A First Approach to the Optimization of Bogotá’s TransMilenio BRT System

Francisco J. Peña, Antonio Jiménez and Alfonso Mateos
Decision Analysis and Statistics Group (DASG)
Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid
Campus de Montegancedo S/N, Boadilla del Monte, 28660, Madrid, SPAIN
Email: francisco.pena.escobar@alumnos.upm.es, {ajimenez, amateos}@fi.upm.es
http://www.dia.fi.upm.es/grupos/dasg/index.htm

Abstract—Bus rapid transit (BRT) systems are massive transport systems with medium/high capacity, high quality service and low infrastructure and operating costs. TransMilenio is Bogotá’s most important mass transportation system and one of the biggest BRT systems in the world, although it only has completed its third construction phase out of a total of eight. In this paper we review the proposals in the literature to optimize BRT system operation, with a special emphasis on TransMilenio, and propose a mathematical model that adapts elements of the above proposals and incorporates novel elements accounting for the features of TransMilenio system.

Keywords—Bus rapid transit systems, optimization, mathematical modeling.

I. INTRODUCTION

Bus rapid transit (BRT) systems are public transport systems with medium/high capacity, high quality service and low infrastructure and operating costs (\cite{1}). They are considered to be a good affordable alternative for developing cities seeking to provide their citizens with a high-quality possible self-sustaining public transport alternative comparable with rail systems, but without the high costs and without taking cities to high levels of debt, leaving the possibility of investing the city funds in priority areas such as health or education.

BRT systems have a lot in common with rail systems, particularly performance and passenger service. The main difference is that operation and implementation costs are 4 to 20 times lower than the costs of a light rail system, and 10 to 100 times lower compared to a heavy rail and metro system (\cite{1}).

They can operate of limited stop services (also called stop-skipping services), in which a bus service omits stops along certain routes. This has great advantages, such as the reduction of travel times due to fewer stops and the reduction of operator costs because they can meet the demand with fewer vehicles thanks to shorter bus cycles (\cite{2}).

BRT systems are now operating in 149 cities, most of which have been built since 2000, and 84 more are planned around the world. TransMilenio is Bogotá’s most important public transportation system and one of the biggest BRT systems in the world. New plans have been made to expand it due to its success, and similar systems have been constructed in other cities of Colombia.

There are very few proposals in the literature focused on optimizing the BRT system operation, mainly because they are relatively recent phenomena, and many of the currently operating BRT systems are far from reaching maximum capacity. To the best of our knowledge, there are not automatic proposals for route design. The closest to this is the model proposed in (\cite{2}) that evaluates and selects the best several routes.

In this paper we review the proposals in the literature to optimize BRT system operation, with a special emphasis on TransMilenio, and propose a mathematical model that adapts elements of the above proposals and incorporates novel elements accounting for the features of TransMilenio system. Specifically, we introduce a new model for evaluating TransMilenio BRT system routes, given the trip demand in the form of an origin-destination matrix.

Section 2 introduces BRT systems and their main elements. Section 3 focuses on the TransMilenio system, Bogotá’s most important public transportation system and one of the biggest BRT systems in the world. In Section 4, we review the different studies in the literature on the optimization of BRT systems and, specifically, on the TransMilenio system. In Section 5, we introduce a new mathematical model approach to the optimization of the TransMilenio system. Finally, some conclusions and future research are discussed in Section 6.

II. BUS RAPID TRANSIT SYSTEMS

A BRT system was defined in \cite{1} as a system based on high quality buses, that provide fast and a comfortable urban mobility and with a favourable cost-benefit through the provision of segregated infrastructure of exclusive use, fast and frequent operations, and marketing and customer/user service excellence.

The first BRT system started operating in Curitiba in 1974, but until the decade of 1990 this type of system was seen as a public transportation system for small cities or as complementary systems of a metro network. Many experts considered that these systems were not able to reach a capacity beyond 12000 passengers per hour per direction (pphpd). This perception radically changed in 2000 with the creation of TransMilenio in the city of Bogotá (Colombia). Nowadays, TransMilenio transports nearly 500 million people yearly (\cite{3}). It introduced a series of improvements that raised the capacity of BRT systems enormously to 45000 pphpd, and has inspired
many cities around the world to implement this type of systems ([1]).

Nowadays, there are 149 cities with BRT systems, and 84 more are planned. The majority of systems were built after the year 2000, which can be attributed to TransMilenio’s success, as illustrated in Figure 1, which shows the evolution in the number of cities with BRT systems per decade and the respective number of kilometers.

A. Comparison with other mass transportation systems

Table I shows the price range for mass transportation systems based on a comparison of infrastructure costs real data ([1]).

<table>
<thead>
<tr>
<th>Type of system</th>
<th>Cost per kilometer (US$ million/km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRT</td>
<td>0.5 - 15</td>
</tr>
<tr>
<td>Tram and light rail transit</td>
<td>13 - 40</td>
</tr>
<tr>
<td>Elevated systems</td>
<td>40 - 100</td>
</tr>
<tr>
<td>Underground metro</td>
<td>45 - 350</td>
</tr>
</tbody>
</table>

TABLE I. CAPITAL COSTS FOR DIFFERENT MASS TRANSPORTATION SYSTEMS

The infrastructure costs for BRT systems are clearly significantly lower than for any other rail-based transportation system. The city of Bangkok is a case in point. This city has an elevated rail system (SkyTrain) and an underground metro system (MRTA), a proposed BRT system (Smartway) and a proposed light rail train. The real costs per kilometer of the metro systems and elevated train were US$ 142.9 million and US$ 72.5 million. The projected costs per kilometer for the proposed light rail and BRT are US$ 25 million and US$ 2.34 million. This means that with a budget of US$ 1000 million they could build 7 km of underground train, 14 km of elevated train, 40 km of light rail train or 426 km of BRT system ([1]).

Unlike rail systems around the world, BRT systems are capable of operating without government subsidies. As a matter of fact, they are profitable, which is the reason why many governments delegate the operation to private companies. This is a great advantage, especially useful for developing cities, where governments have tight budgets and there is nothing better than a self-sustaining mass transportation system thanks to which they can invest resources in other areas such as sewerage, education and health.

BRT systems can be planned and implemented in short time periods, which can be covered in one government term. The two most successful and complete BRT systems (Curitiba and Bogotá) were planned and implemented in a three year span.

Formerly it was thought that bus based services could operate within a range up to 6000 pphpd. If the demand was higher, a light rail based system should be considered, with capacity between 6000 and 12000 pphpd. A heavy metro system had to be considered for a higher demand, since its capacity ranges from 25000 to 80000 pphpd.

The arrival of BRT systems with a capacity range from 3000 to 45000 pphpd changed the situation. BRT systems turned into a real mass transportation alternative for big cities, and the myth that says that BRT system cannot compete with metro systems in terms of capacity was broken. As a matter of fact, it is not necessarily true that big cities need overflowing levels of capacity, an example is the London metro system, which has a capacity of 30000 pphpd, but thanks to its multiple parallel lines it has distributed corridors demand across the entire transportation network. An opposite case is Hong Kong’s metro, whose capacity is 80000 pphpd and there is only one line from Kowloon and New Territories to Nathan Road. But the high level of demand is what makes this metro system profitable ([1]).

B. Main components of a BRT system

A BRT system has seven main components ([5],[6]): busways, stations, vehicles, fare collection, intelligent transportation systems, service and operation plans, and branding elements.

The busways or corridors are the main component of the BRT systems and it is where the vehicles circulate; they are like the rails of a metro system. They are also the most expensive and one of the most visible elements of the whole system. Therefore, they have a direct impact on the image and perception the users have of the system ([6]). The busways must be exclusive for the system buses. Furthermore, the busways must be located in the center and not at the side of the road ([1]).

The stations are the link between the passengers, the BRT system and other transportation systems. They are the element that has most influence on system image, and so, they must have comfortable facilities so that the passengers feel at ease. The stations must accommodate many more people than a bus stop, so they must have a wide infrastructure ([6]) since they are located in high demand busways. Besides, the stations must protect the users from climate conditions. The boarding platforms must be at the same height as the buses floor to ease and speed up passenger’ access. There must be also large capacity header stations at the ends of each busway to integrate busways and feeder routes [1].

The vehicles are the system’s element in which the passengers spend most of their time. They have a direct impact on speed, capacity, comfort and environment friendliness.
are the element that most non-users see, becoming one of the elements with most influence over the public’s perception of the system ([6]).

Currently, there are three types of vehicles: articulated, bi-articulated and simple. **Articulated buses** have the capacity for 160 passengers and operate within the busways. **Bi-articulated buses** have the capacity between 240 and 270 passengers and operate within busways. **Simple buses** have the capacity for 90 passengers and in some cities they operate only in mixed traffic corridors as feeder routes. In other cities with low capacity BRT systems operate in busways.

**Fare collection** has a direct effect on capacity and the system’s income. If fares are collected outside the bus, it decreases passenger waiting time through bus boarding efficiency. This is especially useful for bus routes that have high levels of demand ([6]). The fare may be collected outside the bus at the station entrance. Furthermore, fares should be integrated, that is, users should be able to transfer from one bus route to another, including feeder routes, without having to pay an extra fare [1].

The **intelligent transportation system** is a technological component that helps to improve overall system performance. It is a combination of different technologies to retrieve all kind of data about system operation, from the number of passengers that enter the system to the positioning of every vehicle using GPS, vehicle departure times, traffic conditions, the traffic lights, etc. The goal of this component is to collect and transform all the possible information into useful knowledge for operators, and ultimately benefits for the passengers.

The **service and operation plans** directly affect the user’s perception of the system. A good plan allows to adjust to the levels of demand present along the busways. Frequencies should be high to reduce waiting times, and a good design will also help to reduce the number of passenger transfers. Furthermore, the route maps must be easy to understand for users [6]. The busway and feeder routes must be physically integrated, forming a network. Besides, the entry of other public transport operators must be restricted ([1]).

The BRT systems must have a distinctive **brand** image from other transportation systems. A good marketing strategy can position the brand and improve its image to attract more users. The BRT system should have a positive brand image.

### III. **TransMilenio BRT system**

TransMilenio is Bogotá’s most important mass transportation system and one of the biggest BRT systems in the world. It is based on the Curitiba BRT system, and there are new plans for its expansion, due to its success. Similar systems have been constructed in other Colombian cities. Currently, the system has completed its third construction phase out of a total of eight.

A clear definition of TransMilenio is given in [7]: “TransMilenio is defined as an urban mass transportation system that privately operates high capacity articulated buses that circulate through segregated busways, which are integrated into a system of feeder services that cover circular peripheral services with medium capacity buses. The system has stations with platform level boarding and automatic doors synchronized with the buses, where passengers take or get off the buses and the service is limited for those who have bus tickets. A satellite control system permanently supervises the buses, and the one-payment fee allows the passenger to board both busway and feeder services”.

#### A. Infrastructure

TransMilenio’s infrastructure is composed of three fundamental elements without which operation would be impossible: busways, stations and buses.

TransMilenio buses circulate on exclusive roads called **busways**. Currently the system has 11 busways with a total length of 104.6km ([8]). There are two types of busways: one lane and two lanes busways. The one-lane busways have passing lanes at stations so that buses can pass each other, thereby providing for express routes. The busways are located on the city’s main roads and are physically separated from the mixed traffic lanes. The busways are in the central lanes of the roads [8], [7]. Internal rules of circulation and operation control can be imposed to improve system performance because the lanes are used exclusively by buses.

As a complement for the buses that circulate on busways, there are lower capacity buses that circulate on the other roads of the city. These routes are called **feeder routes** and have predefined stop points ([9]).

TransMilenio has a total of 143 **stations**. These stations form the area where the users can move and board bus routes (the “paid area”). The station platforms are at the same height as the bus doors, and that makes it easy to board the buses ([8], [7]). There are three types of stations: portal, intermediate and standard. The **portal stations** are the main stations of the system and are located at the ends of each busway. They are the starting and final destination points for the buses. Furthermore, they have access to feeder routes, which depart from and arrive at these stations like the busway routes. In this way, the transfer between routes becomes easier. The **intermediate stations** are similar to the portal stations (passengers can transfer between busway and feeder routes) but are located at intermediate points of the busways and not at the ends ([8]). The **standard stations** are smaller than the other two and only allow access for busway routes. They are located along the busways with an average separation of 500 meters. Their size varies and they can serve 1, 2 or 3 buses simultaneously in each direction.

TransMilenio has three types of **busway services**: normal, express and super-express. The **normal services** are routes that stop in all stations along the way. They have a higher frequency than the other types of services. The **express services** stop only at some stations (from 40% and 60%), and have a higher average speed than normal services. The stop plan of these services has been designed according to the levels of demand of the stations along the busway ([9]). The **super-express services** are very similar to the express services. The only difference is that they stop at fewer stations (about 20%) of the stations along the busway. These services are better for users that have long journeys because they stop at few stations ([9]). Currently, TransMilenio has 1392 busway buses (articulated or bi-articulated) and 574 feeder buses.
B. Speed

System buses operate at average speeds of 19 and 32 km/h for normal and express services, respectively ([9]).

TransMilenio increased the average speed of the city’s corridors. For instance, the Caracas corridor had speeds of 12 km/h and Calle 80 of 18 km/h due to the oversupply of private bus operators that there was before TransMilenio and which generated traffic jams.

TransMilenio system has an average speed of 26 km/h ([9], [1], [3]). This means that the average speed of public transport increased by 15 km/h after TransMilenio was implemented.

Nevertheless, the speeds are not the same in all busways for different reasons, such as the number of traffic lights along the busways, the number of lanes and even the type of material the street is made of ([10]). For example, the Eje Ambiental busway, is a cobbled road in the historic center of the city. Table II shows the average bus speeds on the major busways of TransMilenio.

<table>
<thead>
<tr>
<th>Busway</th>
<th>Speed (km/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eje Ambiental (EW)</td>
<td>9.07</td>
</tr>
<tr>
<td>Eje Ambiental (WE)</td>
<td>10.51</td>
</tr>
<tr>
<td>Caracas (SN)</td>
<td>22.05</td>
</tr>
<tr>
<td>Caracas (NS)</td>
<td>22.61</td>
</tr>
<tr>
<td>Caracas Sur (NS)</td>
<td>24.82</td>
</tr>
<tr>
<td>Suba (SN)</td>
<td>24.95</td>
</tr>
<tr>
<td>Suba (NS)</td>
<td>25.08</td>
</tr>
<tr>
<td>Calle 80 (WE)</td>
<td>26.59</td>
</tr>
<tr>
<td>NQS Sur (NS)</td>
<td>27.32</td>
</tr>
<tr>
<td>Caracas Sur (SN)</td>
<td>28.17</td>
</tr>
<tr>
<td>Americas (EW)</td>
<td>28.24</td>
</tr>
<tr>
<td>Americas (WE)</td>
<td>28.37</td>
</tr>
<tr>
<td>Calle 80 (EW)</td>
<td>29.27</td>
</tr>
<tr>
<td>Autonorte (NS)</td>
<td>31.21</td>
</tr>
<tr>
<td>NQS Central (NS)</td>
<td>32.80</td>
</tr>
<tr>
<td>Autonorte (SN)</td>
<td>33.12</td>
</tr>
<tr>
<td>NQS Sur (EW)</td>
<td>33.18</td>
</tr>
<tr>
<td>NQS Central (SN)</td>
<td>36.87</td>
</tr>
</tbody>
</table>

TABLE II. AVERAGE BUSWAY SPEEDS IN TRANSMILENIO

C. Capacity

TransMilenio has a maximum load capacity of 45000 passengers per hour per direction (pphpd), the highest-capacity BRT system in the world and even outperforming many heavy rail or metro systems ([1], [10]), see Table III.

The transit capacity and quality of service manual ([11]) defines the capacity of any route or public transport corridor as “the maximum number of people that can be carried past a given location during a given time period under specified operating conditions without unreasonable delay, hazard, or restriction, and with reasonable certainty”. This capacity is measured in number of passengers per hour.

A system’s capacity is limited by the component with least capacity (i.e., the bottle neck). The three key components of the BRT systems are buses, whose capacity is measured in number of passengers; stations, whose capacity is measured in number of passengers and buses; and busways, whose capacity is measured in number of buses. Whichever of these three components has the least capacity will become the factor that controls the system corridor. Several authors agree that BRT systems capacity is most often limited by the stations ([9], [10], [1]).

As mentioned before, TransMilenio has two types of buses that operate on the busways, articulated buses and bi-articulated buses. Table IV shows the number of passengers that each type of vehicle can carry in a one-lane busway and with one boarding platform stations, the average time that a vehicle occupies a given boarding platform (dwell time) and the average boarding time. TransMilenio increases the system capacity by using multiple boarding platforms in each station ([1]).

Articulated buses carried an average of 1596 passengers in 2006, which is five times the average number of passengers carried by traditional buses. Furthermore, the number of kilometers a bus travels has increased due to the expansion of the busways, the extension of operating hours and the increase of express services. Each bus travelled 370 km daily in 2006 ([9]).

TransMilenio was the first BRT system to include multiple boarding platforms inside each station. In this way, it reached levels of capacity that only heavy rail systems had ([1]). Some TransMilenio stations may have up to five different platforms, each used for a different route.

There are reasons for including multiple platforms in a station [1]. The first one is to offer different types of services, such as normal and express, which can be allocated to different platforms. The second, and most important, is to reduce the saturation levels at stations, which helps to improve the service.

Besides, it is possible to distribute the different routes along each platform in such a way that each route stops only at one platform. It is then easier for users to find routes, because the user will associate each bus route with a platform.

In theory, one station with five platforms may have five times the capacity of a station with only one platform ([1]). To make this possible, the platform saturation level should be between 40% and 60%.
A TransMilenio capacity study was conducted in 2007 ([10]) and revealed which capacity values could be achieved according to the number of boarding platforms at each station, see Table V. Note that it is assumed that each platform has space to keep a vehicle in line (storage space).

<table>
<thead>
<tr>
<th>Type of station</th>
<th>Recommended saturation (%)</th>
<th>Capacity (vehicles/hour)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Station with one boarding platform and no storage space</td>
<td>40</td>
<td>48</td>
</tr>
<tr>
<td>Station with one boarding platform with storage space</td>
<td>60</td>
<td>72</td>
</tr>
<tr>
<td>Station with two boarding platforms and no storage space</td>
<td>40 and 40</td>
<td>96=48+48</td>
</tr>
<tr>
<td>Station with two boarding platforms, one with storage space and the other with no storage space</td>
<td>40 and 60</td>
<td>120=48+72</td>
</tr>
<tr>
<td>Station with two boarding platforms with storage space</td>
<td>60 and 60</td>
<td>144=72+72</td>
</tr>
<tr>
<td>Station with three boarding platforms where just one of them has storage space</td>
<td>40, 60 and 60</td>
<td>192=48+72+72</td>
</tr>
<tr>
<td>Station with three boarding platforms with storage space in each</td>
<td>60, 60 and 60</td>
<td>216=72+72+72</td>
</tr>
<tr>
<td>Station with four boarding platforms with storage space in each</td>
<td>60, 60, 60 and 60</td>
<td>288=72+72+72+72</td>
</tr>
</tbody>
</table>

TABLE V. STATIONS CAPACITY ACCORDING TO THE NUMBER OF PLATFORMS ([10])

In BRT systems, the busway capacity is much higher than the station capacity. Bogotá’s City Council transit and transport administration manual ([12]) states that the saturation flow of the busways is reached when there are between 692 and 750 articulated buses per lane. The interval is between 470 and 730 for right turns and between 465 and 735 for left turns. Note that this capacity does not contemplate elements such as intersections or traffic lights. It is clear that the busways capacity is much greater than the stations capacity.

The basic capacity of each busway is equal to the least capacity station along the busway ([10]). It does not account for questions that may increase system performance, such as express routes. The values for the different busways are (Calle 26 and Carrera décima busways are not considered since they were opened after the date of the study) ([10]):

- Caracas Centro Busway: 192 buses/hour.
- Autopista Norte Busway: 144 buses/hour.
- Avenida Suba Busway: 144 buses/hour.
- Calle 80 Busway: 48 buses/hour.
- NQS Busway: 72 buses/hour.
- Américas/Calle 13 Busway: 144 buses/hour.
- Eje Ambiental Busway: 72 buses/hour.
- Caracas Sur ramal Tunal Busway: 72 buses/hour.

IV. EXISTING STUDIES

There are very few proposals in the literature that focus on optimizing the BRT system operation. This can be explained because BRT systems are relatively recent (until the year 2000 there were only 19 BRT systems in the world ([4])).

Possible reason could be that many of the currently operating BRT systems are far from reaching their maximum capacity. For instance, none of the BRT systems operating in USA has reached maximum capacity and all of them have plenty of space for expanding their operation capacity ([6]).

Nevertheless, there are very interesting proposals that can be used as a starting point to propose a model for optimizing the operation of TransMilenio. To the best of our knowledge, there are no automatic route design proposals. The closest to this is the model proposed in [2]) that evaluates and selects the best several routes.

In the following we review of different proposals in the literature to manage and optimize the operation of different BRT systems and, specifically, for TransMilenio.

A. Proposals for optimizing of BRT systems

Most of the proposals in the literature for optimizing BRT systems are based on bus scheduling and are focused on varying the times between each bus departure (i.e., the headway) of the different bus routes.

In [13] a model for optimizing BRT systems is proposed on the basis of two elements, the headway, which is assumed to be uniform, and the order in which the bus routes depart.

The optimization model is characterized by a set of predefined bus routes (normal, zone and express routes). A random number is generated and assigned to the headway. Then, the algorithm finds an optimal solution to the order in which the routes have to depart that minimizes a cost function. For instance, the algorithm may determine that for a headway of 5 minutes the best departure order is [normal, express, express, normal, zone]. This solution means that a normal bus route should depart at minute 0, an express route at minute 5, a normal route at minute 10 and an express route at minute 15 and a zone route at minute 20. This solution may be better than for example [express, express, zone, express].

The cost function accounts for the passengers waiting at the stations, the waiting time inside the buses and the operating costs. It is very complete and includes several variables, such as the number of boarding/alighting passengers by station, the stops of each route, the monetary value of the waiting time costs and the vehicle operating costs, among others.

The model chooses a headway for the given routes and shuffles the order of departure. A genetic algorithm is used to reach an optimal solution. The article presents a novel codification that is a vital element for the model and includes a combination of the headway and route design variables.

In [14] a very complete model is proposed, with good granularity and with greatly detailed costs. An application to Line 2 of the BRT in Beijing is used to illustrate the model. The total cost of a solution accounts again for passenger waiting time at the station, passenger waiting time inside the bus, and the vehicle operating costs. Passenger walking time from home/office to the station is excluded, because bus scheduling has no influence on that time.

The model considers variables such as the bus departure frequency, the distance between stations, average speed between stations, the rates and boarding times, acceleration
and deceleration times, the number of traffic lights between stations, the traffic lights cycle times and others. Furthermore, it assumes that the waiting time is equal to half of the frequency time or headway.

Fixed costs are removed from the analysis because they are unaffected by bus scheduling. The variable costs are composed of operating cost per kilometer, operating hours, vehicle depreciation, etc.

The decision variables in the optimization model are the route headway and binary variables that represent whether or not stations are skipped. The model seeks to minimize the total costs and is subject to capacity constraints, vehicle availability and headway limitations. To accomplish this an all-stop route and an express route is combined and their headways are calibrated to minimize the total costs.

The algorithm complexity increases exponentially along with the number of stations, and this is the main reason why the authors use a genetic algorithm (it would be too costly to use a deterministic algorithm). Another reason is that genetic algorithms are able to naturally represent binary variables.

[2], [15], [16] introduce an optimization model for the minimization of waiting time, travelling time and operating costs for an express bus service, given the travel demand. A mathematical model is built to minimize costs given a set of stations, the distance between stations, the passenger origin-destination matrix and a set of a priori attractive set of routes. For each suggested route the model outputs the frequency of the services and the size of the buses to use.

For the construction of this model the travel demand is assumed to be fixed and known, represented by a origin-destination matrix for the analyzed stations, which must be satisfied. It is also assumed that passengers arrive at an average fixed rate, passengers choose the route to their destination that minimizes travel time and there is no limit on the available vehicle fleet.

The operating costs are computed on the basis of the cycle cost of a full bus route, the frequency of each route and the operating set of routes. The passengers costs are given by the waiting time at stations, travel time and transfer time.

There are also various proposals to improve the operation of BRT systems through the prioritization of transit signals. [17] describes the mathematical relationship between the departure frequency of a route, the cycle length of the transit signals and the number of different signal states when the buses arrive at an intersection. It proposes various strategies for prioritizing signals that decrease the headway time deviation, i.e., decrease the punctuality deviation of the buses, without having a significant impact in the delay of the mixed traffic.

Other proposals that study the priority control of signals in BRT systems can be found in [18], [19], [20], [21].

B. Proposals for the optimization of TransMilenio

Since the construction of TransMilenio in the year 2000, various proposals have been made to optimize its operation. Most of these proposal focus on the reduction of systems costs, to strike a balance between passengers waiting time and operating costs.

For instance, Petri networks are used to model TransMilenio in [22]. The proposed model is classified as a macroscopic deterministic simulation model, due to its detail level, process and operation representation. The model uses a multiagent approximation to model three important system components: the passenger behavior (how many passengers take the bus per hour), the busway dynamics, and the interaction between the passengers and the buses. Since Petri networks are unable to deal with time, trigger times are added to the nodes to represent temporal relationships.

Three busways are modeled, Avenida Caracas, Autonorte and Avenida de las Américas. The model includes the seven most important stations out of a total of 45 on these busways. Moreover, three routes (a normal route, an express route and a super-express route) that stop at the same stations on their back and forth trips were chosen.

Two Petri networks are designed. One models the whole system and randomly assigning buses to routes, and the other separates the routes from the buses. Random models have the advantage of being able to simulate the interaction between routes. The random model outperforms non-random models to satisfy the demand with the minimum number of resources. Finally, the random model works as an integrated system and it is capable of solving perturbation by itself. The result of this simulation shows that there is a point at which adding more buses to the system does not improve the performance.

In [23] a genetic algorithm is used to find the best frequency for pre-established bus routes that minimizes passenger waiting time. The frequency is determined by the assignment of buses to each route. The model tries to minimize the time the passengers spend on the system, which is composed of the travel time plus the waiting time at the stations.

The genetic algorithm chromosome size is equal to the number of routes and the population is initialized randomly with the constraint that each route has at least one bus assigned. A random matrix is also created along with the initial population, this matrix contains all origin-destination trips. This algorithm assumes the user is “smart” and will always choose the best route to go to his/her destination.

The arrival of buses at the station and the passengers waiting time are modeled by a Poisson process and a distributed Erlang event, respectively. It accounts for the scenario where buses are full and passenger cannot board. In these cases the passengers have to wait to the next bus.

A graph with the routes was designed to measure the time, where each node represents a station and the arcs represent the connections between them. The arc costs are the travel time between the stations that the arcs connect. Additionally, arcs with the possible express routes are included. Dijkstra’s algorithm is then used to compute the shortest routes, and it is executed before running the algorithm.

In [24] a model to evaluate TransMilenio routes is built based on the data provided by a origin-destination matrix. The trip probabilities between stations and passenger arrival rate to each station are computed from this matrix, assuming that the users know which is the best route to reach their destination. The model is implemented in a commercial simulation software package.
The model includes a set of constraints regarding user behavior when choosing their route to reach to their destination. The input data for the evaluation algorithm are the origin-destination pairs, the stops of each service and the quantity of passengers associated with each pair. TransMilenio data is also required, such as existing routes, their frequency, vehicle capacity, speed, the distance between stations and other network characteristics.

The model is composed of three modules. The network module stores information about the physical infrastructure, such as stations, the distance between them, busways, and others; the stations module is in charge of the boarding and alighting at each station and the arrivals module assigns passenger origin and destination.

The time that the passenger spends inside the station is given by the travel time, the bus stopping time and passenger waiting time at the station. The bus arrival times are assumed to be uniform. Therefore, the waiting time of each passenger is equal to half of the route’s headway.

C. Analysis of proposed models

Most of proposed models are far from being able to represent what goes on in the real world because they are not detailed enough to represent what happens within a BRT system. We are going to describe the advantages and drawbacks of each model.

The model presented in [23] does not account for vehicle operating costs. This fact is clearly reflected in the results, where the best solution is to increase bus departure frequencies and use the entire bus fleet. The models do not adequately represent constraints concerning capacities within the system, which is modeled for the whole corridor but not for each station individually. This overlooks the fact that there are some stations that have more demand than others. The model does not account for deceleration and acceleration times, passengers boarding and alighting times or dwell times at signalized intersections. An advantage of this proposal is that it builds a graph that pre-calculates the travel times between each pair of stations for each route. This is helpful to find the optimal route between two stations.

The model presented in [22] has several voids, such as the fact that it does not account for passenger waiting time or vehicle operating costs. Neither does it account for passenger congestion within the stations, vehicle congestion at stations, discriminated speeds between each pair of stations, or assign distribution times to passenger and vehicle arrivals at stations. On the other hand, it has several advantages, such as considering that when buses are full passengers must wait for the next bus. The model is a user-friendly graphic tool that can model a system in which equations are not known.

The model presented in [24] refers to some important constraints but the model does not include any. Other constraints included as assumptions are not necessarily realistic. For instance, it is assumed that if a passenger is going to make a trip that is 5 or less stations of long, he/she will only take normal (all-stops) services. It does not consider decelerating and accelerating times, passenger boarding and alighting times or dwell times at signalized intersections either.

The model presented in [13] assumes the same speed between every pair of stations, which is not realistic. It does not consider the bus passenger capacity, vehicle capacity at the stations and passenger capacity at the stations either. Waiting times at signaled intersections are not considered either. A major drawback of this model is that it uses the number of passengers that board and alight from buses at each station rather than an origin-destination matrix as input. This demand data is not detailed enough to identify passenger behavior. The model accounts for passenger waiting times and vehicle operating costs. The introduction of an innovative variable-size codification and the use of binary variables to indicate whether or not a bus stops at a station.

The model presented in [14] is one of the most complete. In fact, it incorporates most cost variables. It is the only model that includes passenger boarding and alighting times and the stop times at the signaled intersections. Nevertheless, it has some drawbacks. For example, it only considers one express route, i.e., scenarios with several express routes cannot be evaluated. The model is aimed at reaching the best departure frequency for a normal and an express route that operate along the same busway. Its parameters are the origin-destination matrix, the stations in which the express routes stops and the bus fleet size. Note, finally that the model does not consider vehicle congestion or passenger congestion at stations.

The model presented in [2] is also very complete and perhaps the best at representing the costs of a real BRT system. This is accomplished thanks to the inclusion of several express routes on one busway and because it is good at differentiating travel time and operating costs. But it is not free of drawbacks. For instance, it does not include acceleration and deceleration times, boarding and alighting times or stop times at signalized intersections. The proposed model searches the departure frequencies that optimize BRT operation according to a defined cost function.

In summary, none of the reviewed proposals considers vehicle congestion at stations or passenger congestion at stations. This is worrying, because, as stated in [9], [10], [1] the capacity bottleneck of a BRT system is the vehicle capacity at the stations. There are not many proposals that account for this point because hardly any BRT systems have reached maximum capacity, which could be the reason why the proposals have focused mainly on the minimization of passenger waiting times an operating costs, and not on the increase of system capacity.

We found that none of the proposals offer automated route design. [14] and [2], which offer validation models for routes that can be given to the model as a parameter, come the closest. We also found that none of the proposals take a multi-objective approach to the problem.

V. PROPOSED OPTIMIZATION MODEL FOR TRANSMILENIO

In the previous section we reviewed the proposals in the literature for optimizing BRT system operation, with a special emphasis on the TransMilenio system. In this section, we provide a mathematical model for the optimization of TransMilenio that adapts elements of the above proposals, mainly [14] and [2], and incorporates novel elements accounting for the features of that system.
The problem is to find departure frequencies for the established routes that minimize the time passengers spend inside the system and operating costs. This set of frequencies must satisfy the constraints associated with the TransMilenio operation. The problem is analyzed only during the rush hour time window.

A. Available information

The set of stations determines the size of the BRT system. The number of stations is directly related to the complexity of the problem to be solved. Information about the system stations, the busways to which they belong, and each station’s neighboring stations must be considered. The set of TransMilenio stations is denoted by \( E = \{ e_1, \ldots, e_{143} \} \), where \( e_i \) refers to the \( i \)-th station, \( i = 1, \ldots, 143 \).

The routes are paths between two stations (usually main stations) that buses must take and are composed of the set of station at which buses must stop. The set of TransMilenio routes is denoted by \( R = \{ r_1, \ldots, r_{90} \} \), where \( r_j \) is the \( j \)-th route, \( j = 1, \ldots, 90 \).

The station vehicle capacity is very important, even more so in cases where nearing full capacity the BRT system is, like the TransMilenio system. When the system is nearly at maximum capacity the problem is to find feasible solutions that can meet the trip demands. The station vehicle capacity \( e_i \) is denoted by \( k_e^i \), \( i = 1, \ldots, 143 \).

The information required about the buses is their capacity and the quantity of buses in operation. We assume that the buses operating along each route have the same capacity. This information is important in order to impose capacity constraints within the buses and to prevent to operate with more buses than the available. The bus capacity is denoted by \( k_b^j \), \( j = 1, \ldots, 90 \) which is the passenger capacity of the vehicles that operate the \( r_j \)-th route.

The distance between stations is used to compute the travel times between each pair of stations. \( d_{e_i e_j} \) denotes the distance between stations \( e_i \) and \( e_j \), \( i, j = 1, \ldots, 90 \).

Speeds between each pair of stations are very important because not all busways have the same characteristics and therefore the speed is not always the same. Some busways have signaled intersections, whereas others are built over highways where they can travel at faster speeds. The speed between the stations \( e_i \) and \( e_j \) is denoted by \( s_{e_i e_j} \).

Acceleration and deceleration times along with boarding and alighting times are used to compute the total time of a stop at a station. These values are constant and independent of passenger demand level in the system. Based on the model proposed in [14], we assume that the times are the same for all stations. The acceleration and deceleration times are denoted by \( p \).

The boarding and alighting times are used to determine the total stop time of a bus at a station. The stop time increases with the amount of people that board or alight the bus. Based on the model proposed in [14], we can calculate passenger boarding time at a station as \( \alpha_{r_j} \times \tau^a \), where \( \alpha_{r_j} \) is the passenger boarding rate for route \( r_j \) at station \( e_i \). \( \tau^a \) is the passenger boarding time. In the same way, the alighting time is denoted by \( \beta_{e_i e_j} \times \tau^b \), where \( \beta \) is used for alightings.

Costs are usually divided into passenger waiting time costs and the BRT system operating costs. The model that we propose accounts for three types of costs: waiting time at stations, waiting time on buses and vehicle operating costs. Fixed costs, such as station cleaning, electricity, administrative wages, rents, and others, are not considered because they are independent of the BRT system operation ([14]). The unit cost per kilometer, the unit cost for waiting time at the station and the unit cost for waiting time inside the buses, are denoted by \( \mu_O, \mu_S \) and \( \mu_B \), respectively. These values are used in the cost function to evaluate the quality of the sets of routes.

The origin-destination matrix contains information about passenger demand, i.e., the amount of users traveling from station \( e_i \) to station \( e_j \). We use an origin-destination matrix with rush hour data, because this is the time window when the system is closer to maximum capacity. The number of passengers that travel from station \( e_i \) to station \( e_j \) is denoted by \( q_{e_i e_j} \).

Operating hours is the time during which the BRT system is operating, denoted by \( T \).

B. Decision variables

The decision variables for the proposed model are the frequencies associated with each route. The set of frequencies is denoted by \( F = \{ f_1, \ldots, f_{90} \} \), where \( f_k \) is the frequency for the buses of the \( k \)-th route. The frequencies identify how often the buses of a given route depart. The headways can be computed from the frequencies and vice-versa.

C. Cost function

Multiple authors (see [14], [2], [13]) agree that the cost function, \( C \), is composed of the sum of three elements: vehicles operating costs, \( C_O \); passenger waiting time at station costs, \( C_S \), and passenger travelling time costs, \( C_B \). These last two costs can be grouped as the passenger total trip costs ([14]).

Then, the function to be optimized (minimized) is:

\[
\min C = C_O + C_S + C_B,
\]

The operating costs can be calculated by:

\[
C_O = \mu_O \times \sum_{r_k \in R} f_k \times D_k,
\]

where \( \mu_O \) is the unit cost per kilometre for a BRT vehicle, \( R \) is the set of all routes in the system, \( T \) is the BRT system operating hours, \( f_k \) is the frequency of route \( r_k \) and \( D_k \) is the length of the path covered by route \( r_k \). \( D_k \) can be computed from the distances \( (d_{e_i e_j}) \) between the consecutive stations included in the \( k \)-th route.

The waiting time at station costs can be computed as follows:

\[
C_S = \mu_S \times \sum_{e_i e_j \in E} q_{e_i e_j} \times \sum_{r_k \in R} f_k \times x_{e_i e_j}^{r_k},
\]

where \( \mu_S \) is the waiting time unit cost, \( q_{e_i e_j} \) is the passenger trip demand for the \( (e_i e_j) \) origin-destination pair, \( f_k \) is the frequency of route \( r_k \), \( x_{e_i e_j}^{r_k} \) is a binary variable that indicates
whether a route \( r_k \) is a good option for travelling from the station \( e_i \) to the station \( e_j \) (its value is 1 if the route is attractive and 0 otherwise), and \( \epsilon \) is the bus arrival distribution at the stations, which are assumed Poisson distributions.

The travel time costs can be computed by:

\[
C_B = \mu_B \times \sum_{e_i, e_j \in E} q_{e_i, e_j} \times \sum_{r_k \in R} \frac{t_{1r_k}^{e_i, e_j} \times f_{r_k} \times x_{r_k}^{e_i, e_j}}{r_k \in R} \frac{t_{2r_k}^{e_i, e_j}}{r_k \in R} \times x_{r_k}^{e_i, e_j},
\]

where \( \mu_B \) is the travel time unit cost, \( f_{r_k} \) is the frequency of route \( r_k \), and \( t_{1r_k}^{e_i, e_j} \) and \( t_{2r_k}^{e_i, e_j} \) are the travel time and delay time from station \( e_i \) to station \( e_j \), respectively.

\[
\begin{align*}
\frac{t_{1r_k}^{e_i, e_j}}{f_{r_k}} &= \frac{d_{e_i, e_j}}{s_{e_i, e_j}}, \\
\frac{t_{2r_k}^{e_i, e_j}}{f_{r_k}} &= \left( \frac{\sum_{e_i \in F_{r_k}} \alpha_{e_i}^{r_k} \times \tau^{\alpha} + \sum_{e_i \in P_{r_k}} \beta_{e_i}^{r_k} \times \tau^{\beta}}{f_{r_k}} \right) + N_{e_i e_j} \times d,
\end{align*}
\]

and

- \( d_{e_i, e_j} \) and \( s_{e_i, e_j} \) are the distance and average speed between station \( e_i \) to station \( e_j \), respectively,
- \( \alpha_{e_i}^{r_k} \) and \( \beta_{e_i}^{r_k} \) are the boarding and alighting rates at station \( e_i \) for route \( r_k \), respectively,
- \( \tau^{\alpha} \) and \( \tau^{\beta} \) are the boarding and alighting times per passenger,
- \( N_{e_i e_j} \) is the number of stations between station \( e_i \) and station \( e_j \),
- \( P_{r_k} \) is the set of stations at which route \( r_k \) stops,
- \( p \) is the acceleration and deceleration delay at station.

### D. Constraints

The bus passenger capacity constraint ensures that the frequency of bus departure is high enough to prevent overcrowding inside the buses. If this constraint is not applied, buses may be full when the arrive at stations, passengers will have to wait for the next bus.

\[
k_{r_k}^b \times f_{r_k} \geq \sum_{e_i \in P_{r_k}} \sum_{e_j \in P_{r_k}} q_{e_i, e_j} \times \sum_{r_m \in R} f_{r_m} \times x_{r_m}^{e_i, e_j} \\
\forall r_k \in R, \forall a \in P_{r_k},
\]

where

- \( k_{r_k}^b \) is the passenger capacity of the buses circulating along route \( r_k \),
- \( f_{r_k} \) is the frequency of route \( r_k \),
- \( P_{r_k} \) is the set of stations at which route \( r_k \) stops,
- \( q_{e_i, e_j} \) is the passenger trip demand for the \((e_i, e_j)\) origin-destination pair,
- \( x_{r_k}^{e_i, e_j} \) indicate whether a route \( r_j \) is a good option for travelling from the station \( e_i \) to the station \( e_j \). Its value is 1 if the route is attractive and 0 otherwise.

The bus fleet size constraint prevents the set of routes from operating with more buses than are available in the system. This assures that the system is working with the available resources:

\[
\frac{T}{N_{r_j}} \leq \frac{1}{f_{r_j}}, \quad \forall r_j \in R,
\]

where \( T \) is the BRT system operating hours and \( N_{r_j} \) is the number of vehicles that can operate along the route \( r_j \).

The choice of best route constraint helps to model passenger behavior when choosing a route to travel to their destination. It models the possibility of passengers often being able to take more than one route to reach their destination in the same time.

\[
x_{r_k}^{e_i, e_j} = 1 \quad \forall r_k \in R, \forall e_i, e_j \in E.
\]

The station vehicle capacity constraint is very important especially in systems that have great passenger demands and are nearing maximum capacity level. The importance of this constraint is that the vehicle station capacity is the bottleneck of the BRT systems ([9], [1]), like TransMilenio.

\[
k_{e_i}^s \geq \sum_{r_j \in R} v_{r_j}^e \times f_{r_j}, \quad \forall e_i \in E,
\]

where \( k_{e_i}^s \) is the vehicle capacity of station \( e_i \), and \( v_{r_j}^e \) are binary variables that point out whether station \( e_i \) is visited on the route \( r_j \).

### VI. Conclusions and Future Research

TransMilenio is Bogotá’s most important mass transportation system and one of the biggest BRT systems in the world. There are very few proposals in the literature that focus on optimizing BRT system operation, mainly because BRT systems are relatively a recent form of transport and many of the currently operating BRT systems are nowhere near full capacity.

Most of proposals, and specifically for TransMilenio, are based on bus scheduling and focus on varying the times between each bus departure (i.e., the frequencies) of the different bus routes to minimize costs.

In these proposals, the set of routes are part of the available information, along with the stations at which the buses stop, and they remain constant during the execution of the model.

In the mathematical modeling introduced in this paper we also analyze the frequencies of the routes to minimize costs.

Note that an automated design of routes that minimize the cost function is an open research line. Rather than designing new routes, the aim would be to optimize existing routes by...
modifying at most 30% of the stations on the original route. This 30% was fixed by TransMilenio experts at meetings. The reasons for just modifying rather than redesigning routes is that the social impact of modifying the routes is not too high, whereas, the search space is greatly reduced and, therefore, better solutions can be found in less time. An important drawback is that it may not be possible to find a global optimum, because the best routes may have less than 70% of the stations in common with the original routes. In this case it is more important to reduce the social impact on passengers that comes with the modification of the routes.

We are now working with TransMilenio experts on extending and solving the proposed optimization problem. We have selected evolutionary algorithms to solve the problem since they have previously proven to be efficient tools. Additionally, the research team is experience in solving other complex optimization problems using this metaheuristic.

The model we propose is a single objective optimization model since only costs are minimized. However, other objectives could be simultaneously considered, leading to a multi-objective optimization model. Evolutionary algorithms would be then used to identify Pareto optimal solutions, and the expert’s preferences could be incorporated into the search process to reach a compromise (satisficing) solution.

Another future research line that we propose is the possibility of adding transfer times to the model. Transfer time is the time it takes to a passenger to switch from one route to another, usually because the first bus that a passenger takes does not stop at the station for which he or she is heading. These times are normally penalized because transfers are an inconvenience for passengers.

Finally, another open research line, and a key aspect for correctly modeling BRT systems is user’s behavior. It is important to correctly model which routes users given an origin/destination pair will choose. They are likely to choose the fastest route, but this is not always the case, because users may not know which the fastest route is or because the frequency of the fastest route is low and they opt for an alternative route. This is one of the least explored issues in the BRT systems literature.

ACKNOWLEDGEMENT

The paper was supported by Madrid Regional Government project S-2009/ESP-1685 and the Spanish Ministry of Science and Innovation project MYTM2011-28983-C03-03. The authors would also like to acknowledge the Ministry of Information and Communication Technologies of Colombia and its Talento Digital scholarship program for funding the studies of Francisco J. Peña, and Alirio Garcia for the useful information he provided about TransMilenio system.

REFERENCES

