lo cual de acuerdo al criterio establecido indica que comienza un viaje, y el restante 41,5% realiza pago parcial (cambio de modo con distintas tarifas horario puntas) o pago cero de acuerdo al criterio de tarifa plana para transbordo dentro de un intervalo de tiempo de dos horas. Este último porcentaje encubre actividades cortas en el destino, así como también viajes cortos en diferentes direcciones (ida y vuelta).

La distribución según tarifa pagada para transacciones realizadas por tarjetas tipo estudiante en un día laboral promedio varía levemente en relación a las proporciones presentadas para el total de transacciones (54,1% pago tarifa \$130 versus 45,9% pago tarifa cero). Para día sábado esta distribución casi se iguala (50,8% pago tarifa \$130 versus 49,2% pago tarifa cero). En tanto para día domingo los pagos cero superan levemente a los pagos de tarifa completa (50,2% pago tarifa cero versus 49,8% pago tarifa \$130).

Los pagos \$380 para tarjeta tipo estudiante son un porcentaje muy bajo ya que corresponden a aquellos casos en que la transacción fue inmediatamente después y en el mismo sitio que la anterior, lo que corresponde en la práctica a pagarle la tarifa a otra persona pero con la misma tarjeta estudiante por lo que el sistema cobra tarifa normal.

La distribución según tarifa pagada para transacciones realizadas por tarjetas tipo adulto varía levemente en relación a las proporciones presentadas para el total de transacciones, aumentando la proporción de transacciones que pagan tarifa total (para día laboral promedio 59,7% pago tarifa total (\$380 y \$420) versus 40,3% pago tarifa parcial (\$40 y cero)). Esta tendencia se acentúa los días de fin de semana (día sábado

60,8% pago tarifa \$380 versus 39,2% pago tarifa cero y día domingo 62,5% pago tarifa \$380 versus 37,5 pago tarifa cero).

Los días de fin de semana no existe diferencia entre las tarifas adulto de bus y metro en horario puntas, no registrándose transacciones con pagos \$40 y \$420 para dichos días.

Como complemento a la distribución de las transacciones por tipo de servicio (troncales, alimentadores y metro) mostrado en la Figura 3-3, se presenta un análisis más detallado que expone la distribución de las transacciones por modo y por tipo de día (Ver Tabla 3-6).

Para este análisis el modo bus, considera aquellas transacciones que utilizaron bus (pudiendo haber realizado la transacción tanto en bus como en zona paga) y se obtienen de sumar las transacciones realizadas en servicios troncales y alimentadores.

Tabla 3-6: Transacciones según modo y tipo día

Tipo día	Bus	Metro	Total
Laboral promedio	3.719.539	2.099.915	5.819.454
Sábado	2.433.645	1.239.532	3.673.177
Domingo	1.105.464	638.023	1.743.487
Total	7.258.648	3.977.470	11.236.118

En día laboral un 63,9% de las transacciones utiliza bus, mientras que el 36,1% metro, para día sábado los porcentajes varían levemente, inclinándose más hacia bus con un 66,3% y un 33,7% para metro, en cambio para día domingo la partición modal se acerca más a la de un día laboral, 63,4% bus y 36,6% metro.

3.2.1.2 Análisis horario de transacciones, por tipo de día.

Las características de los viajes varían a lo largo del día debido a la naturaleza dinámica de la demanda de transporte. En consecuencia, se hace necesario analizar cada tipo de día por separado, y establecer la distribución de las transacciones a lo largo del día.

Se debe tener claro que no es factible realizar una comparación directa con los resultados de la encuesta Origen–Destino del año 2001 (EOD 2001), ya que la dinámica del sistema y la naturaleza de la información de ese momento, era distinta a la que se presenta hoy en el nuevo sistema.

Sin embargo se estableció una manera que permite realizar una comparación gruesa con dichos resultados, mediante una estimación del número de transacciones que implicaría la estructura de viajes expuestos en la distribución de viajes según modo resultantes de la EOD 2001¹⁶, donde se detalla la composición por modo de estos.

3.2.1.2.1 Día Laboral

La distribución horaria de las transacciones para este tipo de día, se muestra en la Figura 3-6, donde cada intervalo corresponden a 15 minutos que contienen todas las transacciones que se realizaron en ese segmento de tiempo.

Se puede observar que existe una mayor concentración de viajes entre las 07:00 y 09:00 hrs lo que corresponde al período punta mañana. Este peak es más acentuado que el alcanzando entre las 17:00 y 20:00 que se puede identificar como punta tarde. También se observa una pequeña punta mediodía entre las 12:00 y 14:00 hrs. Es factible identificar un valle regular entre los períodos punta mañana y punta mediodía, mientras que en la tarde comienza un aumento paulatino desde las 15:00 hrs hasta llegar al período punta.

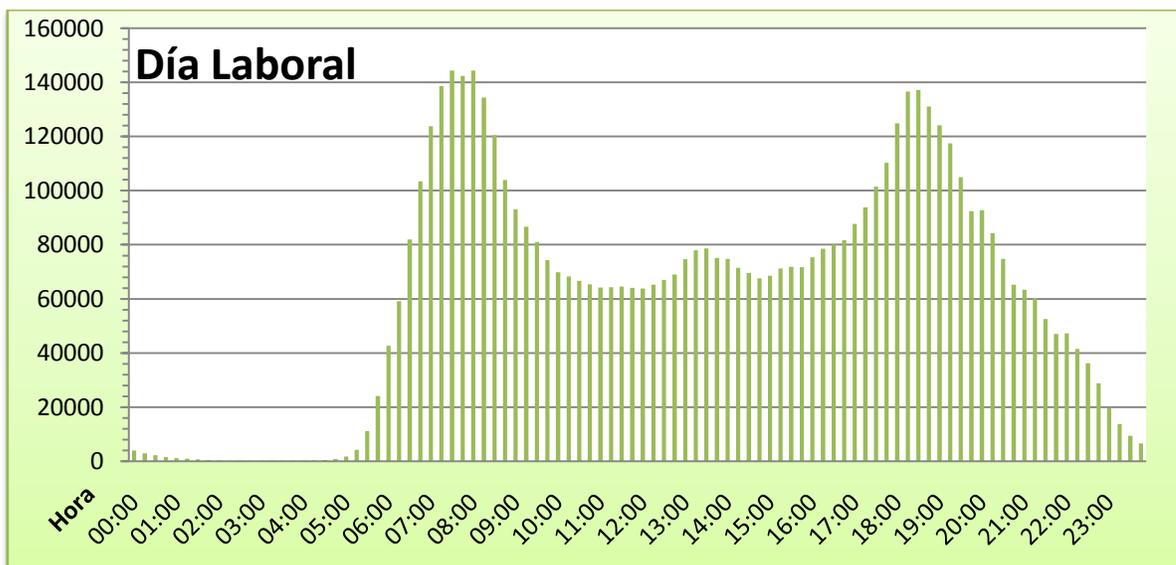


Figura 3-6: Distribución horaria de las transacciones en día laboral

En la Figura 3-7 se presenta la distribución horaria de las transacciones según tipo de tarjeta y pago.

¹⁶ Índices y valores obtenidos a partir del Informe Final EOD2001.

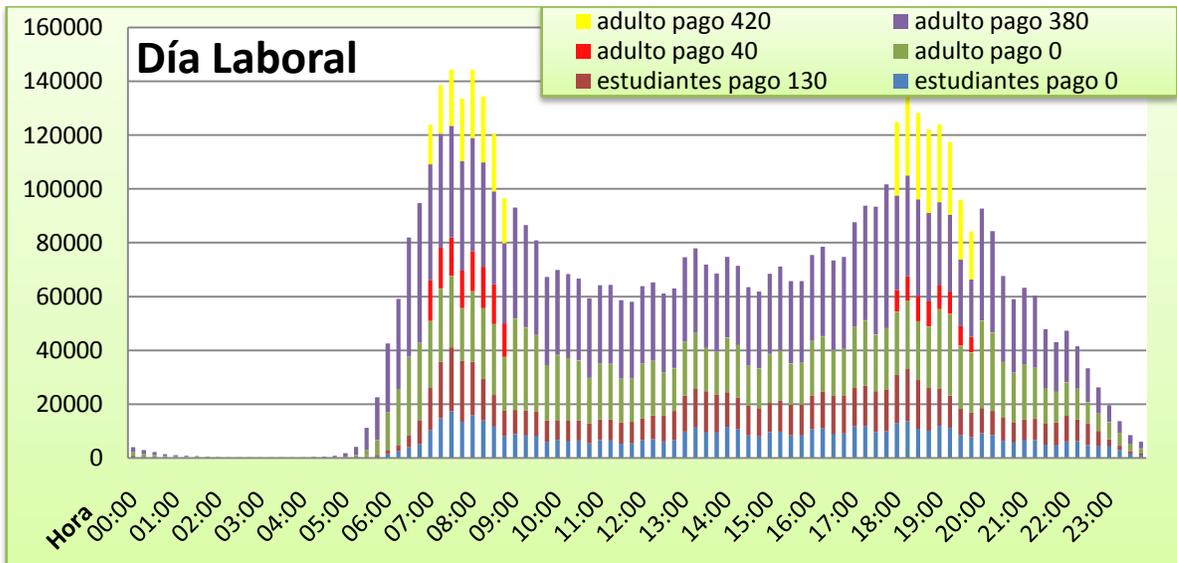


Figura 3-7: Distribución horaria de las transacciones en día laboral según pago y tipo de tarjeta

Como se aprecia en la Figura 3-7 en relación a las transacciones tipo estudiante estas presentan la hora punta mañana mucho más marcada que el tipo adulto y a partir de las 13:00 comienza un constante e irregular aumento hasta aproximadamente las 20:00. Por otro lado en las transacciones correspondientes a adulto, en rojo se aprecia las transacciones que pagan un diferencial ya que vienen de bus y se suben a metro, y en amarillo aquellas transacciones que son en metro horario punta. Los pagos cero se mantienen en proporción para todas las transacciones tipo adulto.

Finalmente en la Figura 3-8 se muestra la distribución horaria de las transacciones por modo (bus y metro). Donde se puede apreciar que la cantidad de transacciones realizadas en metro se mantiene regular excepto en los horarios punta donde aumenta la participación de metro en el total de transacciones.

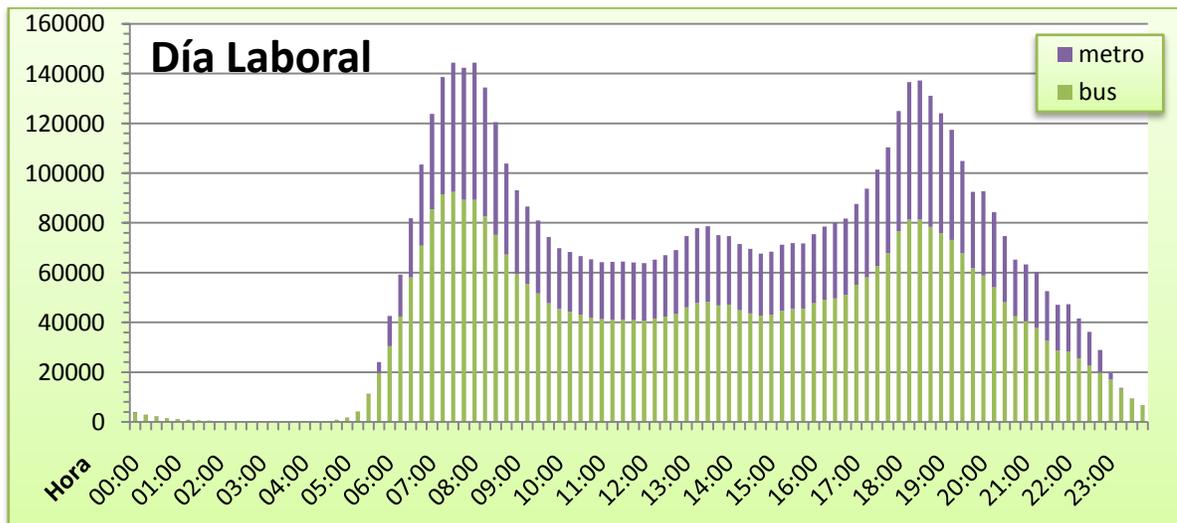


Figura 3-8: Distribución horaria de las transacciones en día laboral por modo

En todas las figuras anteriores se puede distinguir claramente dos períodos punta: punta mañana y punta tarde, siendo más marcado el primero. También se puede apreciar el período punta mediodía, el cual se presenta menos pronunciado que los anteriores. Además, se puede verificar que la forma de la curva se conserva en general y que los períodos no punta se comportan de manera bastante regular durante el día.

En la Tabla 3-7 se muestra la estimación del total de transacciones estimadas a partir de la información de viajes por modo de la EOD 2001 para día laboral en temporada normal.

Tabla 3-7: Estimación de transacciones de viajes EOD, día laboral

Modo		N° de Viajes EOD	Transacciones
Bus		3.674.005	3.674.005
MetroBus		241.632	241.632
Bus	Bus	262.511	525.022
MetroBus	Bus	42.752	85.504
Metro Solo		370.800	370.800
Auto Chofer	Metro	8.551	8.551
Auto Acompañante	Metro	21.049	21.049
Bus o MetroBus	Metro	177.859	355.718
Taxi Colectivo	Metro	68.434	68.434
Taxi	Metro	10.184	10.184
Metro	Tren	2.894	2.894
Otros	Metro	14.845	14.845
Auto Acompañante	Bus o MetroBus	25.192	25.192
Bus o MetroBus	Taxi Colectivo	36.667	36.667
Bus o MetroBus	Taxi	6.620	6.620
Total		4.963.995	5.447.117

Utilizando esta aproximación del total de transacciones que se obtendría de una demanda como la registrada en la EOD 2001 es posible comparar en orden de magnitud la cantidad de transacciones que existen en día laboral con las del actual sistema. Esta estimación arroja un valor de casi 5,5 millones, algo inferior a los 5,8 millones que se obtiene a partir de los datos actuales.

Se observa una variación en la distribución de transacciones por modo, donde las transacciones en modo bus disminuyen de un 84,4% a 63,9%, mientras que para metro aumentan de un 15,6% a 36,1%, esto se traduce en un 20,5% de transacciones que dejan de utilizar bus y se suben a metro (ver Tabla A-4, Anexo A).

3.2.1.2.2 Sábado

Las transacciones realizadas en un día sábado presentan una disminución de un 36,9% respecto de las transacciones realizadas en un día laboral (Ver Tabla 3-3). En cuanto a la distribución horaria de las transacciones (Figura 3-9), ésta es diferente en cuanto a la forma de la curva, ya que no se observan períodos punta muy marcados, observándose que existe mayor cantidad de transacciones durante la tarde, alcanzándose el peak cerca de las 14:00 hrs.

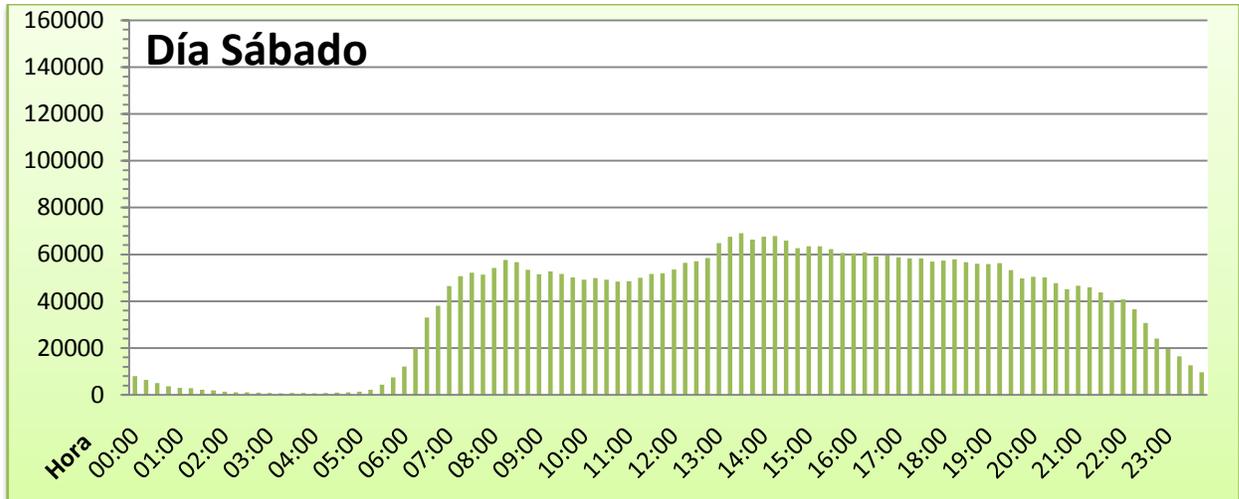


Figura 3-9: Distribución horaria de las transacciones en día sábado

En la Figura 3-10 se puede observar que la forma de la curva de demanda se mantiene, además se aprecia que las transacciones tipo estudiante disminuyen su participación sin presentar claros horarios punta y que se presenta un mayor número de transacciones alrededor del mediodía. También se ve que las transacciones correspondientes a adulto aumentan levemente entre las 13:00 y 15:00 hrs y que existe un mayor número de transacciones a partir del mediodía en comparación con la mañana.



Figura 3-10: Distribución horaria de las transacciones en día sábado según pago y tipo de tarjeta

La distribución horaria de las transacciones por modo (bus y metro) se presenta en Figura 3-11, existiendo mayor número de transacciones en bus que en metro, además se puede apreciar una pequeña punta mañana entre las 07:00 y 09:00 hrs y una punta medio día un poco más pronunciada para metro que bus entre las 13:00 y 15:00 hrs.

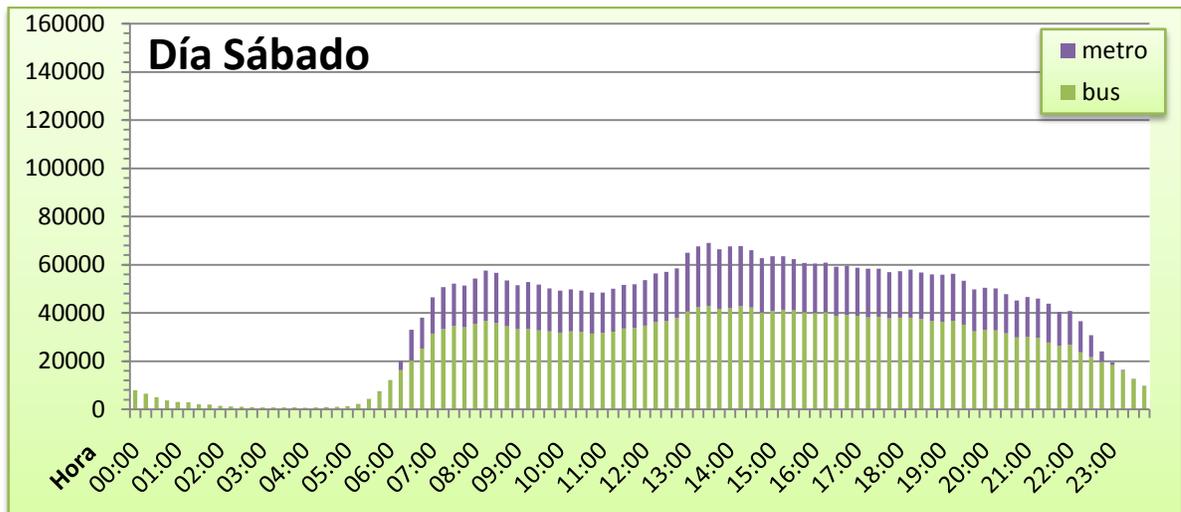


Figura 3-11: Distribución horaria de las transacciones en día sábado por modo

En la Tabla 3-8 se muestra la estimación del total de transacciones estimadas a partir de viajes por modo de la EOD 2001, para día sábado en temporada normal.

Tabla 3-8: Viajes EOD llevados a transacciones, día Sábado

Modo		N° de Viajes EOD	Transacciones
Bus		2.644.979	2.644.979
MetroBus		224.440	224.440
Bus	Bus	247.849	495.698
MetroBus	Bus	31.491	62.982
Metro	Solo	199.740	199.740
Auto Chofer	Metro	2.573	2.573
Auto Acompañante	Metro	10.930	10.930
Bus o Metro Bus	Metro	80.466	160.932
Taxi Colectivo	Metro	21.744	21.744
Taxi	Metro	5.057	5.057
Otros	Metro	17.694	17.694
Auto Acompañante	Bus o MetroBus	19.301	19.301
Bus o Metro Bus	Taxi Colectivo	24.876	24.876
Bus o Metro Bus	Taxi	3.416	3.416
Total		3.534.556	3.894.362

Así se puede ver que los 3.673.177 de transacciones que se realizaron en día sábado tienen relación con los 3.894.362 de transacciones convertidas a partir de los viajes de la EOD, mostrando una leve disminución en la cantidad de transacciones realizadas.

Para el día Sábado se contempla, al igual que para el día laboral, una variación en la distribución de transacciones por modo (levemente mayor a la apreciada en día laboral), donde las transacciones en modo bus disminuyen de un 89,2% a 66,3%, mientras que para metro aumentan de un 10,8% a 33,7%, esto se traduce en un 23% de transacciones que dejan de utilizar bus y se suben a metro (ver Tabla A-5, Anexo A).

3.2.1.2.3 Domingo

Las transacciones realizadas en un día domingo son 1.743.487, presentando una disminución de un 70% respecto de las transacciones realizadas en un día laboral, y una baja en un 52,5% con respecto a las transacciones realizadas en el día sábado. En cuanto a la distribución horaria de las transacciones (Figura 3-12), ésta es diferente en cuanto a la forma de la curva de los días estudiados anteriormente, siendo más plana y regular, no mostrando períodos punta claros.

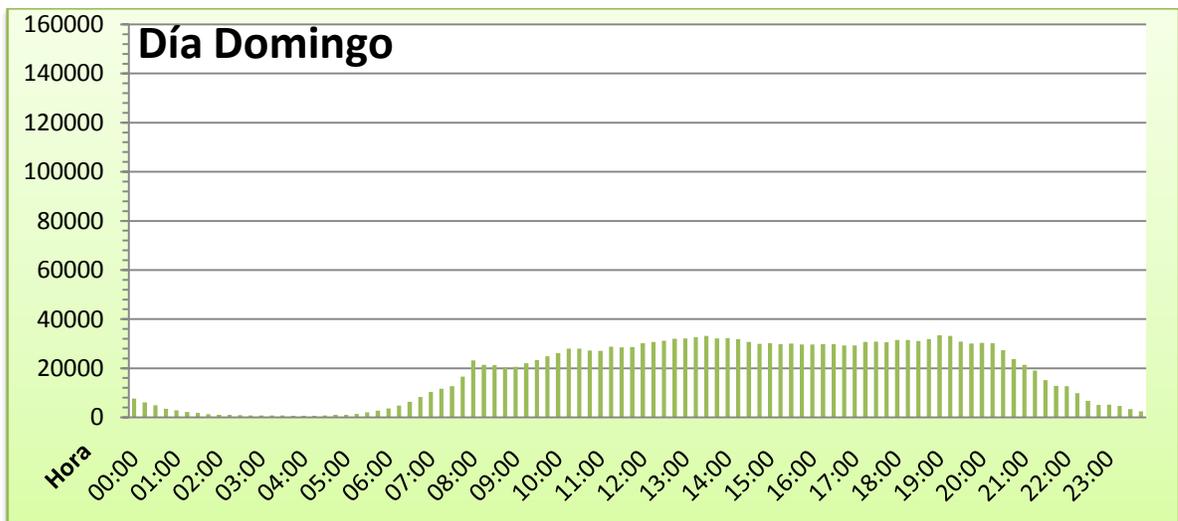


Figura 3-12: Distribución horaria de las transacciones en día domingo

En la Figura 3-13 se aprecia que las transacciones tipo estudiante disminuyen su participación no presentando claramente horarios punta. También se ve que las transacciones correspondientes a adulto se comportan de manera bastante regular entre las 10:00 y las 21:00 hrs.



Figura 3-13: Distribución horaria de las transacciones en día domingo según pago y tipo de tarjeta

La distribución horaria de las transacciones por modo (bus y metro) se presenta en la Figura 3-14, existiendo mayor número de transacciones en bus que en metro, además se puede apreciar que la cantidad de transacciones hechas en bus aumenta gradualmente hasta el medio día manteniéndose durante la tarde y disminuyendo a partir de las 20:00 hrs aproximadamente. En tanto para metro se presenta un pequeño peak local a las 08:00 hrs, aumentando paulatinamente el número de transacciones realizadas en él a lo largo del día alcanzando el máximo de transacciones a las 19:00 hrs.

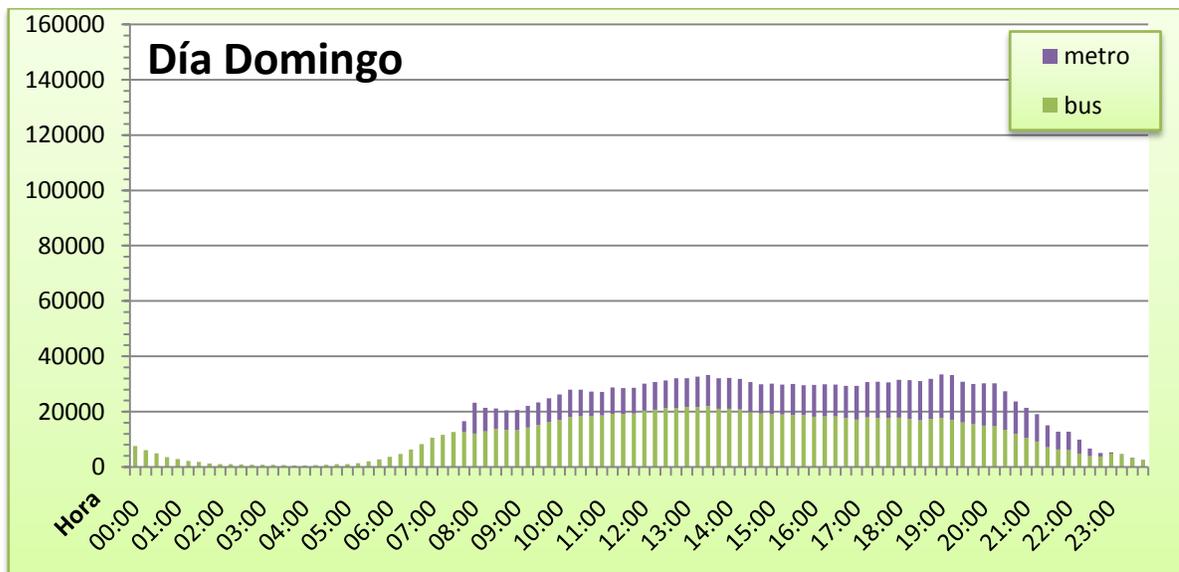


Figura 3-14: Distribución horaria de las transacciones en día domingo por modo

En la Tabla 3-9 se muestra la estimación de transacciones realizada a partir de la distribución de los viajes según modo obtenida de la EOD 2001, para día domingo en temporada normal.

Tabla 3-9: Viajes EOD llevados a transacciones, día Domingo

Modo	N° de Viajes EOD	Transacciones
Bus	2.023.916	2.023.916
MetroBus	191.469	191.469
Bus Bus	228.723	457.446
MetroBus Bus	36.035	72.070
Metro Solo	96.794	96.794
Auto Chofer Metro	1.289	1.289
Auto Acompañante Metro	852	852
Bus o Metro Bus Metro	38.185	76.370
Taxi Colectivo Metro	6.672	6.672
Otros Metro	5.605	5.605
Auto Acompañante Bus o MetroBus	13.810	13.810
Bus o Metro Bus Taxi Colectivo	29.995	29.995
Bus o Metro Bus Taxi	10.195	10.195
Total	2.683.540	2.986.483

Así se puede ver que los 1.743.487 de transacciones que se realizaron en día domingo son casi la mitad de 2.986.483 de transacciones estimadas a partir de los viajes de la EOD 2001.

Si se comparan las transacciones obtenidas de la base de datos y las transacciones que se obtienen a partir de la Tabla 3-9 en base a la EOD 2001 (ver Tabla A-6, Anexo A), existe un 30,3% de las transacciones que cambian de bus a metro, produciéndose una mayor redistribución en comparación a los otros días.

3.2.1.3 Transacciones sin posición

Del total de transacciones, 510.195 de ellas no poseen posición asociada lo cual corresponde a un 1,5%, este valor es bajo dado el volumen de información disponible, pero se traduce directamente en pérdida de información sobre el origen del viaje, así como también influye en la estimación de la posición de bajada ya que esta no es factible de obtener si no se tiene la posición de la siguiente transacción. En consecuencia las transacciones sin posición afectan el conocer la posición de la subida o bajada del viaje.

De las transacciones que no poseen posición, 72.452 (14,2%) se realizan en 35 zonas pagas afectando a 30.261 tarjetas. La falta de posición se explica porque 27 de las 35 zonas pagas no presentan posición asociada al paradero o estación de transbordo ó no presentan código de parada, y en las restantes 8 se da la paradoja que presentan dos posiciones, pero no se asigno ninguna. Este criterio, en el futuro debería ser mejorado.

Las restantes 437.743 (85,8%) transacciones que no muestran posición se realizan en 1719 buses (29% del total de buses con transacciones) y afectan a 370.585 tarjetas. Esta falta de posición se debe a la pérdida de información de posicionamiento GPS.

De los 1719 buses que tienen transacciones sin posición, 203 no se encuentran en la tabla de posicionamiento, es decir la falta de posición para las transacciones realizadas en esos buses se debe a que el gps no funcionó a lo largo de la semana. Para los restantes 1516 buses no es directo, dado el procesamiento de la información disponible, distinguir si la falta de posicionamiento se debe a una intermitencia en las emisiones de posición, perjudicando a algunas de las transacciones realizadas en dicho bus, o bien es una falla prolongada en un periodo de tiempo que afecta a la totalidad de transacciones realizadas en ese bus en dicho periodo de tiempo.

Una primera aproximación para distinguir el motivo de falta de posicionamiento, se obtiene de contar el número de transacciones sin posición diarias por patente y eliminar las patentes que se repiten a lo largo de la semana de manera de conocer el número de buses que tienen cierto rango de transacciones sin posición diarias, lo cual se muestra en la Tabla 3-10.

Tabla 3-10: N° de buses que realizan transacciones sin posición-día.

N° de transacciones sin posición por día	N° de patentes	N° buses
0 - 10	1248	964
11 - 30	500	435
31 - 50	155	137
51 - 100	222	192
101 - 200	280	215
201 - 300	190	143
301 - 400	124	100
401 - 500	95	83
> 500	294	151

Como se puede observar en la Tabla 3-10, un gran número de buses presenta entre 0 y 10 transacciones diarias sin posición. Es altamente probable que estas transacciones se deban a fallas intermitentes del gps, ya que de ser una falla prolongada afectaría un mayor número de transacciones. También se observa que un número bajo de buses presentan un alto número de transacciones diarias sin posición (> a 500 transacciones diarias) lo que correspondería a fallas prolongadas de emisiones de posición. De este análisis, se puede inferir que un número importante de buses presenta fallas intermitentes afectando de manera aleatoria a las transacciones, mientras que un número menor de buses presenta fallas más prolongadas (se deduce de la acumulación de transacciones perjudicadas por bus) que daña directamente a la secuencia de transacciones realizadas en el bus en cierto periodo de tiempo.

Para establecer un análisis más certero sobre la causa de la falta de posicionamiento se debe, en el futuro, determinar los intervalos de tiempo en que el bus no presenta datos de posición

Dado que la pérdida de información está directamente relacionada con el bus, y estos con las unidades de negocio, se presenta en la Tabla 3-11 las transacciones sin posición por unidad de negocio.

Tabla 3-11: Número de transacciones sin posición realizados en intervalos de tiempo, día viernes según tipo de tarjeta

Unidad de Negocio	N° transacciones sin posición	N° transacciones totales	% de Transacciones Totales
Metro	0	12.377.129	0
T1	63.370	2.886.098	2,2
T2	114.812	4.119.307	2,8
T3	65.569	1.592.807	4,1
T4	94.261	4.147.276	2,3
T5	35.655	2.441.980	1,5
B	36.140	811.558	4,5
C	12.359	962.653	1,3
D	6.581	756.114	0,9
E	1.294	347.537	0,4
F	11.995	1.200.910	1,0
G	31.784	363.520	8,7
H	166	207.444	0,1
I	23.589	1.299.356	1,8
J	12.620	1.000.244	1,3
Total Transacciones	510.195	34.513.933	

Como se puede observar en la Tabla 3-11, la troncal 3, y los alimentadores de las zonas B y G destacan por tener un mayor porcentaje de sus transacciones, sin posición, es decir sus buses presentan mayor número de fallas en las emisiones de posición. Estas unidades de negocio eran, según información de Transantiago para el periodo de estudio, las unidades de negocios que tenían buses más antiguos, que presentaban un bajo índice de cumplimiento y que estaban peor evaluadas en el sistema. Con lo cual si se puede identificar un sesgo por unidad de negocio.

La distribución de las transacciones sin posición a lo largo de la semana se muestra en la Figura 3-15.

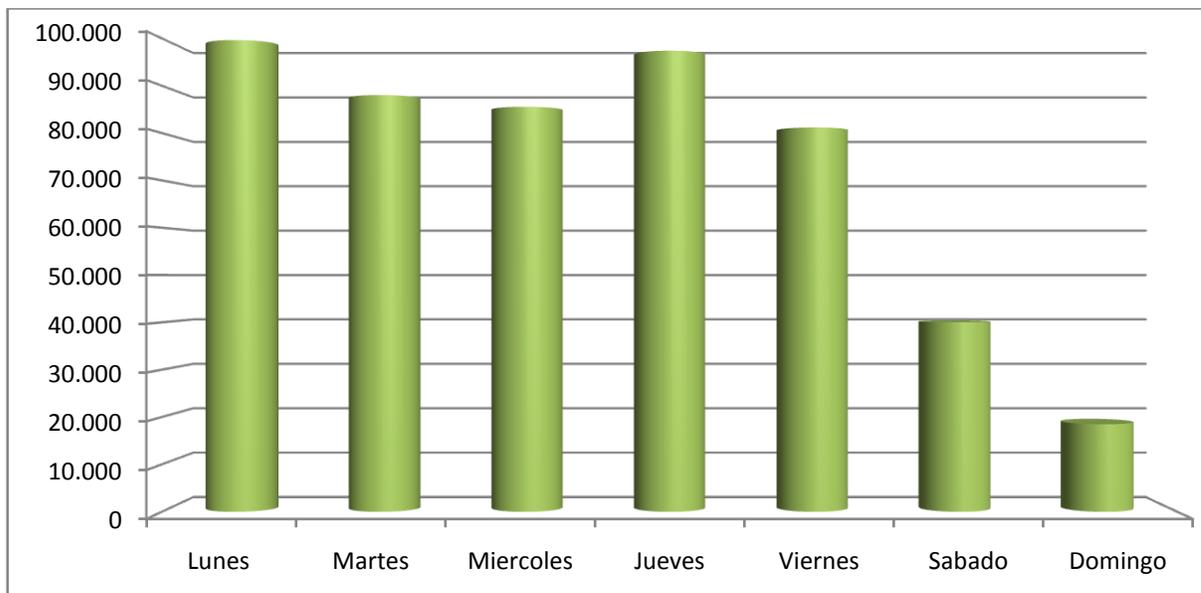


Figura 3-15: Transacciones sin posición por día

Como se puede ver en la Figura 3-15, los días hábiles presentan una mayor pérdida de posición que los días de fin de semana (sobre 80.000 transacciones por día), habiendo una mayor falta de información el día lunes (99.616 transacciones, correspondiente al 1,7% de transacciones del día) y el día jueves (97.366 transacciones, correspondiente al 1,6% de transacciones del día).

En tanto, el día sábado presenta 40.036 transacciones sin posición (0,7% del total de transacciones del día) y el día domingo disminuye esta pérdida a 18.496 transacciones (lo que corresponde a un 0,3% del total de transacciones del día) (ver Tabla A-7, Anexo A).

Para ver si existe algún tipo de patrón horario en la falta de posición de las transacciones, se hizo un histograma para las transacciones sin posición a lo largo del día, durante todos los días de la semana (Figura 3-16).

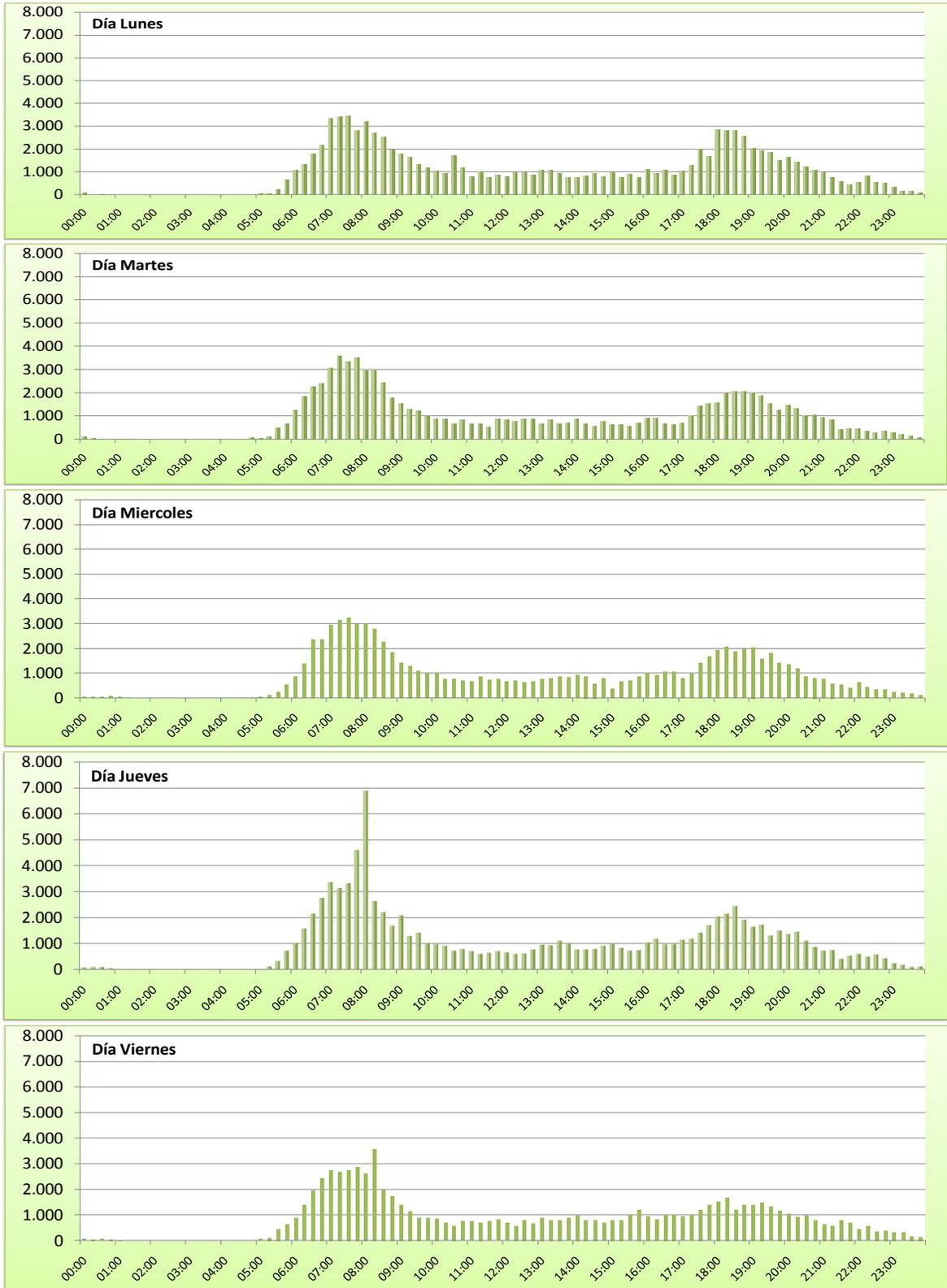


Figura 3-16: Transacciones sin posición a lo largo del día, para todos los días



Figura 3-16: Transacciones sin posición a lo largo del día, para todos los días (continuación)

Como se aprecia en la Figura 3-16, en términos generales la pérdida de información mantiene las formas de las curvas vista en la Figura 3-6, Figura 3-9, Figura 3-12, presentando algunas singularidades.

El día jueves se observa una mayor pérdida de información cercano a las 8 am (en el intervalo entre las 07:45 y las 07:59 hubieron 4.616 transacciones sin posición, mientras que para el intervalo entre las 08:00 y 08:14 aumentaron a 6.914).

También se identifica un aumento en relación a los otros días, particularmente el día viernes cercano a las 8 am (entre las 8:15 y las 8:29 aumentaron a 3.592 transacciones sin posición), y el día lunes tanto en horario punta mañana y en el horario punta tarde, donde se presenta mayor número de transacciones sin posición que el resto de los días.

Debido a lo anterior se analizará en detalle los días lunes, jueves y viernes, en las horas cercanas a las singularidades expuestas, de manera de ver la posibilidad de establecer algún patrón de pérdida. Para ello se examinará la información de manera agregada, según tipo de tarjeta y según tipo de pago.

Tabla 3-12: Número de transacciones sin posición realizados en intervalos de tiempo, día lunes según tipo de tarjeta

Lunes			
hora inicio	horario fin	estudiantes	adulto
07:00	07:14	725	2.637
07:15	07:29	893	2.539
07:30	07:44	861	2.608
08:00	08:14	732	2.456
10:30	10:44	309	1.414
18:00	18:14	632	2.219
18:15	18:29	614	2.196
18:30	18:44	503	2.311
18:45	18:59	491	2.094
19:00	19:14	353	1.691

Tabla 3-13: Número de transacciones sin posición realizados en intervalos de tiempo, día jueves según tipo de tarjeta

Jueves			
hora inicio	horario fin	estudiantes	adulto
07:45	07:59	1.132	3.484
08:00	08:14	1.616	5.298
09:00	09:14	415	1.686

Tabla 3-14: Número de transacciones sin posición realizados en intervalos de tiempo, día viernes según tipo de tarjeta

Viernes			
hora inicio	horario fin	estudiantes	adulto
08:15	08:29	715	2.877

Se observa que el día lunes la variación de transacciones entre los intervalos de tiempos que se muestran en la Tabla 3-12, para las tarjetas tipo estudiante, son menores en comparación a los observados para las tarjetas tipo adulto, que presenta una mayor variación de número de transacciones en los intervalos de tiempo.

Para el día jueves en tanto, las tarjetas tipo estudiante en general mantienen su baja variación exceptuando algunas singularidades en la cercanías de la 08:00 hrs, mientras que las tarjetas tipo adulto presenta mayor variación en los intervalos de tiempo que se aprecian en la Tabla 3-13, finalmente para el día viernes las tarjetas tipo adulto muestran una singularidad bastante marcada en el intervalo descrito en la Tabla 3-14.

Dado lo anterior, se puede afirmar que existe una pérdida de información en las cercanías de las 8 am, común para las tarjetas tipo adulto principalmente los días lunes, jueves y viernes, además del día lunes desde las 18:00 a 19:00 hrs.

A continuación se muestra el detalle realizado según tipo de pago.

Tabla 3-15: Número de transacciones sin posición realizados en intervalos de tiempo, día lunes según pago

Lunes				
hora inicio	horario fin	pago 0	pago 130	pago 380
07:00	07:14	1.352	422	1.588
07:15	07:29	1.231	537	1.664
07:30	07:44	1.468	480	1.521
08:00	08:14	1.144	431	1.613
08:15	08:29	1.231	268	1.203
10:30	10:44	721	183	819
10:45	10:59	490	111	601
18:00	18:14	1.156	288	1.407
18:15	18:29	1.190	285	1.335
18:30	18:44	1.362	209	1.243
18:45	18:59	1.303	189	1.093

Tabla 3-16: Número de transacciones sin posición realizados en intervalos de tiempo, día jueves según tipo de tarjeta

Jueves				
hora inicio	horario fin	pago 0	pago 130	pago 380
07:45	07:59	1.910	592	2.114
08:00	08:14	2.526	919	3.469

Tabla 3-17: Número de transacciones sin posición realizados en intervalos de tiempo, día viernes según tipo de tarjeta

Viernes				
hora inicio	horario fin	pago 0	pago 130	pago 380
08:15	08:29	1599	374	1619

En la Tabla 3-15 se pueden observar las particularidades del día lunes cercanas a las 08:00 hrs, media mañana y en horario punta tarde, donde se puede ver, que para los pagos cero y 380 existe un mayor número de transacciones sin posición. Para el día jueves, la Tabla 3-16 muestra que en los intervalos inmediatos a las 8 am, existe un mayor aumento en los pago \$380, que en los pagos cero y \$130. En la Tabla 3-17 se

muestra el intervalo del día viernes analizado, el cual presenta un aumento de transacciones sin posición para el pago cero y \$380 entre las 8:15 y 8:29.

Se puede concluir que existe una mayor pérdida de información de posición en las cercanías de las 8 am el cual afecta de modo transversal el pago, principalmente para los días lunes y jueves, mientras que para el día viernes en el mismo horario se ven mayormente perjudicados el pago cero y tarifa normal, no identificando un sesgo por pago. También se identifica un intervalo de pérdida más pronunciado para el día lunes cercano a las 10:30 hrs y en la tarde desde las 18:00 a 19:00 hrs, principalmente para pago cero y \$380. No es claro identificar un sesgo horario en la información, ya que las pérdidas aumentan en horarios punta, mostrándose porcentual al número de transacciones no presentando diferencias significativas.

3.2.2 Tabla Posicionamiento

Del total de observaciones un 91,7% corresponde a emisiones realizadas con el motor del bus encendido útiles para el seguimiento de los buses. De esta cantidad de observaciones un 61,4% de ellas presenta asignación de servicio, lo que se traduce directamente en información más completa que permite un mayor conocimiento sobre el comportamiento de las tarjetas.

Del total de buses presentes en la tabla de posicionamiento (Ver Tabla 3-2), 5725 emiten observaciones con ignición, los restantes 12 buses sólo emiten registros de posición con motor apagado.

Existen algunos registros que resultan no tener explicación lógica y que identifican errores en la base de posicionamiento, por ejemplo, aquellos casos donde el motor del bus está apagado (ignición No) pero el bus presenta asignación de servicio (Ver Tabla 3-18), ó presenta velocidad estando apagado (Ver Tabla 3-19).

Tabla 3-18: N° de registros con/sin ignición y con/sin servicio.

	Con Servicio	Sin Servicio	Total
Ignición Si	42.725.031	26.880.827	69.605.858
Ignición No	64.209	6.212.267	6.276.476
Total	42.789.240	33.093.094	75.882.334

También se presentan casos en los cuales se puede tener más de una causa para estas “anormalidades”, como cuando el bus está con motor prendido y no presenta velocidad (Ver Tabla 3-19). Lo cual se puede explicar cómo emisión de observaciones GPS en momentos en que el bus se encuentra detenido tomando pasajeros, o en momentos donde el bus se encuentra en terminal a la espera de salir a servicio.

Tabla 3-19: Casos especiales, ignición.

	N° registros	%
Ignición No, velocidad > 0	90.897	0,1
Ignición Si, velocidad = 0	29.342.416	38,7

En la Tabla 3-20 se presenta el número de buses operando (con ignición) que tenía cada concesionario durante la semana y el número total de observaciones de posición que emitieron esos buses. Destacándose Buses Gran Santiago S.A (T3), Express de Santiago Uno S.A (T4), Subus Chile S.A. (T2) en número de buses y emisiones de registros de posición.

Tabla 3-20: Distribución de buses numero de observaciones por concesionario.

Concesionario		N° buses	N° Observaciones
Nombre	Unidad de Negocio		
Inversiones Alsacia S.A.	T1	614	6.784.613
Subus Chile S.A.	T2	792	10.556.616
Buses Gran Santiago S.A.	T3	801	8.502.560
Express de Santiago Uno S.A.	T4	895	9.565.887
Buses Metropolitana S.A.	T5	493	5.790.217
Buses Gran Santiago S.A.	Zona B	272	3.476.316
Redbus Urbano S.A.	Zona C	286	3.878.590
STP Santiago S.A.	Zona D	193	2.618.361
Union de Transporte S.A.	Zona E	180	2.698.742
STP Santiago S.A.	Zona F	273	3.807.471
A5	Zona G	136	1.348.024
Trans Araucarias S.A.	Zona H	164	2.316.386
Comercial Nuevo Milenio S.A.	Zona I	397	5.024.029
Comercial Nuevo Milenio S.A.	Zona J	229	3.238.046

Al revisar la base de datos, se encontró que existían patentes que presentaban más de un servicio asignado dentro de un mismo período, es por eso, que se confeccionó la Tabla 3-21, donde se muestra la cantidad de observaciones emitidas con motor encendido y que presentan más de un servicio asignado.

Tabla 3-21: Número de registros con ignición con cero, uno ó más servicios asignados.

N° de servicios asignados a un registro	N° de registros (Ignición Si)
0	26.880.827
1	42.102.443
2	620.605
3	1.960
4	23

Se observa que un 38,6% de los registros con ignición no poseen servicio asignado, un 60,5% tiene asignado un servicio, un 0,9% tiene asignado dos, los demás casos son menores al 0,003%.

Al observar el intervalo de tiempo entre registros de posicionamiento de los buses, se detectaron que dichos intervalos no son regulares para la totalidad de los registros emitidos. En la Tabla 3-22, se expone el detalle de los intervalos de tiempos encontrados en la base de datos de posicionamiento, junto con el total de registros emitidos con el motor encendido que presenta cada intervalo.

Tabla 3-22: Número de registros según intervalos de tiempo entre registros.

Intervalo de tiempo entre registros de posición [seg]	Total
0	7.259
1	10.774
9	72.435
10	8.786.335
11	81.723
12	13.882
20	45.613
23	5.353
25	5.692
28	15.678
29	1.066.579
30	57.665.820
31	1.294.287
32	24.265
60	44.531
90	14.202
299	810
300	1.162
301	347

Como se observa en la Tabla 3-22, usualmente se emiten registros de posición cada 30 segundos, dicho intervalo presenta el 82,8% del total de registros de posición con ignición. También se destacan los intervalos de 10, 29 y 31 segundos con 12,6%, 1,5% y 1,9% de registros respectivamente. Los restantes intervalos presentan 1,1% de los registros. En Anexos se encuentra el desglose de la Tabla 3-22 por unidad de negocio.

Cabe mencionar que los registros de aquellos intervalos, menores o cercanos a 30 segundos presentan mayor precisión en la trayectoria del bus. En cambio, para aquellos intervalos cuyas emisiones son más distanciadas en el tiempo, se pierde precisión y certeza sobre la trayectoria que el bus realizó.

3.2.3 Otras

Existen para este período, según la base de datos de paraderos, 10.218 paradas en el sistema, de las cuales 6.717 eran para alimentadores y 3.501 para troncales. Además hay 36 estaciones de transbordo (tipo de parada) y existen 159 zonas paga, de las cuales solo 149 presentan asignación de validadores, lo que significa que durante la semana en estudio un 93,7% de las zonas pagas estaban operativas, funcionando principalmente en horas punta. En cuanto a metro, este tiene 5 líneas con un total de 92 estaciones, de las cuales 7 son estaciones de combinación para las líneas de metro.

3.3 Síntesis

La información disponible presenta fortalezas en relación a la cantidad de información que entrega y a la calidad de esta. En tanto, las debilidades de la información derivan de características de la implementación de la tarjeta, ya que no es personal y es transferible, pudiendo un usuario tener más de una tarjeta lo que se traduce en que no se puede establecer la relación una tarjeta un usuario.

Las bases de datos permiten obtener gran cantidad de información completa sobre el origen (posición y hora) de 34.003.738 de etapas de viaje (98,5% del total de transacciones), y además conocer qué servicio abordó en 28.766.938 de ellas. Este número no se traduce directamente en posibles estimaciones de punto de bajada, ya que las transacciones sin posición afectan la realización de la estimación de aquellas etapas de viaje en que la transacción posterior es una transacción sin posición.

Existen 20.343.686 transacciones realizadas en bus, de las cuales 19.905.943 tienen posición. Las transacciones sin posición se deben a errores en la información de posicionamiento del bus, lo que deriva en que no es posible realizar el seguimiento de este y por lo tanto de la tarjeta. El total de etapas realizadas en bus es la suma de las transacciones realizadas en bus y zona paga, sumando 22.136.804 etapas de viaje

para las cuales la estimación de bajada que se puede obtener cuenta con posición y tiempo debido a la información gps.

La posición de las transacciones realizadas en metro, no presenta errores ya que la ubicación de las estaciones de metro es fija y no se presentan fallas de información en relación a la ruta de este. Por lo que el total de etapas de viaje realizadas en metro (12.377.129) son potencialmente candidatas a obtener información sobre el punto de bajada. Está información es precisa en relación a la posición del punto de bajada, pero se debe realizar una estimación para obtener el tiempo de viaje en metro.

Dada la dinámica de las zonas pagas, las etapas de viaje iniciadas en ellas (1.793.118) presentan un grado de incertidumbre sobre el servicio que se abordó, lo cual afecta directamente a las posibilidades de obtener una correcta estimación del punto de bajada. Además existe un pequeño número de transacciones (72.452) que no presentan posición de subida, debido a error de la posición del paradero.

En relación a la distribución de las transacciones a lo largo del día, se distinguen claramente los horarios punta mañana (07:00 – 09:00 hrs) y punta tarde (17:00 – 20:00 hrs), y en menor medida una pequeña punta mediodía (12:00 – 14:00 hrs) para los días laborales. Lo que permitiría en el futuro realizar un análisis por horario del comportamiento de las tarjetas, estableciendo parámetros según períodos del día.

Para el día sábado la distribución cambia existiendo mayor demanda durante la tarde y distinguiéndose una pequeña punta mediodía comprendida entre las 13:00 y 15:00 hrs. En tanto para el día domingo no se observan horarios punta pronunciados, distribuyéndose las transacciones a lo largo del día de manera plana y regular.

El intervalo de tiempo entre registros de posicionamiento de los buses es un indicador de precisión de estos datos, se considera que intervalos superiores a 60 segundos generan información imprecisa y que no permiten realizar un correcto seguimiento del bus (0,02% del total de observaciones de posicionamiento), debiendo identificar aquellos buses que presentan dicho comportamiento especificando el horario en que se produce este tipo de registros.

La calidad de los datos es un factor importante que se debe conocer para el posterior trabajo que se desea realizar, ya que permite tener una estimación sobre el real alcance de la metodología a plantear. Por esto resulta de gran importancia conocer las bondades y fallas de la información con que se trabaja, lo cual se estudió en éste capítulo.

4 Metodología para la Extracción y el Procesamiento de la Información: Exploración, procedimiento y criterios propuestos.

En este capítulo se presenta el desarrollo de este trabajo, el cual deriva en el planteamiento de la metodología para la extracción y el procesamiento de la información proveniente de la operación del sistema de transporte público de la ciudad de Santiago. El capítulo se encuentra dividido en dos partes.

Primero se realiza un análisis exploratorio exhaustivo sobre una pequeña muestra de tarjetas que permite identificar los principales escenarios que aparecen en la información entregada, determinar la factibilidad de estimar bajada en cada etapa de viaje, identificar parámetros que ayuden a reconocer transbordo, así como también etapas de viaje, calcular tiempos de viaje, transbordo, espera y caminata, y establecer a partir de estos, criterios que permitan verificar la estimaciones realizadas.

Luego se plantea una metodología para la totalidad de la base de datos, con la cual se puede estimar, el punto de bajada de la etapa de viaje que comienza con cada transacción, los tiempos de viaje entre origen y destino (incluidos tiempos de transbordo, caminata, espera), diferenciar entre etapa de viaje y viaje propiamente tal mediante criterios determinados a partir del análisis exploratorio realizado en la fase previa en conjunto con los parámetros sugeridos en la literatura consultada.

4.1 Análisis Exploratorio: Estudio de una Muestra Pequeña de Tarjetas

Para tener una idea acabada de cómo se debe extraer la información, los posibles escenarios que se presentan en ella, las fallas de información que se pueden reconocer, cómo se pueden enfrentar, de qué manera estimar la bajada, como aprovechar al máximo la información semanal y diaria disponible, entre otros, se realizó un análisis exploratorio en el cual se buscó tarjetas BIP! que presentaran un comportamiento regular en la semana observada, es decir, que realizaran el mismo número de transacciones diarias a lo largo semana.

Para ello se escogieron 17 tarjetas que efectúan desde dos hasta seis transacciones diarias por tarjeta, y que abarcan todas las combinaciones de modo que se permiten en el sistema. El objetivo de este enfoque es captar y adquirir sensibilidad con los comportamientos que se presentan en la información.

El estudio realizado a esta pequeña muestra se presenta en dos etapas. La primera detalla la extracción de información relevante, desde la base datos, de las tarjetas a estudiar (Ver Figura 3-2), y el posterior tratamiento que se debe realizar a los datos para estimar el punto de bajada y calcular tiempos de viaje, caminata y espera, estableciendo el comportamiento del usuario.

La segunda etapa muestra los resultados obtenidos al implementar el procesamiento mencionado, permitiendo realizar un análisis de ellos, y así establecer la calidad de las estimaciones, y los rangos de valores en que se encuentran las distancias y tiempos de transbordo, así como también los tiempos de viajes.

4.1.1 Extracción y procesamiento de la información

La extracción de información tiene como fin conocer los atributos importantes de las transacciones de cada tarjeta y rastrear la tarjeta en el modo de transporte que abordó. Para ello, primero se extrae de la base de datos Transacciones la información correspondiente a cada tarjeta escogida para la muestra mediante el número de identificación de ellas (id), seleccionando:

- Hora y día en que se hizo la transacción
- Lugar donde se realizó: El cual corresponde a la patente del bus, si la transacción se realizó en el bus, el nombre de la estación de metro, si se validó en metro, o en el caso de zona paga (zp) se le asocia su código.
- Posición de la transacción: Coordenadas UTM de donde se realizó la transacción.
- Tipo de Servicio: Indica, en caso de abordar un bus, si este es Troncal o Alimentador.
- Servicio: Número de recorrido que el bus tiene asignado al momento de la transacción, en caso de ser metro corresponde al nombre de la estación donde se validó, y para el caso de zona paga aparecen los recorridos que tienen parada en dicha zona.
- Pago: Monto que paga el usuario al validar.

En la Tabla 4-1 se muestra el resumen con las principales características de las tarjetas que conforman la muestra.

Tabla 4-1: Principales características de la muestra.

N° de transacciones diarias	Usuario	ID	Total transacciones semanales	Lugares donde validó
2	1	1016146244	12	bus
	2	1002914474	10	bus, metro, zp
	3	1006235226	15	bus, zp
	4	1002993498	10	metro
3	5	3108959572	15	bus
	6	1006232954	16	bus, metro
	7	1046831445	18	bus, metro, zp
	8	1022373956	18	bus, metro, zp
4	9	704045850	24	bus
	10	1521061684	24	bus, metro
	11	1157875018	21	bus, metro
5	12	2557778756	25	bus
	13	1909622228	27	bus, metro
	14	2757250644	25	bus, metro
6	15	4176352388	30	bus
	16	1160430346	35	bus, zp
	17	3606733028	35	bus, metro

A partir de esta información, se realiza un seguimiento de los viajes que comienzan con las transacciones de cada tarjeta, para ello se definen los siguientes conceptos:

- **Origen:** Posición donde comienza el viaje.
- **Destino:** Posición donde termina el viaje.
- **Viaje:** Desplazamiento entre un origen y un destino para realizar una cierta actividad en el destino, en un cierto período. Puede estar compuesto por etapas.
- **Etapas:** Segmento de viaje que se identifica al existir diferentes modos de transporte ó servicios utilizados. La presencia de etapas en el viaje constituyen transbordos.
- **Punto de subida:** Posición donde comienza la etapa.
- **Punto de bajada:** Posición donde termina la etapa.
- **Hora origen:** Hora en que se realizó la transacción en el origen del viaje.
- **Hora destino:** Hora estimada en que se llegó al destino del viaje.
- **Hora punto de subida:** Hora en que se realizó la transacción que da inicio a la etapa.
- **Hora punto de bajada:** Hora estimada en que se terminó la etapa.

- **Tiempo de viaje:** Duración del viaje, diferencia de tiempo entre hora origen y hora destino.
- **Tiempo de etapa:** Duración de etapa del viaje, diferencia de tiempo entre hora punto de subida y hora punto de bajada.
- **Tiempo de transbordo:** Diferencia de tiempo entre etapas de un viaje. Compuesto por tiempo de espera y tiempo de caminata.

Cabe destacar que las posiciones y los tiempos de destino y punto de bajada son estimaciones realizadas por la metodología planteada, basada principalmente en el seguimiento y comportamiento de las tarjetas.

Para hacer el seguimiento de los viajes se debe rastrear la tarjeta desde el momento que efectúa alguna transacción. La información útil a extraer depende del lugar donde se realiza la transacción (bus, metro, ó zona paga).

Para el caso bus, los datos se obtienen desde la base de datos Posicionamiento, recuperando la trayectoria (posición y tiempo) realizada por la patente del bus donde se efectuó la transacción durante el período de tiempo en que había asignación del servicio utilizado.

Si no se conoce el servicio que se abordó, la búsqueda de la trayectoria se realiza considerando el tipo de servicio (alimentador ó troncal) registrado. Para los buses que sirven recorridos alimentadores se busca en una ventana de tiempo comprendida desde el tiempo de la transacción hasta una hora más tarde. Si el bus es del tipo troncal la ventana de tiempo considera dos horas desde el tiempo de la transacción. Las transacciones que no tienen posición no son rastreables ya que no se tiene la posición de subida al bus.

Si se trata de una transacción en metro se debe consultar la información de posicionamiento y operación de las estaciones de la red de metro, que se encuentra en la base de datos Estaciones Metro.

Finalmente, si la transacción se realizó en zona paga, se tiene la posición de la parada donde se abordó algún bus, pero se desconoce el bus al que se subió. Identificar el bus se convierte en un desafío, debido a la incertidumbre acerca del servicio que tomó el usuario (en cada zona paga pasa más de un servicio) ya que los recorridos de los servicios pueden tener distintos destinos - rutas ó bien compartirlos. Este proceso se aborda de manera teórica en la metodología planteada, por lo que para las transacciones realizadas en zona paga de la muestra no se rastrearon las tarjetas.

Extraída la información de posicionamiento y transacciones, éstas se agrupan por día y id para facilitar el procesamiento de la información, el cual busca estimar el punto de bajada, obtener origen y destino, identificar etapas dentro de un viaje, así como también el tiempo de viaje, el tiempo de cada etapa, los tiempos de transbordo, espera y caminata.

Se considera que cada transacción da comienzo a una etapa de viaje, estimando un punto de bajada para cada una de ellas. La estimación del punto de bajada se basa en dos supuestos básicos:

- Para una cierta etapa de viaje, la siguiente transacción marca el término de esta, es decir, se supone que el punto de bajada de la etapa actual es cercano al punto de subida de la próxima etapa.
- Para estimar el punto de bajada de la etapa de viaje que comenzó con la última transacción del día, se considera que esta termina en un punto cercano de donde se realizó la primera transacción de dicho día, es decir, en el último viaje del día se regresa al origen del primer viaje del día.

Para conocer el punto de bajada se necesita conocer la posición de la transacción posterior a la transacción actual, en caso de que ésta sea la última transacción del día, se debe entonces conocer la posición de la primera transacción de dicho día. Conocida la posición se buscará el punto más cercano a ella en la etapa de viaje iniciada con la transacción actual.

Si la transacción actual fue realizada en bus se busca en la trayectoria de este (registros gps), el punto que esté más cerca de la posición de la siguiente transacción ó en caso que ésta sea la última transacción del día, de la posición de la primera transacción de dicho día. Para obtener el punto más cercano, se utiliza la distancia euclidiana dada por la siguiente fórmula:

$$d(tr_{post}, p_{bus}) = \sqrt{(x_{tr_{post}} - x_{p_{bus}})^2 + (y_{tr_{post}} - y_{p_{bus}})^2}$$

Ecuación 4-1

Donde

tr_{post} : Posición de la transacción posterior a la transacción actual, ó en caso de que la transacción actual sea la última transacción del día, corresponde a la posición de la primera transacción de dicho día.

p_{bus} : Posición emitida mediante sistema GPS, del bus abordado por la transacción actual.

Luego, se calcula la distancia entre tr_{post} y cada p_{bus} y se busca la mínima distancia de ellas, verificando que la hora de la emisión no sea posterior al momento en que se realiza la siguiente transacción. De ser así el punto de bajada queda estimado como el p_{bus} más cercano a tr_{post} , en caso contrario, se busca en la trayectoria, el punto anterior al p_{bus} encontrado como mínimo, que cumpla con la condición temporal establecida. Para ambos casos, el punto de bajada estimado tiene posición (x_b, y_b) y hora (t_b) .

En el caso de que la transacción se realizó en metro, se calcula la distancia euclidiana entre tr_{post} y cada estación de metro de la red, y se busca la mínima distancia entre dichos puntos, estimando el punto de bajada como la estación a la cual corresponde la mínima distancia. Para este caso la estimación de bajada sólo cuenta con la posición de ella (x_b, y_b) .

Al realizar la estimación de bajada existen etapas que no presentan estimación de punto de bajada, en estos casos se registra la causa que no lo permite las cuales se identificaron como:

- Falta de posición de la transacción actual.
- tr_{post} sin posición.
- Fallas en los registros gps.
- Transacción única durante el día.

Una vez realizado el proceso de estimación de bajada se revisan aquellas etapas sin estimación, con el propósito de deducir la información faltante a partir del comportamiento del usuario observado y estimado durante la semana.

Primero se intenta inferir la posición para las transacciones sin posición, ya que al deducir dicha posición se recupera la posibilidad de estimar bajada a aquellas transacciones en que ésta correspondía a tr_{post} , para ello:

- 1) Se compara el comportamiento de la tarjeta a lo largo de la semana buscando transacciones que muestren coincidencia temporal (rangos de tiempos de las transacciones) con la transacción sin posición, así como también concuerden con la secuencia de modos utilizados por la tarjeta el día en que existe la transacción sin posición.
- 2) De las transacciones que cumplan las condiciones y criterios definidos, se escoge aquella cuya diferencia temporal con la transacción sin posición sea menor, adjudicando a la transacción sin posición la ubicación de la transacción seleccionada.

- 3) Recuperada la posición de la transacción se realizan los procesos de estimación de bajada establecidos anteriormente, de acuerdo al lugar donde se realizó la transacción.

Para los casos de error en la información de posicionamiento de buses, no se puede estimar la bajada porque no se tiene la trayectoria del bus. En este caso se intenta deducir el punto y hora de bajada de las etapas de viaje afectadas a partir de:

- 1) Comparar el comportamiento de la tarjeta a lo largo de la semana buscando transacciones que muestren coincidencia espacial y temporal (rangos de tiempos de las transacciones) con la transacción que da comienzo a la etapa de viaje perjudicada, y que además concuerden con la secuencia de modos utilizados por la tarjeta el día en que se presenta la falla en la información de posicionamiento.
- 2) De las transacciones que cumplan las condiciones y criterios definidos, se escoge aquella cuya diferencia temporal y espacial sea menor con respecto a la posición e instante de la transacción asociada a la etapa de viaje que presenta falla en la información de posicionamiento. Asignando como punto de bajada de la etapa afectada, el punto de bajada estimado para la etapa de viaje asociada a la transacción seleccionada.
- 3) La hora del punto de bajada se deduce a partir del tiempo de viaje promedio de las etapas de viaje relacionadas a las transacciones que cumplen las condiciones de coincidencia espacial, temporal y de modo establecidas.

Para las transacción únicas durante el día no es factible deducir el punto de bajada debido a los múltiples supuestos que se deberían realizar para asegurar una estimación razonable, ya que no se cuenta con mayores antecedentes para ese día.

Una vez obtenidos los posibles puntos de bajada para cada etapa, junto con la distancia euclidiana (Ecuación 4-1) entre el punto (x_b, y_b) determinado ó inferido y el punto tr_{post} (a esta distancia se llama Δd), se establece cuan cerca están dichos puntos determinando si el punto de bajada estimado permite un correcto seguimiento de la tarjeta.

Para todas las transacciones se calculan los tiempos entre transacciones (t_{t-t}) de un mismo día y tarjeta como:

$$t_{t-t} = t_{tr_{post}} - t_{tr_{cur}} \quad \text{Ecuación 4-2}$$

Este tiempo ayuda a establecer el criterio para separar etapa de viaje, de viaje propiamente tal.

Para las etapas de viaje que presentan estimación del punto de bajada se calculan tiempo de etapa de viaje (t_{etapa}), tiempo entre etapas (Δt), distancia recorrida (d_r) y distancia entre etapas (d_{e-e} que es igual Δd) como:

$$t_{etapa} = t_b - t_{tr_{cur}} \quad \text{Ecuación 4-3}$$

$$\Delta t = t_b - t_{tr_{post}} \quad \text{Ecuación 4-4}$$

$$d_r = \sqrt{(x_b - x_{tr_{cur}})^2 + (y_b - y_{tr_{cur}})^2} \quad \text{Ecuación 4-5}$$

$$d_{e-e} = \Delta d = \sqrt{(x_{tr_{post}} - x_b)^2 + (y_{tr_{post}} - y_b)^2} \quad \text{Ecuación 4-6}$$

Donde

t_b : Hora estimada punto de bajada.

$(x_{tr_{cur}}, y_{tr_{cur}}, t_{tr_{cur}})$: Posición y tiempo en que se realizó la transacción actual, corresponde a la posición y tiempo punto de subida.

Debido a la información disponible al momento de realizar este análisis exploratorio, solo se puede conocer la hora estimada del punto de bajada para el caso en que la transacción se realizó en bus. Con lo cual solo se obtienen t_{etapa} y Δt solo para dichos casos.

Luego, se procede a identificar los viajes que realiza cada tarjeta, identificando cuantas etapas lo componen, origen, destino, tiempos de transbordo (t_{tr}) y tiempo de viaje (t_v). Estableciendo condiciones espacio-temporales que determinan si es una etapa o un viaje (utilizando t_{t-t} , Δt y Δd), e identificando valores que servirán de parámetros para la metodología.

En caso de existir transbordo, el tiempo de transbordo (t_{tr}) corresponde a Δt y está compuesto por tiempo de caminata (t_{cam}) y tiempo de espera (t_e), los cuales se establecen a partir de la distancia de transbordo (d_{tr}) que corresponde d_{e-e} ó Δd , mediante la siguiente fórmula:

$$t_{cam} = d_{tr} / v_c \quad \text{Ecuación 4-7}$$

$$t_e = t_{tr} - t_{cam} \quad \text{Ecuación 4-8}$$

Donde

v_c : Velocidad de caminata, considerada como 1,5 m/s para un peatón promedio¹⁷.

4.1.2 Resultados obtenidos

En esta sección se muestran los resultados obtenidos del análisis exploratorio realizado a las 17 tarjetas seleccionadas. Se comienza por presentar el estado de las transacciones realizadas por dichas tarjetas, para luego analizar los resultados obtenidos de estimar los puntos de bajada de cada etapa, estableciendo los criterios utilizados que resultaron de dicho análisis.

Posteriormente se calculan los tiempos de etapa, se identifican los viajes y se determina la forma de reconocer transbordo, calculando tiempos de caminata, espera y viaje. Verificando además, que los puntos de bajada estimados estén correctos.

En la Tabla 4-2, se presenta el total de transacciones por tarjeta y se identifica cuántas de ellas poseen o no posición.

Tabla 4-2: Transacciones por tarjeta.

Usuario	Total transacciones semanales	N° de transacciones con posición	N° de transacciones sin posición
1	12	12	0
2	10	10	0
3	15	15	0
4	10	10	0
5	15	14	1
6	16	16	0
7	18	18	0
8	18	17	1
9	24	24	0
10	24	23	1
11	21	21	0
12	25	25	0
13	27	27	0
14	25	24	1
15	30	29	1
16	35	35	0
17	35	35	0

¹⁷ Valor obtenido del manual de recomendaciones para el diseño del espacio vial urbano REDEVU.

Como se puede observar, las 17 tarjetas realizaron en su totalidad 360 transacciones a lo largo de la semana (Lunes a Domingo), de ellas un 98,6% tiene posición asociada, existiendo un 1,4% que no posee posición, dicho porcentaje puede duplicarse cuando se estiman los puntos de bajada, ya que las transacciones sin posición afectan directamente la factibilidad de estimar el punto de bajada de las etapas que comienzan con dichas transacciones, así como también aquellas en que dicha transacción corresponde a la utilizada para estimar el punto de bajada de alguna etapa.

La distribución de las transacciones según el sitio donde se realizaron se puede apreciar en la Figura 4-1.

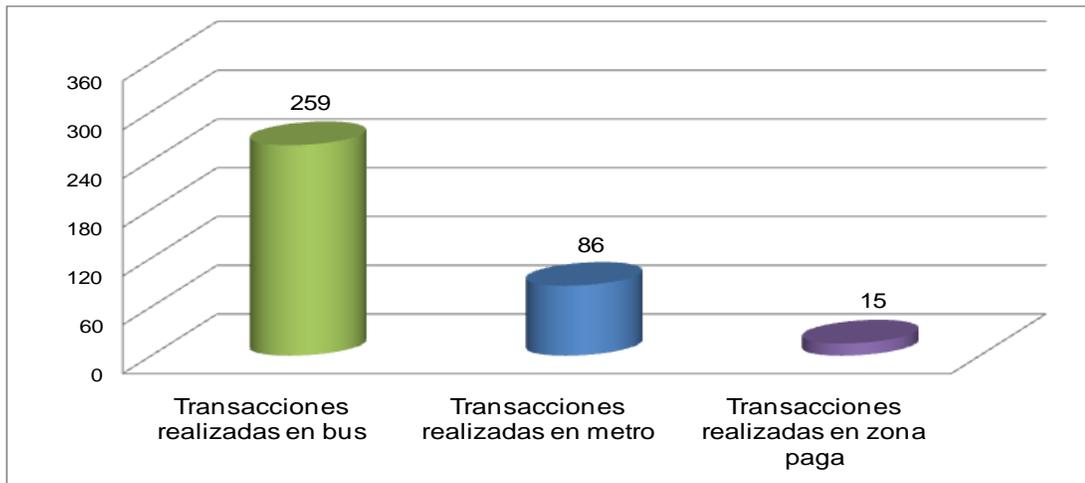


Figura 4-1: Transacciones según lugar donde se realizaron.

Se observa que para la muestra, las transacciones realizadas en bus triplican a las de metro, mientras que las hechas en zonas pagas corresponden a la sexta parte de estas. De las transacciones realizadas en bus, cinco de ellas no poseen posición (1,9%) y un 85% posee asignación de servicio, es decir se conoce el servicio-sentido (recorrido) que está sirviendo al momento de la transacción.

Al calcular el punto de bajada para cada etapa iniciada por una transacción se obtiene a priori que, a un 91,7% de la totalidad de etapas es factible calcular un punto que resulta ser el más cercano, en el seguimiento de la tarjeta, a la siguiente transacción. La distancia Δd varía en un amplio rango de valores que se muestran en la Figura 4-2. El restante 8,3% corresponde a etapas de viaje en que no es posible estimarle bajada en una primera instancia, las causas se detallan más adelante.

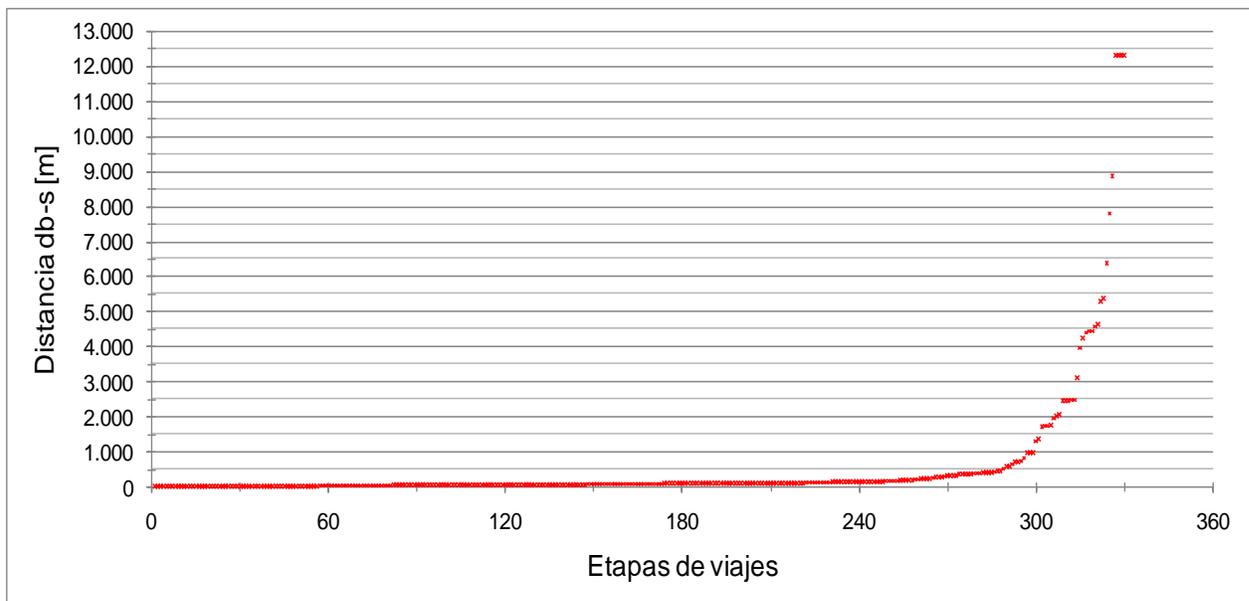


Figura 4-2: Rango de valores distancia Δd .

Como se puede observar en la Figura 4-2, las distancias se concentran en valores menores a 500 metros (87,3% de las etapas), pudiendo reconocer un umbral en los 1.000 metros, valor a partir del cual el número de etapas que se encuentran a distancias mayores disminuye. Es decir, un 87,3% de las etapas terminan su viaje a menos de 500m donde realizan la siguiente transacción, si se considera una velocidad de caminata promedio de 1,5 m/s, el punto de bajada está a 6 minutos de caminata del próximo punto de subida, un 90,6% caminan menos de 1.000m u 11 minutos, y un 9,4% de las etapas terminan en un punto que dista a más de 1km del punto de la siguiente transacción.

Se considera que puntos de bajada a distancias mayores a 1.000m, no son fiables ya que es difícil asegurar que este punto de bajada sea el correcto dadas las distancias y tiempos de caminata, y a la incertidumbre sobre cómo se accedió al punto de subida de la siguiente transacción. Esto último sugiere la idea de etapas que acercan a los usuarios desde un punto de bajada en el actual medio a su destino final mediante otros medios que no se registran en el sistema (bicicleta, colectivo, taxi, auto) ó bien etapas en las cuales se evade la validación o pago del viaje.

En consecuencia los puntos de bajada estimados se dividieron en dos rangos según el umbral determinado en la muestra:

Caminable	:	$0 m \leq \Delta d \leq 1.000 m$
No caminable	:	$1000 m < \Delta d$

Los puntos de bajada “caminable”, son aquellas estimaciones que permiten realizar un seguimiento de la tarjeta y de sus etapas de viaje. Son dichas estimaciones

las que finalmente quedarán como punto de bajada. Por otro lado, los puntos de bajada estimados y clasificados como “no caminable”, quedan como etapas sin estimación de punto de bajada, debido a que no se puede realizar un seguimiento de la tarjeta.

Además de aquellos puntos de bajada estimados que resultaron “no caminable”, existen también otros motivos para no obtener el punto de bajada.

En algunos casos puede existir falla en la información, como lo son emisiones de GPS erróneas (posición incorrecta), emisiones que entregan la misma posición a lo largo de cierto período, o información que sea intermitente (en tiempo) a lo largo de la ruta.

En otros casos el motivo puede ser la falta de información que presentan las transacciones sin posición, y aquellas transacciones únicas del día (las primeras no permiten estimarles punto de bajada a la etapa que dan comienzo, así como también, a alguna etapa en que dicha transacción corresponda a la utilizada para estimar el punto de bajada, mientras que a las segundas no se les puede hacer un seguimiento durante el día).

A todos estos motivos se les denominó “falla de información”.

El la Figura 4-3 se muestra la cantidad de etapas que no presentan estimación de bajada, agrupadas de acuerdo a las categorías establecidas anteriormente, incorporando las etapas de viajes cuya transacción se realizó en zona paga y por consiguiente en este análisis no se realizó estimación de bajada.

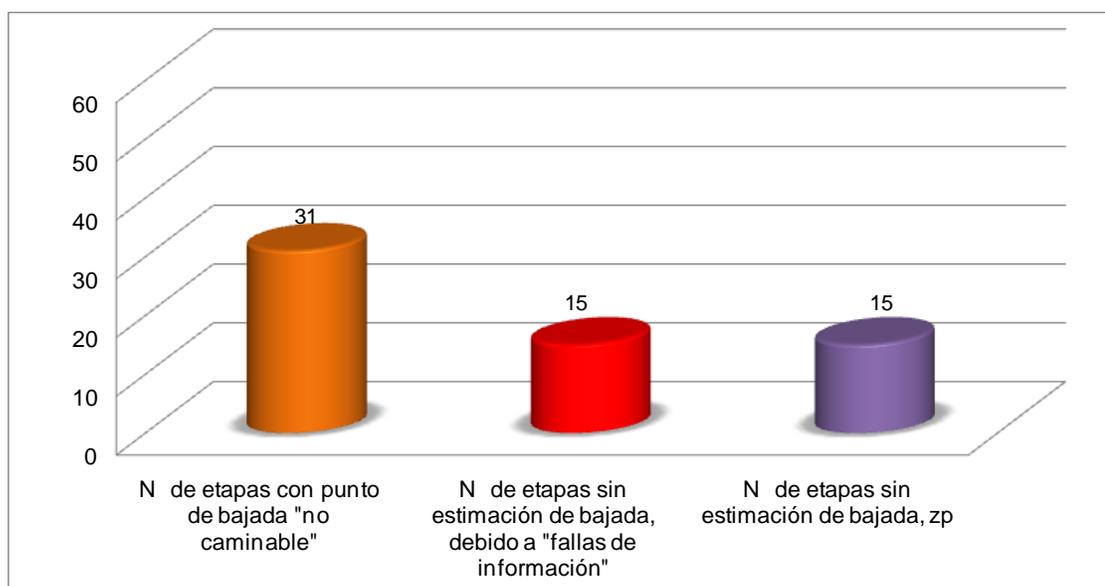


Figura 4-3: Distribución según motivo, de etapas sin estimación de punto de bajada

Se observa que del total de transacciones realizadas por la muestra de tarjetas escogidas, un 83% presenta estimación de punto de bajada existiendo una sexta parte de las transacciones a las cuales no es posible estimar el punto de bajada. La mitad de estos casos corresponden a aquellos no rastreables (“No caminable”), una cuarta parte a “fallas de información” y la misma cantidad a etapas cuya transacción se realizó en zona paga.

De la causa “fallas de información”, dos tercios de estos casos corresponden a etapas cuya transacción actual no tiene posición o bien cuya transacción posterior no la tiene (se duplica el efecto de las transacciones sin posición), el tercio restante corresponde principalmente a aquellos casos en que sólo hay una transacción en el día, y solo existe un caso donde el motivo resulta ser error de registros gps.

Al revisar aquellas etapas sin estimación debido a “fallas de información”, se intenta recuperar información de origen y del punto de bajada en aquellos casos donde el comportamiento de la tarjeta lo permite.

Como se identificó en el capítulo 3, las transacciones sin posición se deben a errores en la información de posicionamiento del bus, por lo que no es factible realizar el proceso de estimación de bajada para estas transacciones. En consecuencia si desea conocer el punto de bajada de estas etapas de viaje éste se debe inferir del comportamiento del usuario.

Para las 5 transacciones sin posición se recuperó la posición en la totalidad de los casos mediante la aplicación del tratamiento descrito en el punto 4.1.1 (pág. 60), donde se verificó que la tarjeta presentara transacciones los otros días que tuvieran la misma estructura de modos, y que se encontraran en una ventana de tiempo $[t_{tr_{cur}} - 15min, t_{tr_{cur}} + 15 min]$ para días laboral, y $[t_{tr_{cur}} - 60min, t_{tr_{cur}} + 60 min]$ para días de fin de semana.

De las transacciones que cumplían dichas condiciones se seleccionó aquellas dos transacciones más cercanas (temporalmente) a $t_{tr_{cur}}$, luego, se comparó la distancia de las transacciones escogidas con la posición de tr_{cur} , si las distancias eran \leq a 500m se estableció como posición de tr_{cur} la posición de la transacción con menor diferencia de tiempo a $t_{tr_{cur}}$, y si las distancias eran $>$ a 500m no se realizó recuperación de información.

Al recuperar la totalidad de las transacciones sin posición, se pudo realizar el proceso de estimación de bajada para todas las etapas de viaje cuya transacción posterior no tenía posición, resultando 4 de ellas con punto de bajada “Caminable” y uno “No caminable”.

El tratamiento que se le da a las transacciones sin posición es similar al establecido para etapas de viaje que presentan fallas de rutas, ya que ambos se deben al mismo problema: error de registros gps. La recuperación del punto de subida de estas etapas se realiza utilizando el criterio descrito para transacciones sin posición, en tanto para la deducción del punto de bajada se realiza el proceso explicado en el punto 4.1.1 (pág. 61) donde se adjudica como punto de bajada de la etapa que presenta problemas, la posición del punto de bajada estimado para la etapa que se determinó cumplía con los criterios espacio-temporales establecidos.

La deducción del tiempo de bajada se realiza tomando como t_v el tiempo de viaje promedio para las etapas de viaje realizadas por la tarjeta los otros días y que se cumplieran con la pertenencia a la ventana de tiempo establecida. En consecuencia t_b se obtiene de:

$$t_{b_{\text{sin pos/fallaruta}}} = t_{tr_{cur}} + \bar{t}_v \quad \text{Ecuación 4-9}$$

Mediante este procedimiento fue posible deducir el caso de “falla ruta” que existía en la muestra y los cinco casos de transacciones sin posición, todos resultan “caminable”.

A continuación se presenta la distribución final (incluye proceso de deducción) de éstas categorías (“caminable”, No caminable”, “falla de información”) con respecto al total de transacciones realizadas según el lugar donde se realizó la transacción.

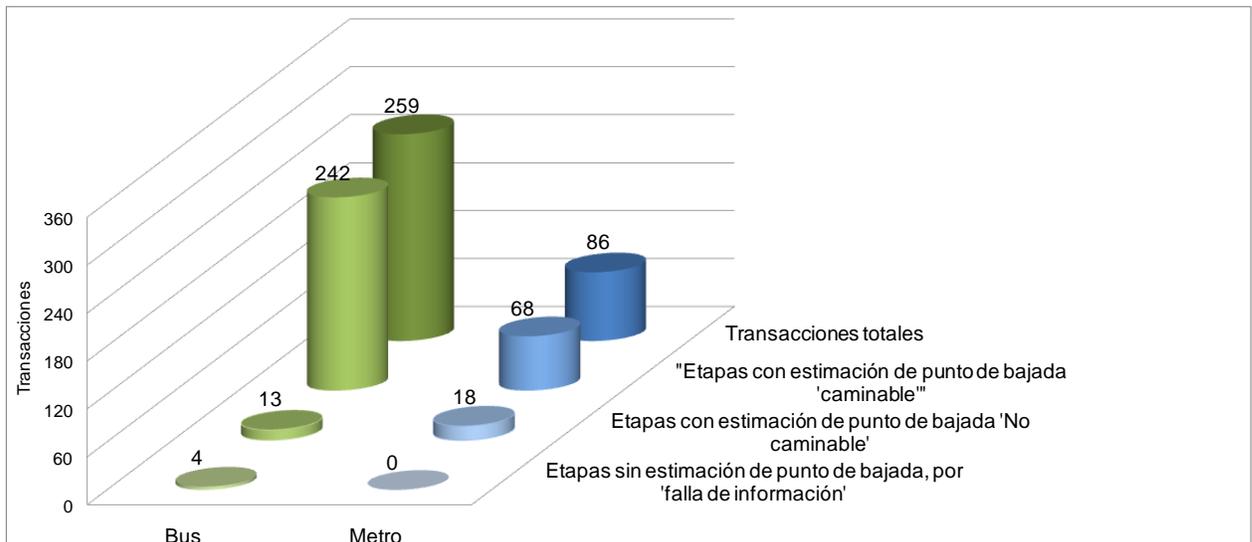


Figura 4-4: Distribución según lugar donde se realizó la transacción, de clasificación de punto de bajadas de las etapas.

En la Figura 4-4, se puede ver que existen variaciones en la participación de cada categoría según modo, destacando el aumento de los casos “No caminable” para metro en relación a los presentes en bus (15,9% más) lo que se explica debido a que para

transacciones realizadas en bus existe mayor posibilidad de encontrar un recorrido cercano al punto de destino dado que la red de buses es más flexible, mientras que para metro esta posibilidad disminuye al ser la red fija.

De la causa “fallas de información”, el total de ellos corresponden aquellos casos en que sólo hay una transacción en el día.

Del total de la muestra se obtienen 310 etapas con información sobre el punto de subida y bajada, de las cuales 242 presentan estimación de tiempo de bajada (realizadas en bus).

Los tiempos de viaje estimados se muestran en la Figura 4-5, donde se observa que un 68,2% de las etapas de viaje que presentan tiempo de estimación tienen tiempo de viaje inferior a media hora, en tanto un una cuarta parte de estos tienen tiempos de viajes entre media y una hora, y un 6,6% presenta tiempos de viaje mayores a una hora y menores a una hora cuarenta y cinco minutos.

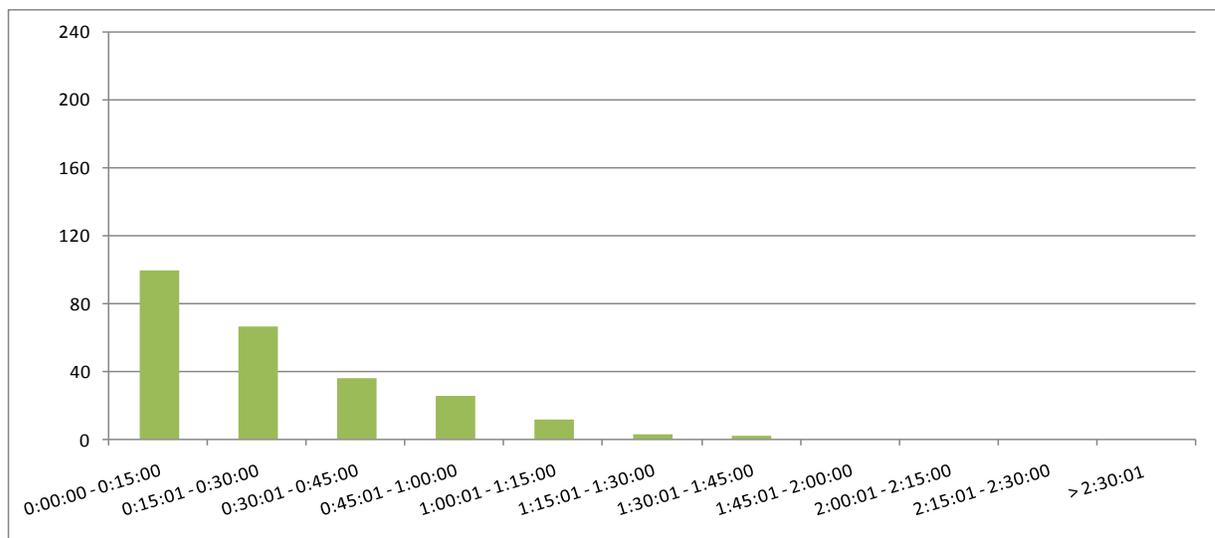


Figura 4-5: Distribución de tiempos de viaje para muestra usuarios.

El tiempo entre etapas (Δt) calculado según la Ecuación 4-4, se obtiene para todas aquellas etapas de viaje que presentan estimación de tiempo de bajada (242). Para la última transacción del día se tomo la diferencia temporal en valor absoluto con respecto a la primera transacción de dicho día. La distribución de los Δt obtenidos se muestra en la Figura 4-6.

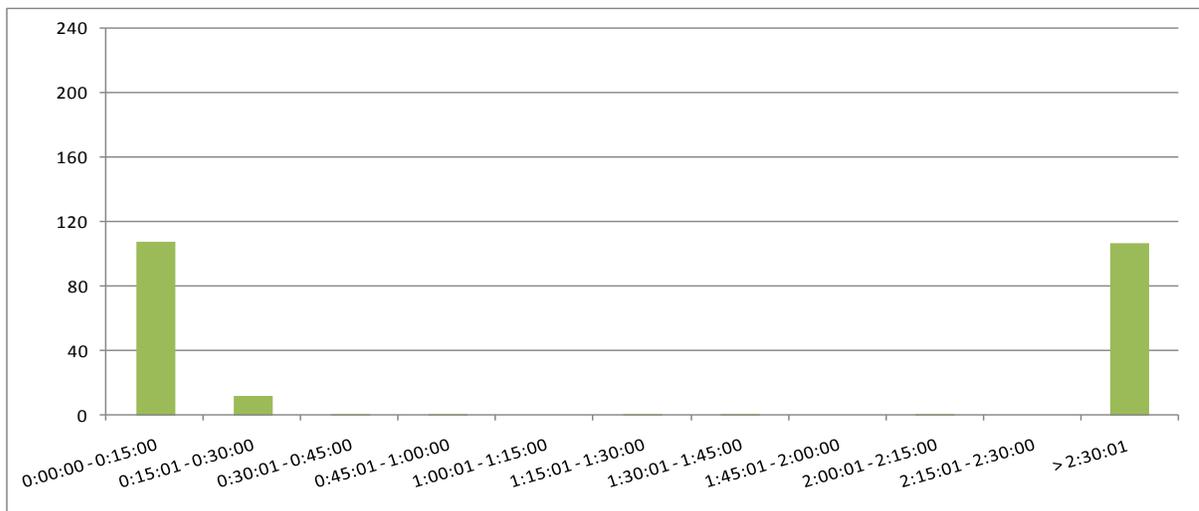


Figura 4-6: Distribución de tiempos entre etapas Δt para muestra usuarios.

Se observa que un 51,7% de los Δt son menores a media hora y cerca de un 46% presenta Δt de más de dos horas. Es claro identificar transbordo para Δt menores a 15 minutos, y actividades de larga duración cuyo Δt es mayor a dos horas. Entre 15 y 30 minutos, es necesario realizar un análisis más detallado para identificar actividades de corta duración.

Para la muestra se determinó que se realizaron 166 transbordos, de los cuales 124 presentan estimación del tiempo de transbordo (Δt) (no se realizó estimación de tiempo de bajada para metro).

También es importante conocer los intervalos de tiempo que se presentan entre transacciones, ya que permiten establecer parámetros para identificar etapas de viaje, al asumir que al estar un tiempo prolongado en un lugar se desarrolló una actividad que da término al viaje, esto se ilustra en la Figura 4-7, donde se observan claros intervalos de tiempo que diferencian transbordo, de actividad en el destino.

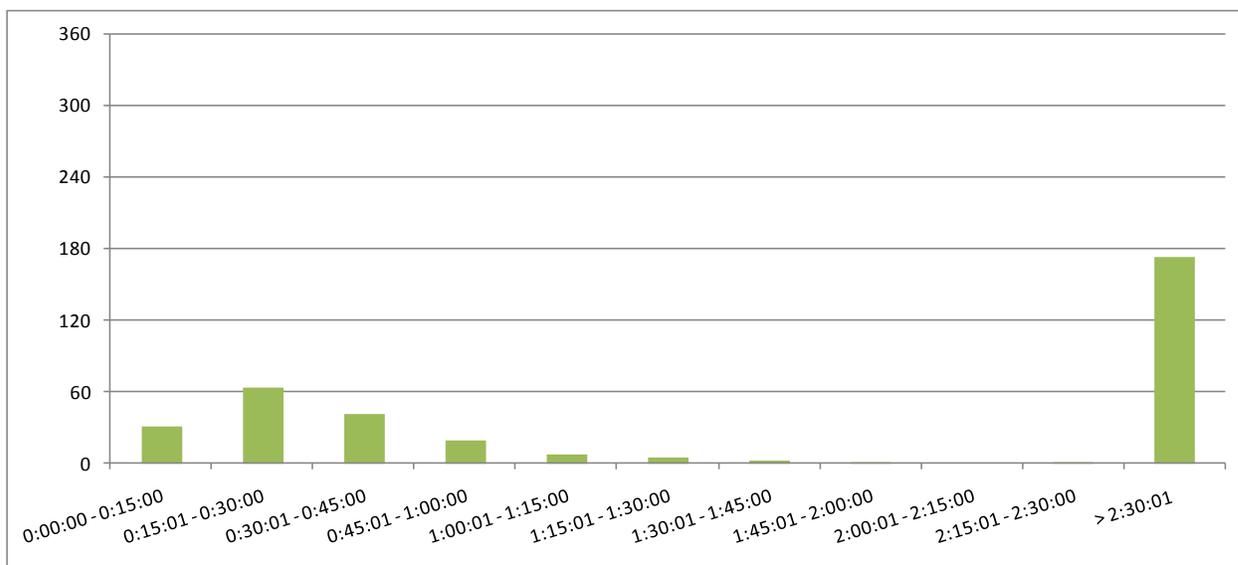


Figura 4-7: Distribución de tiempos de entre transacciones para muestra usuarios.

Con los tiempos de viaje, tiempos entre transacciones y tiempos de transbordo calculados, se identifica la estructura de los viajes realizados por las tarjetas de la muestra. Considerando etapa de viaje cuando los tiempos entre bajada y subida son menores a media hora, y determinado término de viaje cuando la siguiente transacción tiene una diferencia de tiempo mayor a dos horas.

En la Tabla 4-3 se presenta la estructura de viaje encontrada para cada usuario.

Tabla 4-3: Estructura de viajes, muestra usuarios.

Tarjeta	N° viajes	N° etapas por viaje	N° transbordos identificados
1	12	1	0
2	10	1	0
3	11	1	0
	2	2	2
4	10	1	0
5	5	1	0
	5	2	5
6	6	1	0
	5	2	5
7	6	1	0
	6	2	6
8	6	1	0
	6	2	6
9	12	2	12
10	12	2	12
11	1	1	0

Tarjeta	N° viajes	N° etapas por viaje	N° transbordos identificados
	10	2	10
12	1	1	0
	6	2	6
	4	3	8
13	2	1	0
	5	2	5
	5	3	10
14	5	2	5
	5	3	10
15	10	3	20
16	3	1	0
	1	2	1
	10	3	20
17	1	2	1
	11	3	22

En anexos se encuentra el detalle de los viajes para los 17 usuarios escogidos.

4.2 Metodología Propuesta

La metodología que se plantea, tiene por objetivo principal reconstruir los viajes realizados por cada tarjeta BIP presente en el sistema, basado en la estimación de destino a partir de la información disponible. Con ello se puede generar una nueva tabla que permita extraer y realizar análisis de comportamiento, origen destino, perfiles de carga, realización de actividades, de una manera más fácil y directa.

El trabajo que se realizó con la muestra pequeña de usuarios, sirvió para sensibilizar con el tipo de información que es posible encontrar en las bases de datos y ayudar a determinar la forma de abordar la metodología.

En la Figura 4-8 se presenta de manera general la metodología propuesta, donde se establecen 5 procesos principales que permiten conocer los destinos de las etapas así como la composición de los viajes y características de éste. Estos procesos apuntan a enriquecer la información disponible generando herramientas que facilitan el posterior análisis de ella.

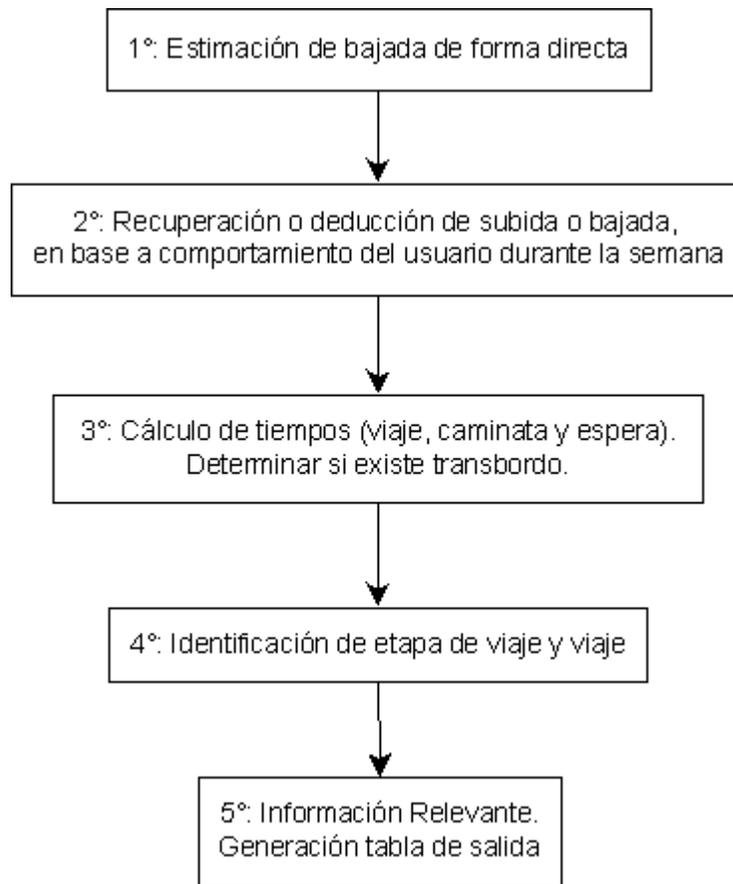


Figura 4-8: Metodología propuesta.

Cada uno de estos pasos se detalla en secciones posteriores dentro de este capítulo. A continuación se presenta una breve introducción con el objetivo de cada uno de estos pasos y sus alcances.

- **Estimación de bajada:** Este paso lleva a cabo la ejecución del algoritmo que tiene por objetivo obtener el destino de cada transacción realizada que cuenta con información de posición (transacción actual, posterior ó primera del día según corresponda). Aquí se estiman todas aquellas etapas de viaje que presentan estimación de punto de bajada caminable.
- **Recuperación y deducción de información:** Estimación de subida o bajada, en base a comportamiento del usuario durante la semana. Con esto se pretende recuperar información de aquellas transacciones en las cuales por falta ó falla de información no era posible realizar el seguimiento de la tarjeta. Recupera porcentaje de estimaciones en las cuales el primer proceso no arrojaba resultados.
- **Cálculos de tiempos:** Este proceso se realiza una vez estimada la bajada. En él se calculan tiempos de viaje, se determina la existencia de transbordo, se

obtienen los tiempos de caminata, espera, y transbordo propiamente tal, además de la distancia entre transacciones. Una vez conocidos los tiempos y posición de subida y bajada (origen-destino) se puede realizar un seguimiento de las transacciones de la tarjeta para identificar si realiza transbordo ó actividades en el punto de bajada estimado.

- **Identificación de etapa de viaje y viaje:** Una vez determinada la existencia de transbordo se etiqueta el tramo de viaje observado como viaje o etapa según corresponda.
- **Información relevante:** Detalla el output de esta metodología, el cual consiste en una tabla que contiene la información relevante para realizar análisis y consultas sobre el comportamiento del usuario, perfiles de carga, patrones de viaje, etc.

4.2.1 Estimación de bajada

La metodología a plantear presenta diferencias con el tratamiento que se realizó en el análisis exploratorio, principalmente en la formulación de la estimación de bajada para etapas de viaje en buses, donde se mejora el criterio de mínima distancia integrando el tiempo de viaje asociado.

La estimación del punto de bajada que se realiza para todas las etapas de viaje, considera los siguientes supuestos:

- Para una cierta etapa de viaje, la siguiente transacción marca el término de esta, es decir, se supone que el punto de bajada de la etapa actual es cercano al punto de subida de la próxima etapa, y el tiempo del punto de bajada es anterior a la hora del próximo punto de subida.
- Para estimar el punto de bajada de la etapa de viaje que comenzó con la última transacción del día, se considera que esta termina en un punto cercano de donde se realizó la primera transacción de dicho día, es decir, en el último viaje del día se regresa al origen del primer viaje del día.

El modelo propuesto se basa en el modelo establecido por Trépanier et al. (2007) y se ilustra en la Figura 4-9.

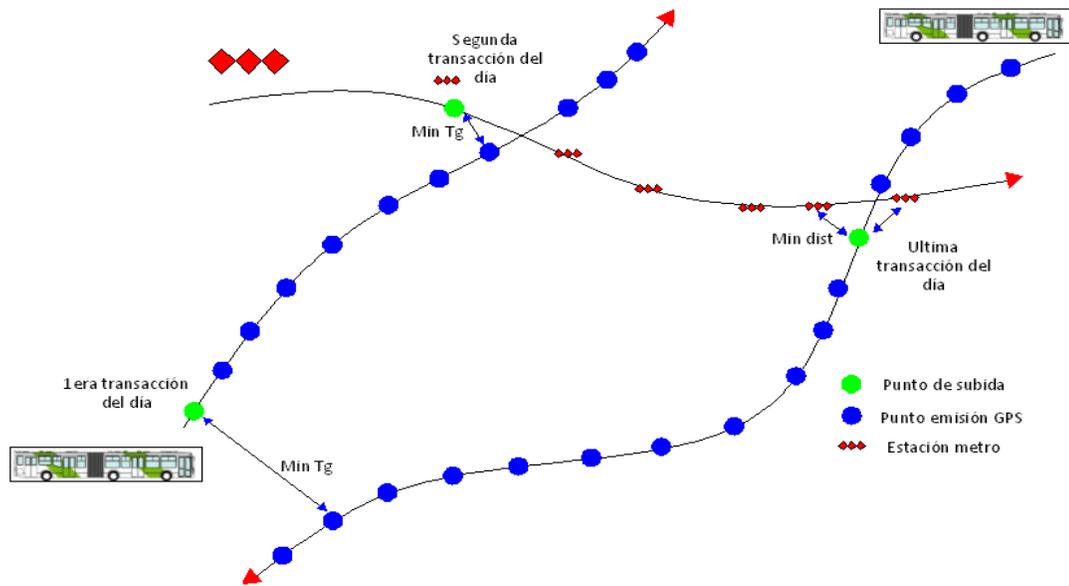


Figura 4-9: Modelo de estimación de bajada.

Dada la integración de modos presente en el sistema, la formulación para determinar el punto más cercano a la siguiente transacción se plantea de acuerdo al lugar donde se realizó la transacción (bus (sección 4.2.1.1), metro (sección 0), zona paga (sección 4.2.1.3)).

Para los casos bus y zona paga (etapas de viaje en bus) la formulación planteada se basa en optimizar el tiempo generalizado, el cual se define considerando el tiempo de viaje en el bus y el tiempo de caminata al punto de la siguiente transacción. Con este concepto se elimina el problema que se presenta en trayectorias cortas donde en las ventanas de tiempo de búsqueda se registran posiciones gps muy cercanas pero que corresponden a distintas direcciones de viaje, y que el buscar aquel punto de mínima distancia no asegura determinar un punto gps que sea de acuerdo a la dirección del viaje del usuario, pudiendo sobreestimar el tiempo de viaje de este.

Para el caso metro la formulación es distinta, ya que la red e información disponible es de diferente naturaleza.

El tratamiento general de la información presente en la tabla de transacciones para obtener la estimación del punto de bajada se muestra en la Figura 4-10, el cual define el algoritmo computacional a seguir.

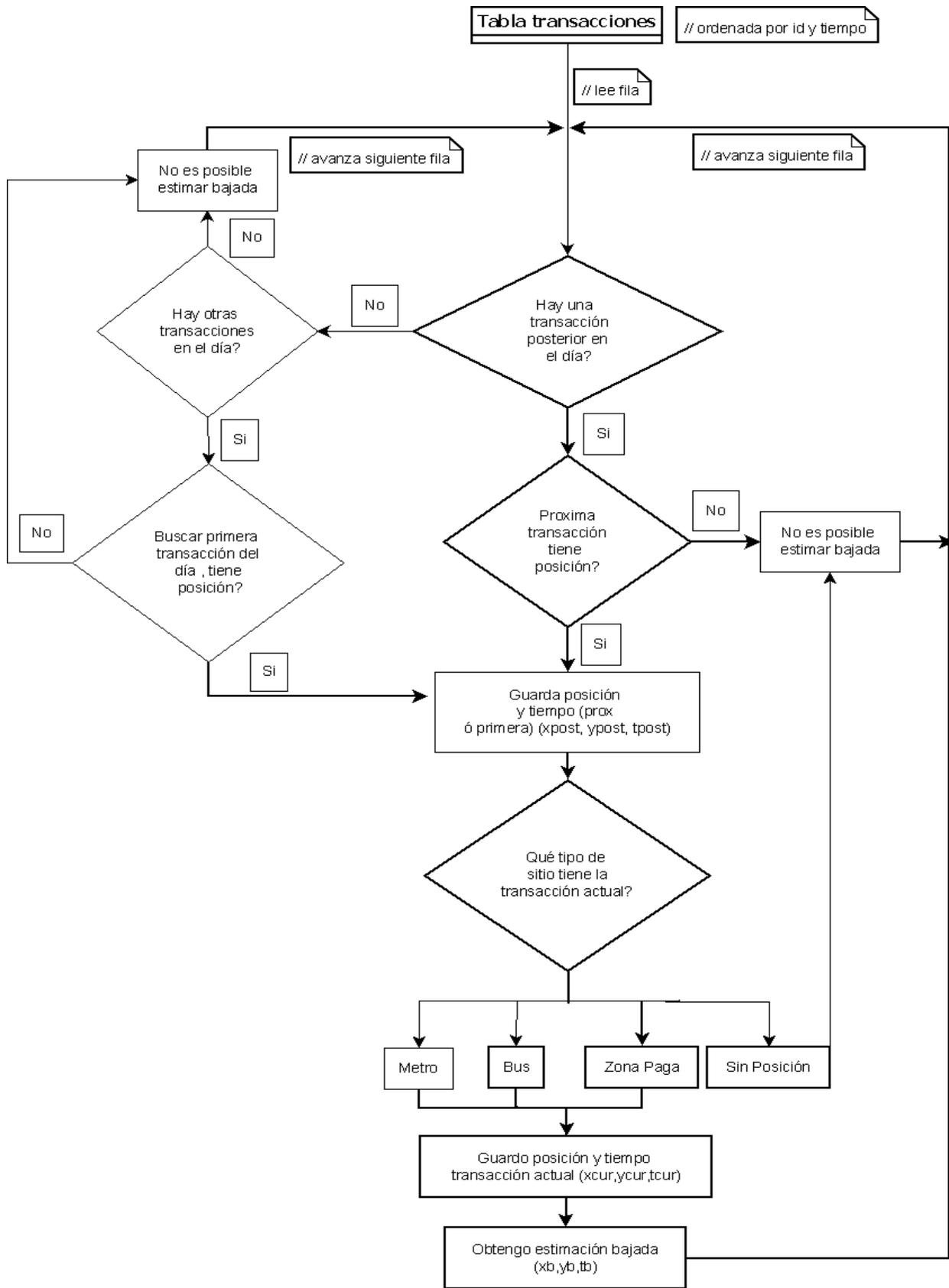


Figura 4-10: Diagrama general de tratamiento de la información para la estimación del punto de bajada.

Como se puede observar en la Figura 4-10, el procesamiento de la información reconoce aquellas transacciones actuales y transacciones posteriores que no tienen posición, así como también se reconocen aquellas transacciones que son únicas en el día, identificando para todos estos casos que no es posible estimar el punto de bajada, y registrando el motivo por el cual no se presenta estimación.

4.2.1.1 Caso Bus

Para las transacciones realizadas en bus, el desafío que se presenta al estimar el punto de bajada, se encuentra en la precisión y calidad de la estimación a realizar, dada la gran cantidad de datos de posicionamiento para cada bus y las diferentes configuraciones de recorridos de la red.

Se busca un punto de bajada (x_b, y_b, t_b) dentro de la trayectoria del bus (emisiones gps) que sea el mejor candidato a ser una estimación correcta (Ver Figura 4-11). Para ello se define la medida de Tiempo generalizado para el punto gps i (Tg_i) como:

$$Tg_i = t_i + f_c \cdot \frac{d_{i-tr_{post}}}{V_c}$$

Ecuación 4-10

Donde

- i : Punto gps con coordenadas (x_i, y_i) .
- t_i : Tiempo asociado al punto (x_i, y_i) .
- $d_{i-tr_{post}}$: Distancia euclidiana entre (x_i, y_i) y (x_{post}, y_{post}) .
- V_c : Velocidad promedio de caminata, considerada como 1,5 m/s.
- f_c : Factor de peso, que evidencia el desagrado de caminar en relación al tiempo de viaje en vehículo.

La Ecuación 4-10 considera la hora (tiempo) de emisión del punto gps i y el tiempo de caminata estimado desde dicho punto al punto de subida de la siguiente transacción, recogiendo la molestia que perciben los usuarios por caminar a cambio de ir abordo del bus.

El problema consiste entonces en buscar un punto gps i talque

$$\forall i, \quad t_i \in \begin{cases} [t_{cur}, t_{cur+60 \text{ min}}], & \text{si tipo de servicio es troncal} \\ [t_{cur}, t_{cur+30 \text{ min}}], & \text{si tipo de servicio es alimentador} \end{cases}$$

$$\begin{aligned}
 & \text{Min } Tg_i \\
 \text{s.a } & d(i, tr_{post}) \leq 1000m \\
 & t_i < t_{post}
 \end{aligned}$$

Ecuación 4-11

Donde

$d(i, tr_{post})$: Distancia euclidiana (Ecuación 4-1) entre el punto gps i y la siguiente transacción (x_{post}, y_{post}) .

t_{cur} : Tiempo de la transacción actual.

t_{post} : Tiempo de la transacción posterior.

Aquel punto gps i que tenga $Min Tg_i$ asociado se identifica como $Tg (x_{tg}, y_{tg}, t_{tg})$.

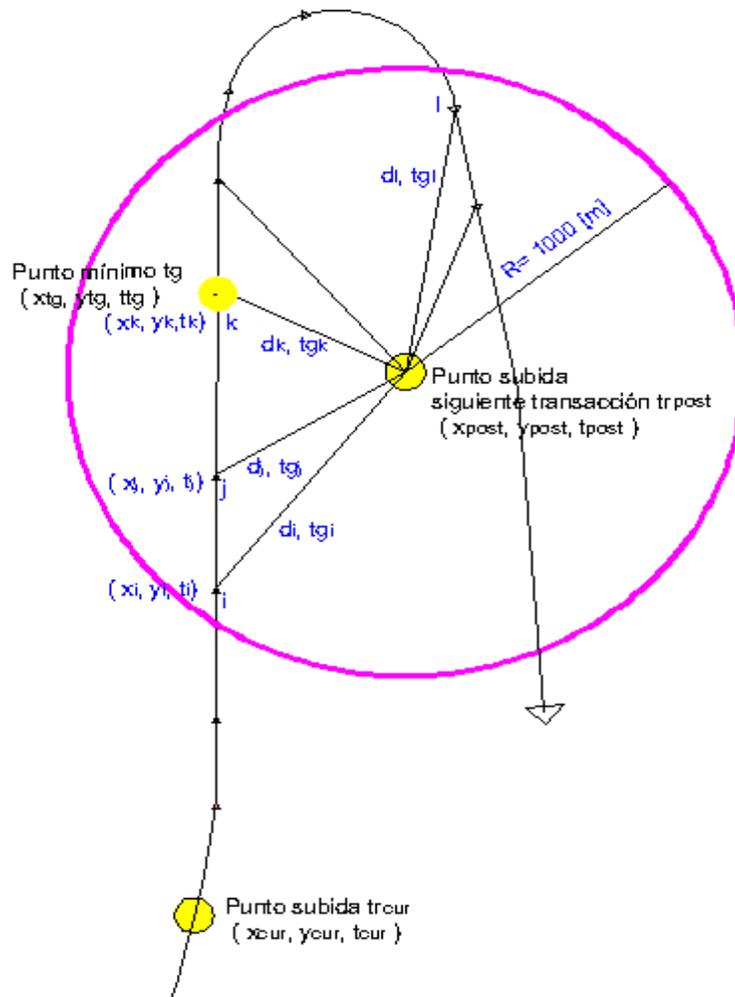


Figura 4-11: Esquema búsqueda de punto gps que minimiza el tiempo generalizado.

Las ventanas de tiempo consideradas para realizar la búsqueda de posiciones gps de la trayectoria son diferentes de acuerdo al tipo de servicio del bus, estableciéndose en relación a lo observado en la muestra estudiada anteriormente. Se establecen ventanas de tiempo para optimizar y hacer más eficiente la búsqueda dada la cantidad de información.

Para el caso en que el punto de bajada esté fuera de la ventana de tiempo en que se realiza la búsqueda, se activa la restricción, encontrando el último punto gps de la trayectoria como el mínimo Tg , en este caso se realiza una nueva búsqueda en una nueva ventana de tiempo que va desde el tiempo del punto gps encontrado Tg más 30 ó 60 minutos según el tipo de servicio que sea.

El que no exista resultado para la Ecuación 4-11, indica que no existen puntos gps que cumplan la condición de distancia establecida, por lo que la etapa de viaje queda sin estimación de punto de bajada, registrando como causa “No caminable”.

Para las etapas de viaje realizadas en bus cuya transacción cuente con asignación de servicio, se puede mejorar la estimación del punto de bajada al contar con la ruta establecida que el bus realiza (eliminado errores asociados a curvas en trayectoria gps). Para ello, se recuperan los puntos gps anterior (Tg_{ant}) y posterior (Tg_{post}) al punto Tg identificado, proyectándolos ortogonalmente a la ruta obteniendo ($Tg_{ant_{proy}}$, Tg_{proy} , $Tg_{post_{proy}}$), luego se calcula la distancia euclidiana entre cada uno de los puntos proyectados y la siguiente transacción. Finalmente, aquel que presente la mínima distancia se establece como estimación de punto de bajada (x_b , y_b , t_b).

Para las etapas de viaje en bus cuya transacción no cuente con asignación de servicio, se establece como estimación de punto de bajada (x_b , y_b , t_b) el punto Tg identificado.

El método descrito se muestra en la Figura 4-12, el cual establece el algoritmo computacional a seguir.



Figura 4-12: Diagrama de la estimación del punto de bajada, para transacciones realizadas en bus.

4.2.1.2 Caso Metro

En este caso la estimación del punto de bajada corresponde a una estación de metro, la cual se determina buscando la estación que esté más cercana a la transacción posterior (mínima distancia $d(tr_{post}, p_{metro})$) en un radio de 1000m (parámetro establecido para determinar precisión de la estimación de bajada). Si no existe una estación en dicho radio se establece que la etapa de viaje queda sin estimación de punto de bajada, registrando como causa “No caminable”.

Una vez estimada la estación de bajada punto (x_b, y_b, em_b) con x_b, y_b coordenadas de la estación y em_b nombre de la estación de bajada. Se busca determinar el instante de llegada a dicha estación, para ello primero se determina la secuencia de estaciones que representa el camino más corto entre las estaciones inicio y fin del viaje, implementando el algoritmo Dijkstra que utiliza los tiempos entre estaciones y tiempos de detención en cada estación como pesos. Los transbordos son incorporados, como nodos auxiliares en la codificación de las intersecciones de las líneas. Luego teniendo la secuencia de estaciones se calcula el tiempo de viaje en metro (Ver Figura 4-13).

$$tv_{metro} = 2t_{ct-a} + t_{ea} + t_{vtren} + t_{tranb}$$

Ecuación 4-12

Donde

t_{ct-a} : Tiempo de caminata desde torniquete a andén ó desde andén a torniquete.

Tiempo de espera en andén (t_{ea}): $\frac{1}{2f} - \frac{1}{2}t_{det_{estini}}$

Tiempo de viaje en tren (t_{vtren}): $\frac{1}{2}t_{det_{estini}} + t_{rec_{ini-fin}} + t_{det_{sig.ini-ant.fin}}$.

t_{tranb} : Tiempo de transbordo entre líneas.

Como se observa el tiempo de espera en andén se determina en base a la frecuencia de los trenes (f).

En tanto el tiempo de viaje en el tren considera el promedio del tiempo detenido del tren en la estación de inicio ($t_{det_{estini}}$), los tiempos de recorrido entre las estaciones definidas como inicio y fin ($t_{rec_{ini-fin}}$), y el tiempo de detención del tren en cada estación a partir de la estación posterior a la de inicio y hasta la anterior de la estación fin ($t_{det_{sig.ini-ant.fin}}$) (se considera bajada instantánea del pasajero del tren).

Se incluyen también los tiempos de transbordo, en caso de existir, estos deben ser obtenidos ya sea de mediciones o estimaciones realizadas en base a la distancia de combinación entre líneas.

Además se considera dos veces el tiempo de caminata desde el acceso de la estación (torniquete) al andén, dado que esta acción se reitera en la estación de subida y en la estación de bajada pero con distinta dirección (torniquete-andén/andén-torniquete). Las magnitudes de estos tiempos se deben obtener de mediciones en terreno, considerando tiempos promedios dadas las diferentes configuraciones de estaciones de la red.

Finalmente se estima el tiempo asociado a la estación de bajada como:

$$t_b = t_{cur} + tv_{metro} \quad \text{Ecuación 4-13}$$

Donde

t_{cur} : Tiempo de la transacción actual.

tv_{metro} : Tiempo de viaje en metro.

El proceso descrito se ilustra en la Figura 4-12.

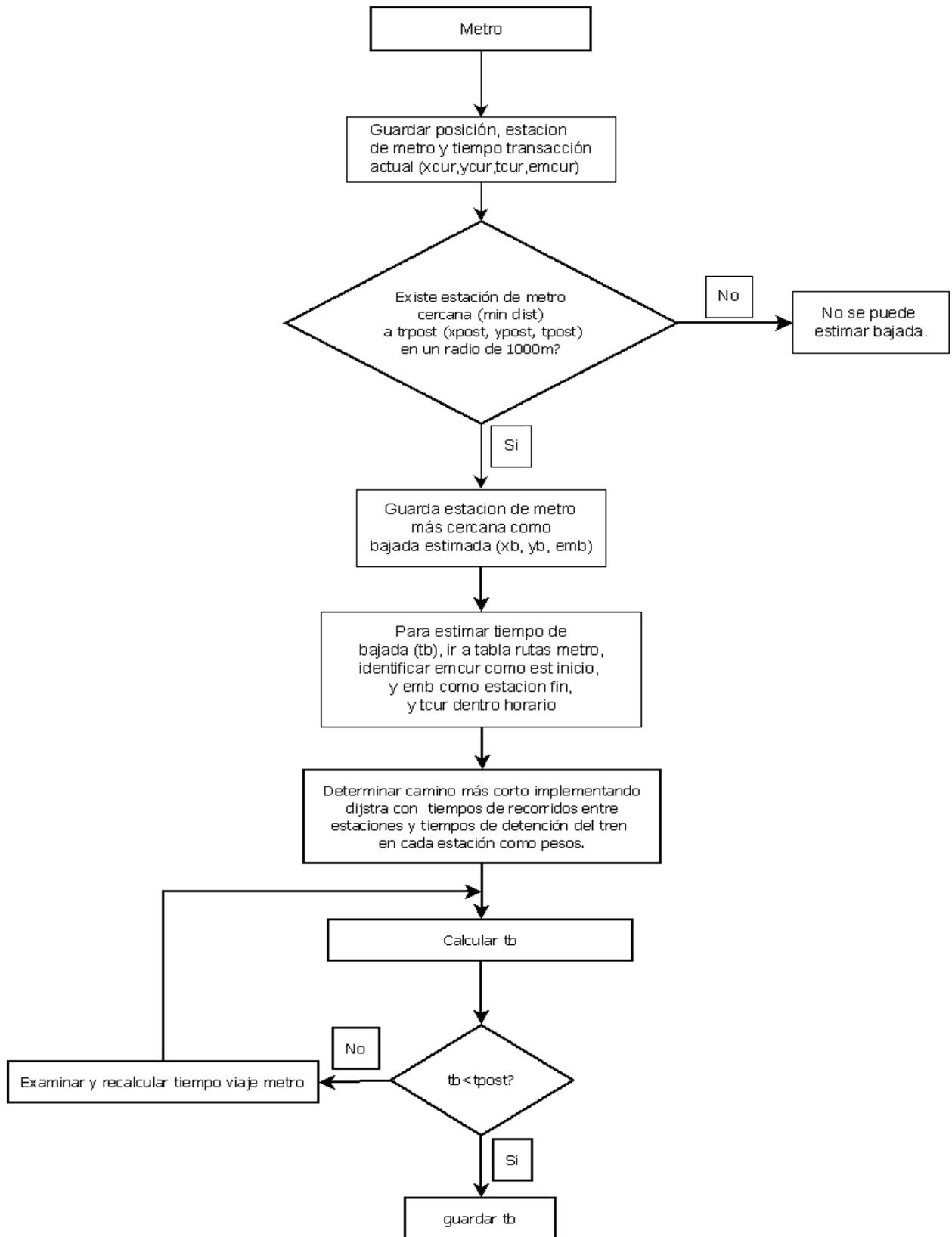


Figura 4-13: Diagrama de la estimación del punto de bajada, para transacciones realizadas en metro.

4.2.1.3 Caso Zona Paga

Para el caso de zona paga el proceso es más complejo ya que no se conoce el bus que se abordó, y se sabe que resulta imprescindible contar con la asignación de algún bus a la transacción realizada por la tarjeta, para poder estimar el punto de bajada de la etapa iniciada con dicha transacción.

La información disponible para zonas pagas detalla los servicios que pasan por cada una de ellas (Ver Figura 3-2), con lo cual se conoce a priori los posibles servicios abordados por la transacción. Aprovechando que se determinó que el punto de bajada a estimar debe ser cercano a la siguiente transacción, se filtran en un primer paso, aquellos posibles servicios (que pasan por la zona paga donde se realiza la transacción) con algún paradero de su ruta que esté en un radio menor a 1000m con respecto a la posición de la siguiente transacción.

Una vez obtenido el conjunto de servicios que le sirven a la transacción actual dada la cercanía de las rutas con el punto de la siguiente transacción, se busca asignar la etapa de viaje a un servicio. Si no existe servicio que cumpla con el filtro de distancia la etapa de viaje queda sin estimación de punto de bajada, registrando como causa “No caminable”.

La asignación de transacción a un servicio se realiza en base al enfoque de ruta mínimas sin restricción de capacidad propuesto por Chriqui y Robillard (1975). Se considera que el usuario (transacción) elige su alternativa dependiendo de los tiempos de espera y tiempos de viaje promedio observados históricamente, no considerando restricción de capacidad en paraderos.

Para el problema se considera:

- Esperanza de tiempo de espera: $\left\{ \frac{k}{f_1}, \frac{k}{f_2}, \frac{k}{f_3}, \dots, \frac{k}{f_n} \right\}$

Donde k : constante de espera = $\frac{1}{2}$

f_i : frecuencia servicio i (teórica), estimada a partir de la velocidad promedio observada de la ruta, en horario establecido para zona paga.

- Tiempo en vehículo: $\{tv_1, tv_2, tv_3, \dots, tv_n\}$

Donde tv_i : tiempo de viaje en ruta i (estimado a partir del mínimo tiempo generalizado, el cual se obtiene de $\min_p t_p + 2,5 \frac{d_{p-b}}{v_c}$, Ver Ecuación 4-10)

p : paradero en la ruta i .

Luego se minimiza el tiempo total esperado de viaje TVE (asignación rutas mínimas), para encontrar el conjunto de líneas comunes (L_a).

$$\text{Sea } X = \{x_1, x_2, \dots, x_n\} \quad x_i = \begin{cases} 1 & \text{si } l_i \in L_a \\ 0 & \text{si no} \end{cases}$$

$$\min_x TVE(r, b) = \min_x \left\{ \frac{k + \sum_i tv_i \cdot f_i \cdot x_i}{\sum_i f_i \cdot x_i} \right\}$$

s. a

$$x_i = \{0,1\}$$

Ecuación 4-14

Dado que para los servicios de zonas pagas se tiene la secuencia de paraderos de las rutas de dichos servicios, y se cuenta con el perfil de velocidades para ellas, se decide trabajar con paradas en vez de puntos gps, esto con el fin de mejorar la modelación del problema ya que se acerca más a la forma de operación del sistema (bajada y subida en paraderos formales).

Una vez encontrado el conjunto de líneas comunes, se asume que la transacción toma el servicio que primero pasa, es decir se asigna aquel bus que primero pasa de los servicios establecidos como líneas comunes, identificándolo mediante su PPU (Ver Figura 4-14).

Para estimar el punto de bajada se realiza el proceso de estimación de bajada para el caso bus descrito en la sección 4.2.1.1 (Ver Figura 4-12).

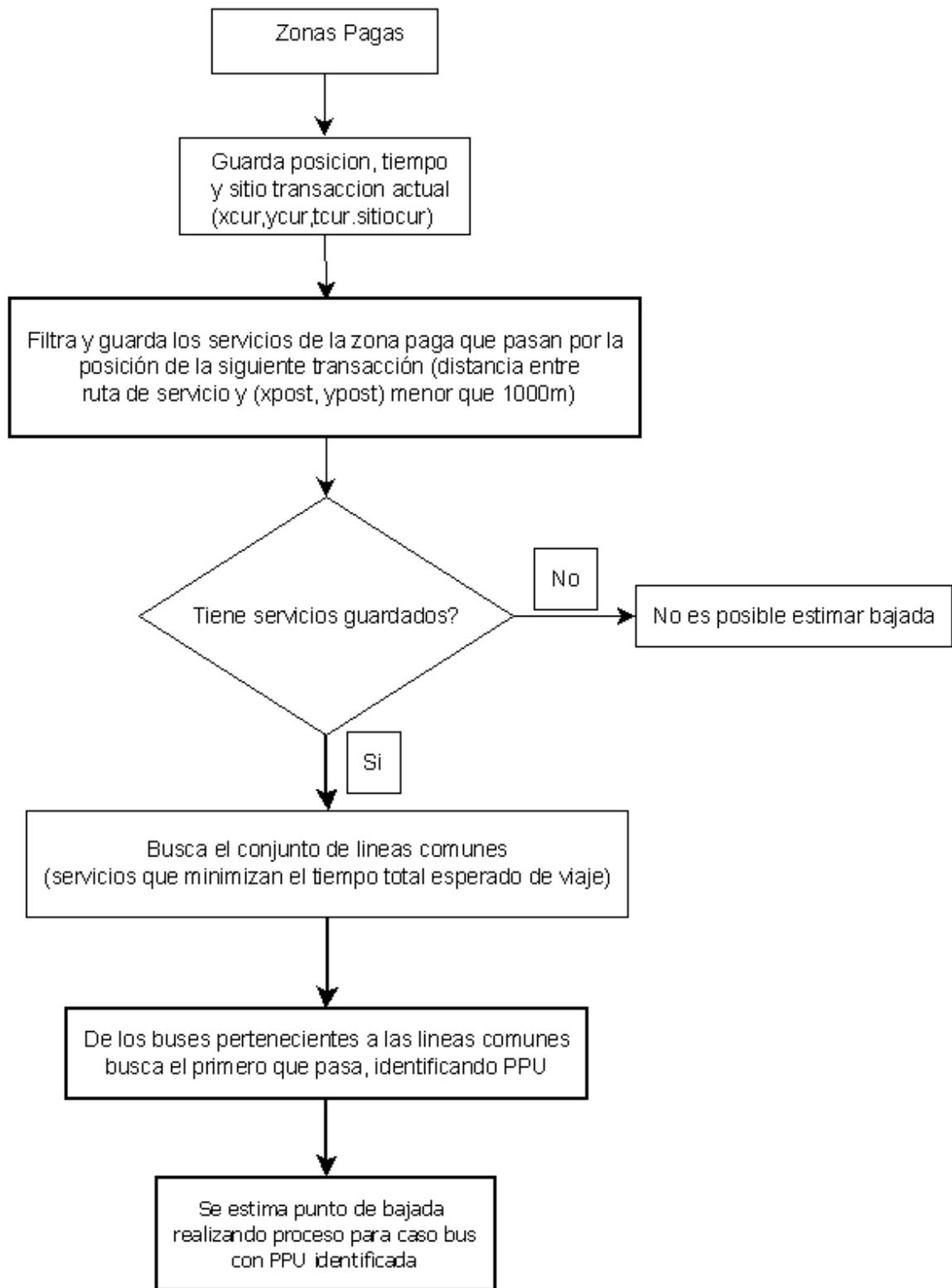


Figura 4-14: Diagrama de la estimación del punto de bajada, para transacciones realizadas en zona paga.

4.2.2 Recuperación y deducción de información

La recuperación y deducción de información se realiza considerando las causas que se detectaron que provocan información incompleta o errónea: transacción sin posición, transacción posterior sin posición, solo una transacción diaria, falla ruta.

De acuerdo a las causas encontradas, se plantean dos maneras de enfrentar la pérdida ó insuficiencia de información: deduciendo la posición de origen ó infiriendo la posición de bajada.

Deducir la posición de origen se aplica para aquellas transacciones sin posición o bien para aquellas que presentan falla en la ruta (información de posicionamiento con errores).

La deducción de la posición de origen para transacciones sin posición, da solución a los casos cuya transacción posterior correspondía a una transacción sin posición, ya que se recupera la posición de la transacción posterior. Esto permite realizar el proceso de estimación de bajada para aquellas etapas de viaje que no presentaban posición en la transacción posterior.

En tanto para las etapas de viaje comenzadas con una transacción sin posición realizada en bus o una transacción hecha en algún bus que tenga deficiencias en el registro de posicionamiento (falla ruta), no se puede realizar el proceso de estimación de bajada ya que no se puede realizar el correcto seguimiento del bus. Debiendo inferir y deducir el punto y hora de bajada a partir del comportamiento de la tarjeta si es posible.

Para aquellas etapas de viaje comenzadas con una transacción sin posición realizada en zona paga, una vez deducida la posición de origen se realiza el proceso de estimación de bajada para zona paga (Ver Figura 4-14) recuperando la estimación del punto de bajada.

En la Figura 4-15 se muestra en general el procedimiento descrito.

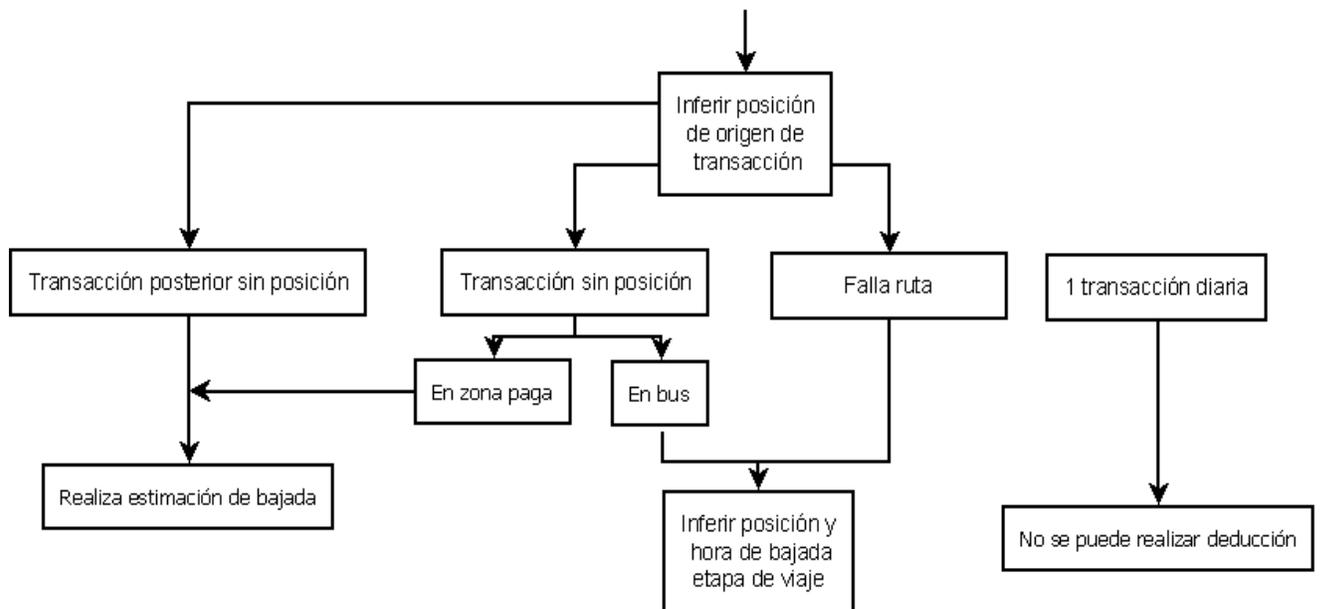


Figura 4-15: Diagrama global para deducción y recuperación de información.

Los criterios para inferir se detallan a continuación:

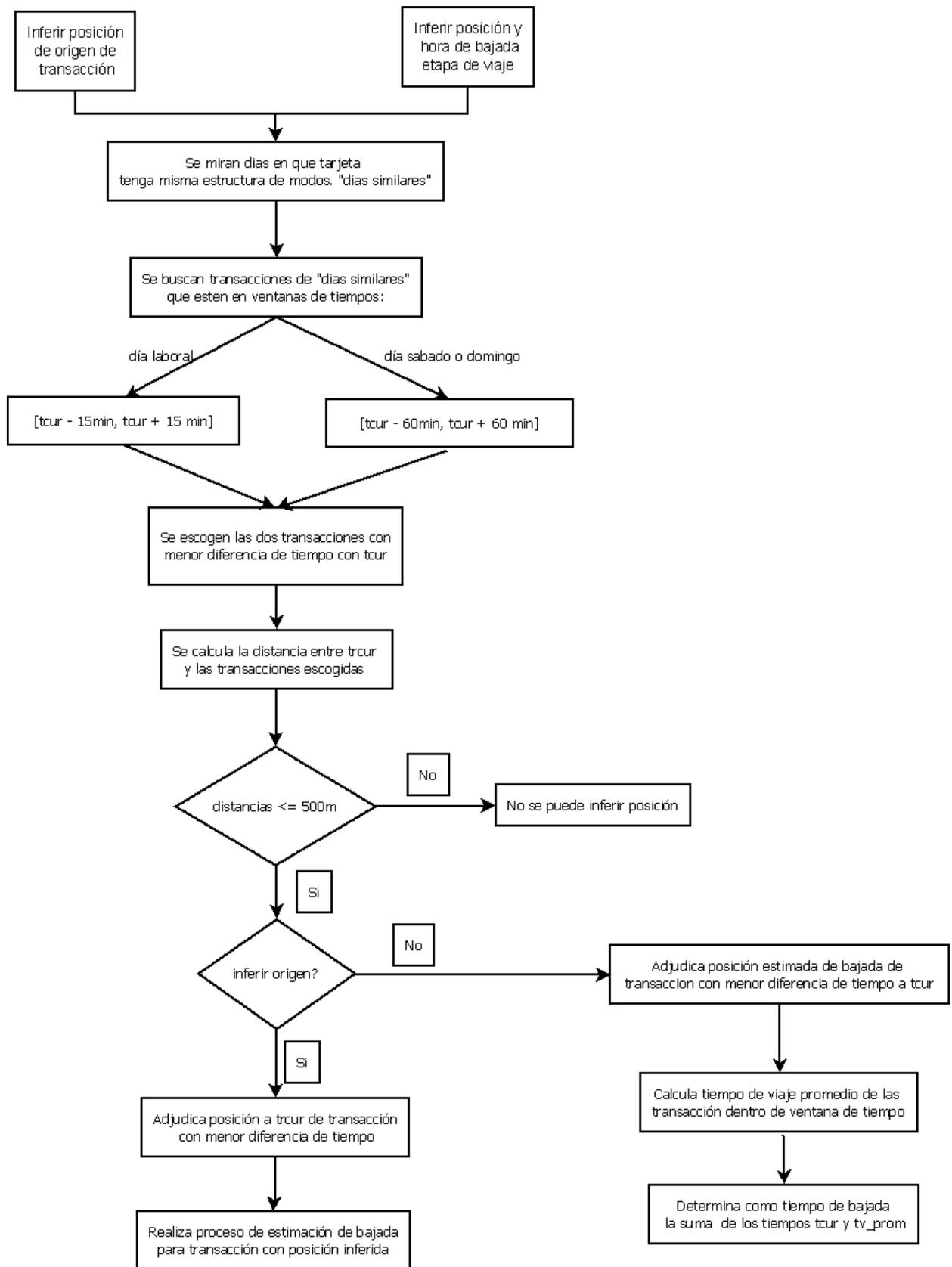


Figura 4-16: Proceso de recuperación y deducción de información.

Cabe destacar que las causas determinadas y reconocidas en esta metodología no son únicas, y que los parámetros de coincidencia temporal y espacial son arbitrarios, debiendo en el futuro ser validados.

4.2.3 Cálculo de tiempos

En esta etapa se determina la existencia de transbordo mediante el cálculo de tiempos de viaje de la etapa, tiempo entre etapas, distancia entre etapas. Se considera que las etapas de viaje cuya diferencia de tiempo entre la bajada actual de etapa y la subida de la transacción posterior es mayor a 45 min, realizan una actividad en el lugar de bajada.

Este valor se escogió considerando que para realizar alguna actividad con algún propósito se necesitan al menos 30-40 minutos. Y que además el caso más desfavorable es aquel que d_r son 1000m, tomando un tiempo de caminata de 11 minutos a la posición de la siguiente transacción, con lo cual se está considerando un máximo tiempo de espera de 34 minutos, valor que parece razonable debido a que los tiempos de espera en el sistema fluctúan dependiendo del tipo de servicio y de la hora del día. Estos valores son totalmente arbitrarios y deben ser validados en el futuro.

Al calcular t_{cam} se verifica que el tiempo de caminata obtenido sea menor al tiempo de transbordo, permitiendo un rango de error de 2 minutos, el cual pretende recoger la incertidumbre sobre la precisión del tiempo de bajada estimado.

Que el tiempo de caminata sea mayor al tiempo de transbordo solo se explica si la velocidad de caminata es muy superior a la considerada, por lo cual se concluye que en estos casos existe una etapa de viaje no registrada, ya sea porque se hizo en algún modo que no registra el sistema (taxi, bicicleta, auto) o porque no hay validación de esta etapa en el sistema.

Los tiempos y distancias se calculan según las ecuaciones 4-3 a 4-8.

El proceso de este paso de la metodología se muestra en la Figura 4-17 exhibida a continuación.

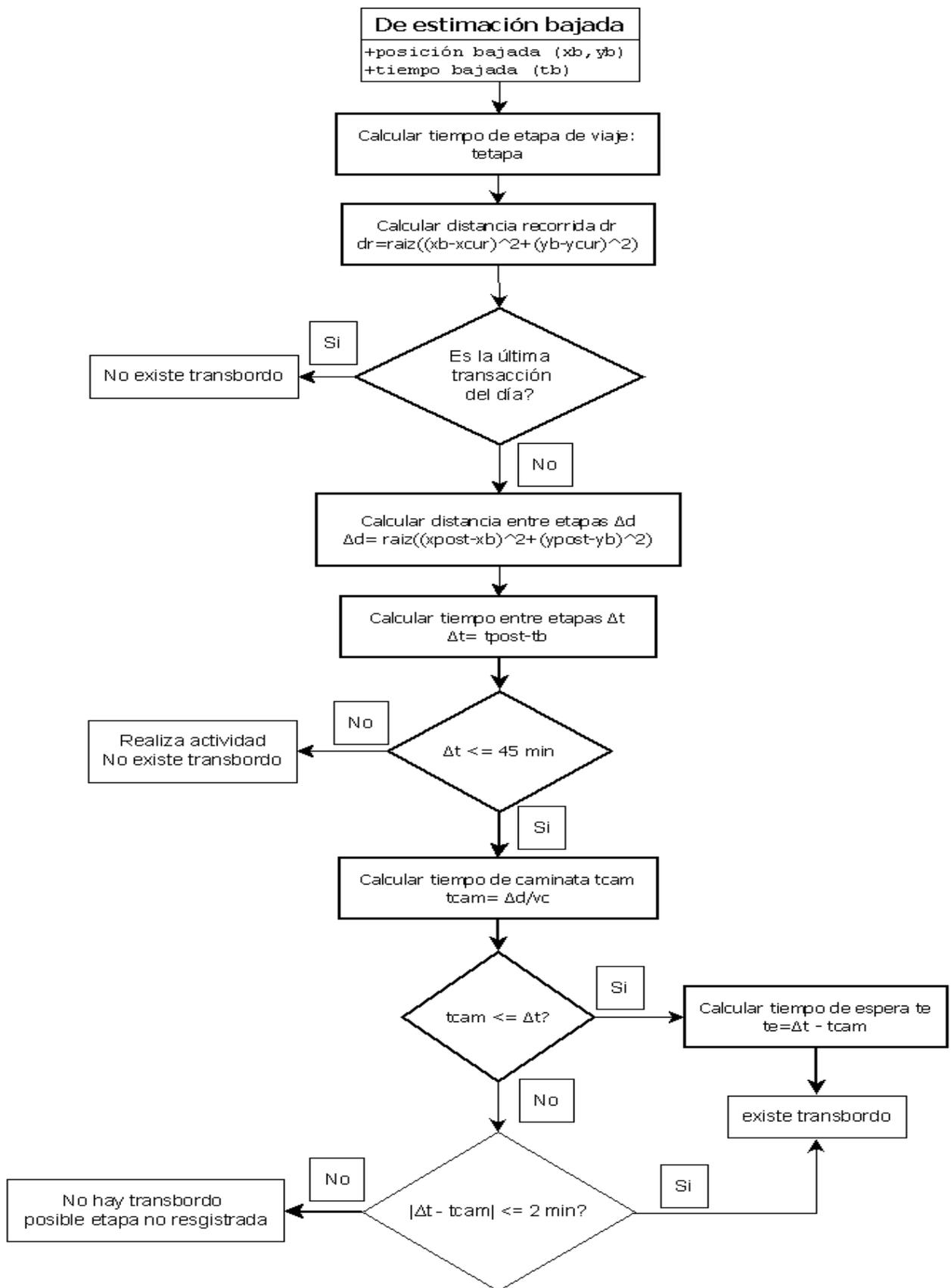


Figura 4-17: Diagrama para determinar la existencia de transbordo.

4.2.4 Identificación de etapa de viaje y viaje

Una vez determinada la existencia de transbordo para aquellas etapas que presentan estimación de bajada, conociendo además aquellas cuyo destino no se conoce porque Δd es mayor a 1000m y la diferencia de tiempo entre la transacción actual y la transacción posterior, se pueden reconstruir los viajes de las tarjetas, identificando cuantos viajes se realizan y la estructura de estos (N° de etapas) (Ver Figura 4-18).

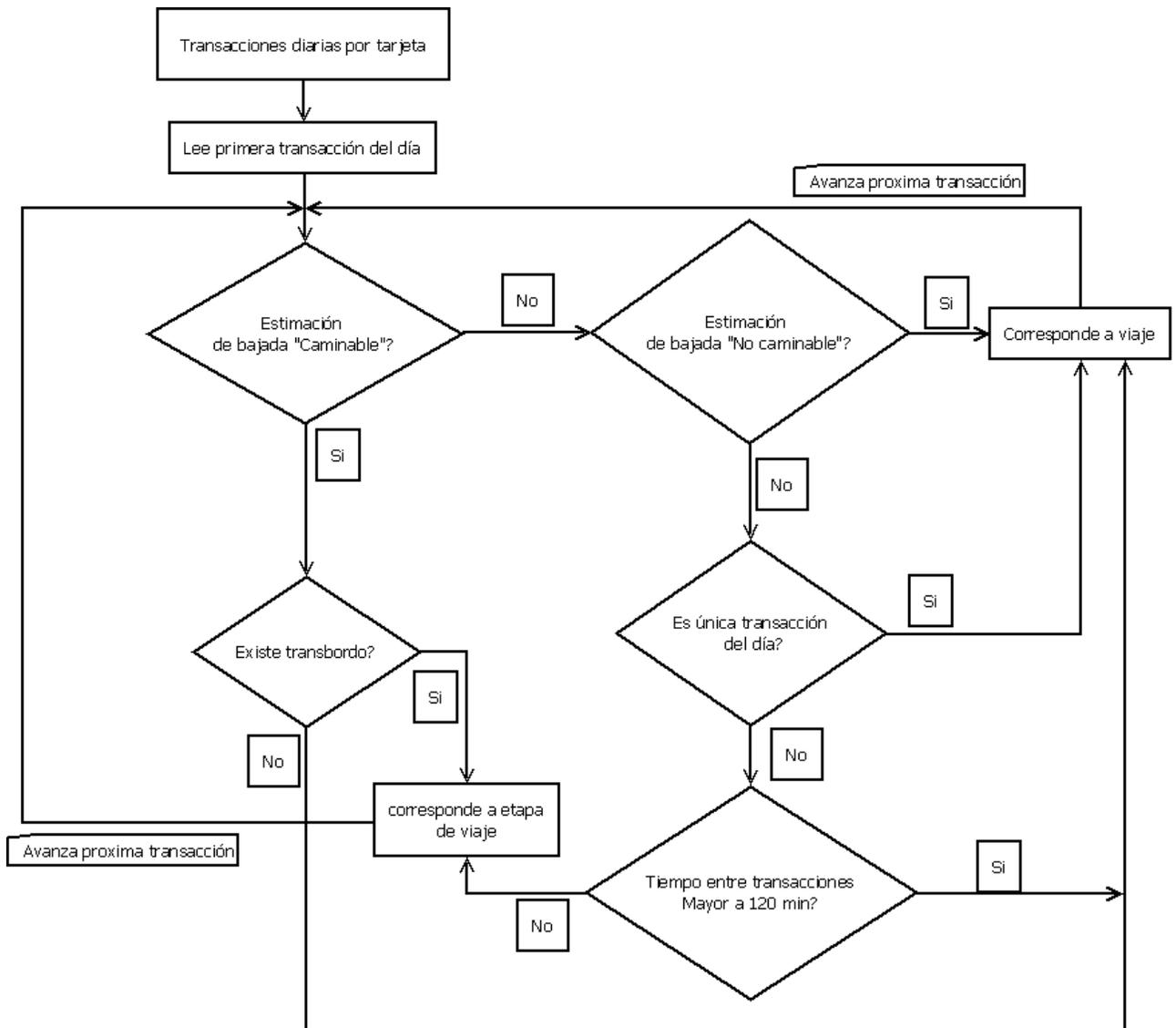


Figura 4-18: Diagrama para identificar etapa de viaje de viaje.

Para el caso en que no se tiene información sobre la bajada de la etapa de viaje, se adopta el criterio del sistema donde toda transacción posterior dentro de una ventana de tiempo de 2 horas se considera transbordo.

4.2.5 Información relevante

Una vez realizados los procesos anteriores se confecciona una nueva tabla donde se reúne la información completa sobre las transacciones de cada tarjeta, presentando el seguimiento realizado.

En esta tabla cada línea representa un viaje realizado por cada tarjeta, el cual está compuesto por segmentos (etapas) determinadas mediante la metodología. Los atributos de esta tabla se detallan a continuación (Tabla 4-4).

Tabla 4-4: Campos de información tabla de salida.

Output Metodología	
<ul style="list-style-type: none"> • ID • Tipo de tarjeta (Adulto/Estudiante) • Fecha • N_v • Origen <ul style="list-style-type: none"> • X_{origen} • Y_{origen} • $hora_{origen}$ • Destino <ul style="list-style-type: none"> • $X_{destino}$ • $Y_{destino}$ • $hora_{destino}$ • Características Viaje <ul style="list-style-type: none"> • t_v • d_r • N° de etapas • N° de transbordos 	<ul style="list-style-type: none"> • Etapa 1 <ul style="list-style-type: none"> • Servicio • Sitio • Punto de subida <ul style="list-style-type: none"> • $X_{puntosubida}$ • $Y_{puntosubida}$ • $hora_{puntosubida}$ • Punto de bajada <ul style="list-style-type: none"> • $X_{puntobajada}$ • $Y_{puntobajada}$ • $hora_{puntobajada}$ • Características Etapa <ul style="list-style-type: none"> • t_{etapa} • d_{retapa} • Δt • Δd • t_{cam} • t_e • • • Etapa N

Donde

- N_v : Número correlativo de viaje (para cada tarjeta) dentro del día.
- t_v : tiempo de viaje.
- d_r : Distancia recorrida en el viaje
- N° de transbordo: Número de etapas menos 1.

- Servicio: Para viajes/etapas en bus (transacciones realizadas en bus o zona paga) corresponde a línea de bus abordada (recorrido). Para metro línea de metro abordada.
- Sitio: Lugar donde se realizó transacción que da comienzo a viaje/etapa. Corresponde a placa patente única del bus, paradero zona paga, estación de metro.
- t_{etapa} : tiempo de etapa.
- d_{retapa} : Distancia recorrida en etapa
- Δt : Tiempo entre etapas, compuesto por tiempo de caminata (t_{cam}) y tiempo de espera (t_e).
- Δd : Distancia física entre etapas, corresponde a la d_{b-s} obtenida del proceso.

A partir de esta tabla se puede generar perfiles de carga e identificar lugares de interés, realizar una serie de estadísticas sobre los viajes y su estructura como lo son: histograma de viajes diarios por tarjeta, histograma de etapas por viaje, histograma tiempos de viajes, histograma de transbordo y tiempos de caminata. Estos análisis también se pueden realizar para un subconjunto de la tabla de acuerdo a alguna agregación, como tipo de tarjeta, períodos específicos del día, origen o destino determinados.

5 Conclusiones y líneas futuras de investigación

5.1 Comentarios y Conclusiones

La información disponible a partir de la utilización de la tarjeta BIP, en conjunto con los datos entregados por el sistema de posicionamiento global (GPS) existente en los buses y el conocimiento de la operación del sistema de transporte, otorgan una gran cantidad de información la cual permite realizar un seguimiento de las tarjetas a través del tiempo y en la ciudad.

De la información disponible se determinó que las principales fortalezas radican en los altos porcentajes de información correcta y útil que presentan las bases de datos, mostrando un 97% de información completa para la ubicación de transacciones (lo que implica la posibilidad de tener dicho porcentaje de estimaciones de punto de bajada), y un 87% de registros de posición útiles que permiten el seguimiento de los usuarios. Además se identificó las causas de los errores que se presentan en los datos, los cuales se deben a fallas de los registros de posición gps, siendo estas principalmente fallas intermitentes que afectan aleatoriamente a transacciones en diferentes buses a lo largo de la semana. También existe un pequeño porcentaje de buses (3,5%) que presenta falta de información de posicionamiento a lo largo de la semana, lo que indica que estos buses no funcionaron con su gps prendido o bien no contaban con él.

Este trabajo plantea una metodología de cinco etapas, y en cada una de ellas se lleva a cabo un proceso que permite enriquecer la información disponible con el objetivo final de realizar un seguimiento de la tarjeta que proporciona información completa sobre los viajes realizados por un usuario.

La metodología establecida considera un análisis minucioso de la interacción entre los elementos presente en el sistema. Resolviendo problemas tanto de baja dificultad (supuestos base para estimar punto de bajada), como de mayor complejidad (asignación de pasajero a bus en zona paga). La estimación de bajada se plantea en base al lugar donde se realiza la transacción (bus, metro o zona paga) de acuerdo a la dinámica del sistema, estableciendo claramente el procedimiento para cada caso, incorporando las características de sitio y operación del sistema en la metodología.

La principal fortaleza de esta metodología es la forma de realizar la estimación de bajada, la cual enriquece los métodos encontrados en la literatura, modelando la elección del usuario entre caminar una distancia un poco menor y estar más tiempo sobre el bus versus, tener un mayor tiempo de caminata y permanecer menos abordo del bus. El concepto de tiempo generalizado que recoge esta elección, se aplica para conocer el punto de bajada de las etapas de viaje realizadas en bus, determinando como punto de bajada aquel punto gps que minimiza el tiempo generalizado del usuario. Este criterio para encontrar el punto más cercano a la siguiente transacción

incorporando la variable tiempo de viaje, permite solucionar problemas que se diagnosticaron con el criterio de mínima distancia, donde en casos de puntos gps cercanos pero con diferencias de tiempo grandes no se distinguía el punto que correspondía a la trayectoria en la dirección del viaje del usuario.

La metodología de estimación de bajada se puede plantear para tener como resultado final paradas de subida o bajada en lugar de puntos gps, este resultado es factible de obtener para aquellas etapas de viaje realizadas en buses que presentan asignación de servicio (y por lo tanto se puede conocer la secuencia de parada definidas para cada ruta), esto permite una mejor modelación del problema, ya que formalmente los usuarios del transporte público se suben en paraderos, pero se pierde precisión en la estimación de posición y de tiempo de bajada ya que para obtener las paradas se debe proyectar ortogonalmente a la ruta el punto de subida ó bajada estimado y calcular la distancia euclidiana entre el punto y los paraderos inmediatos a él. Para el caso del punto de subida es claro asignar el paradero anterior al punto de subida (ya que no se puede validar la transacción antes de haber subido al bus) teniendo como resultado final una parada de subida. En tanto para el caso del punto de bajada la asignación de paradero depende de la magnitud de la distancia entre el punto y la parada anterior y el punto y la parada posterior.

Para plantear la metodología con paradas de subida y bajada se debe tener claro la repercusión de las asignaciones de paradas, en los tiempos y distancias entre etapas, ya que pueden afectar directamente la correcta estimación de ellas. Para la implementación se hace necesario contar con las correctas secuencias de paradas para cada ruta, incorporando los desvíos que existen en ellas algunos días de la semana o bien las diferentes rutas que pueden existir según periodos del día, además se debe disponer de información que presente una correcta y completa asignación de servicio.

También se destaca la solución generada para el caso zonas pagas, donde no se conoce a priori el bus que se abordó, por lo cual se plantea un proceso que combina el conocimiento de saber que la ruta del bus que abordó pasa cerca de la siguiente transacción junto con la asignación de servicio mediante el método de líneas comunes y asignación a rutas mínimas.

Para los criterios de asignación se supone que no existe congestión en paraderos y que los buses no tienen restricción de capacidad, este supuesto se extiende para los tiempos de viaje en metro, donde tampoco se considera congestión en andenes, asumiendo que el usuario puede abordar el bus o tren de metro en cualquier momento.

Dada la gran cantidad de información presente en el sistema, y los múltiples comportamientos de los usuarios que se presentan en la información disponible, resulta ambicioso identificar cada comportamiento. Este trabajo se enfocó en estudiar casos de

comportamiento regular a lo largo de la semana, dirigiendo el planteamiento de la metodología principalmente a estos casos. No obstante ello, el procedimiento para estimar bajada propuesto se aplica a la totalidad de la base de datos.

En la etapa de identificación de la información disponible se analizaron de manera gruesa la composición de la información existente, diagnosticando comportamientos irregulares, pero con cierto patrón a lo largo de la semana, y casos de transacciones seguidas realizadas por una misma tarjeta. En este análisis breve que se realizó se diagnosticó que aproximadamente 68.000 tarjetas realizan el mismo número de transacciones de lunes a viernes, 131.000 tarjetas realizan el mismo número de transacciones de lunes a jueves, 172.000 tarjetas realizan el mismo número de transacciones los días lunes, miércoles y viernes y 492.000 tarjetas el mismo número de transacciones los días martes y jueves. Con los resultados de estimación de bajada que se obtengan al implementar la metodología se debe analizar y profundizar en los comportamientos que se presentan, determinando el alcance de la metodología en el reconocimiento de patrones de comportamientos.

En cuanto a las transacciones realizadas por una misma tarjeta con un intervalo de tiempo entre transacciones menor a 30 segundos, se identificaron aproximadamente 300.000 tarjetas que presentan dicho comportamiento, siendo casi un 50% de ellas tarjetas que presentan dos transacciones en menos de 5 segundos.

De los 5 procesos establecidos, sólo el proceso de estimación de bajada está siendo programado, los 4 restantes, Recuperación y deducción de información, Cálculo de tiempos, Identificación de etapa de viaje y viaje, e Información relevante se encuentran a nivel conceptual.

En este trabajo se pretendía obtener resultados de la implementación de la metodología planteada, permitiendo un proceso de retroalimentación y validación de parámetros usados en ella. Esto no se logro debido al tiempo destinado a estudiar la información disponible y al proceso de implementación de esta metodología que aún no arroja resultados.

El resultado de implementar esta metodología debería proporcionar información completa que permite determinar comportamiento del usuario, perfiles de carga, generar matriz O-D, todos objetivos específicos de este trabajo. Además de reconstruir los viajes y generar patrones de éstos, de manera de conocer el comportamiento de las tarjetas en el sistema, y ser una herramienta de apoyo para la planificación del sistema de transporte público.

5.2 Líneas Futuras de investigación

El campo de estudio en que se enmarca este trabajo presenta una amplia gama de líneas de investigación futuras que resultan interesantes profundizar y que continúan con el desarrollo de este trabajo, algunas de las que se recomiendan son:

Analizar de resultados una vez implementada la metodología, validando el planteamiento de cada uno de los procesos establecidos y ajustando los valores de los parámetros establecidos arbitrariamente. Analizar y estudiar el comportamiento presentado por los usuarios, identificando patrones presentes en ellos los cuales pudieran ser identificados de manera automática en el futuro mediante la adaptación de los parámetros según horario del día o tipo de servicio.

Generar una muestra de control, donde usuarios identificados registren una bitácora de viajes con horarios, origen, destino, propósito y estructura de los viajes realizados, comparando dichos viajes con los resultados arrojados para la muestra por la metodología, retroalimentado el planteamiento de ella.

Realizar seguimiento a las tarjetas reconociendo patrones de viaje de los usuarios según períodos de tiempo dentro del día, de la semana, ó según tipo de usuario (estudiante, adulto), generar perfiles de carga a lo largo del día como espacialmente, identificando aquellas zonas de atracción y generación de actividades.

Identificar de los perfiles de carga, las paradas “naturales” existentes en cada ruta, contrarrestando con la secuencia de paradas establecida para dicha ruta, estableciendo los rangos de precisión y error asociados a utilizar punto gps en lugar de parada.

Profundizar en errores y sesgos encontrados, tratando de identificar aquellas transacciones sin posición producto de fallas intermitentes y que afectan aleatoriamente a diferentes usuarios en distintos buses, así como también aquellas que son producto de la falla prolongada en la emisión de registros gps y que afectan a todas las transacciones realizadas en cierto en cierto periodo de tiempo. Establecer que campos de la información se deben mejorar en el procesamiento original de la información para tener herramientas directas que permitan evaluar y reconocer dichas transacciones.

Establecer la metodología para generar la matriz O-D a partir de la información procesada y estimada en esta investigación, obteniendo factores de expansión y complementándola con encuestas que integren información acerca de los viajes no motorizados y en modos particulares (autos, taxi), comparar los resultados obtenidos con las encuesta OD 200, identificando similitudes y diferencias.

Mejorar la información proporcionada por metro, realizando mediciones en las distintas estaciones de tiempos de transbordo y tiempos andén-torniquete, incorporándolas en la obtención de los tiempos de viaje.