The City of Edmonton

Transit Service Delivery Models

Transit Strategy Guiding Perspectives Report

Sustainable Development;Transportation Strategies
6/24/2016

TABLE OF CONTENTS

1. INTRODUCTION

Public transportation, like most services, is far more complicated than it might first appear, with many potential goals competing for limited resources and just as many types of service that can be provided. Even among those who work for transit agencies, few are fully aware of the trade-offs that must be made between conflicting goals or the reasoning behind them. This paper is intended only as an overview of the potential objectives, service types and operation of transit systems and how they are interconnected.

Before understanding how a system works, it is important to understand what the system is supposed to do. For example, an intricately connected mass of tiny gears and springs would appear incomprehensible without first knowing that they are part of a portable device meant to keep precise time. In addition, it is also important to understand the requirements and limitations imposed on the system before passing judgements on how well it is functioning. For example, if the watch is off by several minutes, it may appear to be poorly made unless the watch was made before more precise timekeeping was possible.

Part one of this paper will introduce a few key terms and concepts; part two will discuss the tradeoffs which must be made regarding the purpose and type of service provided; part three will provide an overview of how the different types of service lines and modes work together to form a synchronized network; and part four will provide a brief description of the Edmonton Transit System.

1.1. THE FUNDAMENTAL TRADEOFF: ACCESS VS. MOBILITY

The type of transit service provided is the result of many tradeoffs. At first, it might appear to simply move riders between locations within the city, which poses no obvious conflict. However, funding is limited and choices must be made as to how to allocate funding. Even in the example of moving riders, there are several choices such as: which riders should the system move around?, and what locations should they be able to travel between? In the process of making these choices, others questions will come up such as: what times should service run?, and who should pay for the service?. Each of these will then lead to more questions and so the complexity quickly becomes apparent.

Many of these choices stem from a conflict that is fundamental to all transportation systems, the opposing goals of "Access" and "Mobility". In this context, "Access" describes how easy travelers can get to and use the service while "Mobility" describes the ability of the service to move passengers from one location to another. While both access and mobility are always required, increasing one tends to decrease the other. A common example is bus stop spacing where closely spaced stops improve access by reducing walking distance, but frequent stops reduce mobility by increasing travel times. Several other examples are shown in Table 1. [1]

TABLE 1: ACCESS AND MOBILITY

[1]	**Ways to improve access**	**Ways to improve mobility**
Primary Focus:	Focus on geographic and temporal availability and accommodating special needs	Focus on increasing travel capacity, speed and improving reliability
Stop Spacing:	Closer stop spacing to reduce walking distance	Increased stop spacing to reduce travel time
Roads serviced:	Service along smaller roads to provide service closer to homes and businesses	Service along arterials dedicated ROW where travel speed is higher
Local Service:	Service to all areas, regardless of demand to maximize available connections	Service only where warranted by demand

Long distance Connections:	All routes provide local service to all areas they travel through to improve connectivity of each stop	Express lines provide connections with few stops between distant locations
Length:	Longer routes with more destinations so that fewer transfers are required	Moderate length of routes requiring transfers to some areas but service is more reliable.
Permanence:	Routes are frequently revised to meet changes in demand and the road network	Routes are permanent to allow for upgraded rights of way and upgraded stops
Frequency:	High frequency service regardless of ridership to minimize wait time	Reduce frequency to improve reliability, unless required by high demand or transfer timing
Service time:	Longer hours of service regardless of ridership to maximize availability	Service only when warranted by demand.
Payment:	Multiple options including cash during boarding, as well as tickets and passes	Off-vehicle payment only to minimize time spent boarding and alighting
ROW upgrades:	Improve stop amenities to improve comfort and safety of waiting passengers	Separated ROW to reduce conflicts with other travel modes and improve performance
Vehicle upgrades:	Accessibility features such as kneeling, ramps, wheel chair accommodations and additional seating	Fewer seats thus allowing for higher loads and potentially faster boarding and alighting.

1.2. A FEW KEY TERMS

It is important to clarify a few common terms that are used in a way that might seem unusual. To avoid confusion, throughout this paper the following conventions will be used:

Arterials, collectors and local roads: As many forms of transit share the roads with other traffic the type of road is occasionally mentioned. Due to the different purposes of roads, they tend to follow a hierarchy with three broad types. Arterial roads are intended to move a large number of vehicles quickly over long distances, and are usually linear with many lanes, high travel speeds and limited direct access to homes or businesses. In contrast, local roads provide direct access to homes in neighborhoods and are smaller, have slower speed limits and often use curves and speed bumps to prevent speeding. Between these are collector roads which are intended for access to nearby destinations and to provide connections between all three road types. Collectors are usually larger than local roads and smaller than arterials and can be either linear or curving.

Lines & Routes: Typical convention is to describe rail service as a “line” and bus service as a “route”. This can cause confusion when discussing both rail and bus service, and the term “route” can also be used to describe the travel path. Throughout this paper both rail and bus service will be described as “lines” and the term “route” will be used only to describe the pathway or alignment.

Stops, Stations, Transit Centres & Terminals: Buses and trains use a range of different facilities where passengers can board and alight, which can be described in a variety of ways. The term “stop” is used broadly to indicate any designated location where passengers can board transit vehicles. The term “station” is used to indicate upgraded stops with amenities, which includes most stops along rail lines. Hubs and Transit Centres are stops that are designed to accommodate many lines at once for efficient transfers between lines. Finally, the term “terminal” will only be used to describe the end point of a line. While not required, terminals are often either transit centers or stations

Trains, Buses, Vehicle and Transit Units: Many of the topics discussed in this paper are applicable to both bus and rail service, and so the terms "Vehicle" or "Transit Unit" (TU) will be used as a general term, while "bus" or "train" will be used when discussing topics only relevant for that mode.

Trips & Travelers: The most basic units of transportation are "trips" and "travelers" which are both commonly understood terms. The term "trip" is used to denote a single journey from a specific starting location to a specific destination. For example, if a person travels from home to work and back, this is two trips, the first from home to work and the second from work to home. This is most often described as either a journey made by a single passenger (passenger-trip) or by a specific vehicle (vehicle-trips). When describing transit vehicles, the entire line is typically traversed on a single vehicle-trip. The number of passengers (also referred to as "travelers", "riders", "users" and "customers") is typically described in terms of the number of the number of passenger-trips made over a specific period of time. [2]

1.3. Efficiency, Performance & Level of Service

The terms "Efficiency", "Performance", and "Level of Service" (LOS) will be used throughout this paper. Often these terms are used in relation to personal vehicles or facilities that are intended for personal vehicle transportation, such as fuel consumption, acceleration or vehicles per hour through intersections. Transit, however, operates very differently from personal vehicles and so it is important to have a better understanding of what is meant for a transit system to be considered "efficient" or provide a higher "Level of Service", and how any of this relates to performance. [3]

Personal vehicles, including taxis, only need to meet the needs of the individuals in the vehicle. As such, drivers are bound only by their chosen locations and are free to leave at any time, travel along any route, stop at any location along the way and may change any of these at any time. Transit must accommodate many users with a wide range of different needs, traveling between locations over a large geographic area. The operation of a train is a good analogy for how all transit lines must operate to accomplish this. Trains arrive and depart on a predictable, pre-determined schedule and travel along a pre-determined route, only stopping at specific pre-determined locations. Any change to the route or schedule is likely to frustrate travelers. In order to connect locations, the transit system must include many lines which connect to form a network. This network must be designed to accommodate high demand along the most popular connections as well provide connections to many other areas. [4]

Transit is more restricted in when, where and how it can operate but it is able to move many passengers using less space on roads and less fuel than automobiles. [2] This is because transit vehicles are larger and able to hold dozens or even hundreds of travelers with only a single driver required and so fewer vehicles are needed. When separated from other traffic, fewer large vehicles create much less congestion than many small ones and can travel faster through very busy areas. In addition, parking is not required at the destination and passengers are free to read, text, play games or even sleep while they travel. [5]

From a theoretical perspective, any attempt to measure either performance or efficiency should be closely tied to the intended purposes. [6] As will be discussed in part two, the intended purpose is not as clear or simple as might be expected. Many important goals are difficult to quantify because they are subjective (such as customer experience) or only being supported by transit (such as land use, social equality & economic impacts). As such, most items related to performance, efficiency and level of service are limited to operational and financial characteristics. [3]

1.3.1. Efficiency

Transit efficiency is measured as a ratio between the amount of service provided and the resources consumed to provide it. This is often considered in terms of the entire system due to limitations in data and the complex interaction between various lines and areas. [3] In addition, efficiency information is often used by higher levels of management who are typically most interested in the overall system. A variety of measures are commonly used, although most follow a common pattern where increased passengers on each vehicle and faster travel speeds tend to increase efficiency. [2] Another common relationship is that efficiency tends to decrease when multiple lines have overlapping or redundant service unless accompanied by additional demand. [7]

In the most general sense, the resource consumed is money and is often considered in terms of the service area population (i.e. $156 per capita) or revenue generated (i.e. 45% revenue recovery). Such cost measures are discussed in the Guiding Perspectives report on Transit Fares and Subsidies, and are primarily useful for high level decisions regarding fares and government subsidies. In order to assess the actual efficiency or cost effectiveness, cost should be compared to the quantity of service provided, resulting in measures such as the average cost per passenger (i.e. $3.18/Passenger-Trip), or cost per hour of service provided (i.e. $126.27/Vehicle-hour). [7]

Financial measures are of limited use to planners because they are impacted by many unrelated aspects of service. A wide range of operational measures of efficiency are often used by planners to better understand and evaluate efficiency. For example, average speed can provide insight into passenger boarding times and traffic congestion. Measures for utilization often reflect passengers per hour, ridership per capita and how much of the provided service is actually being used. Many other indicators can also be used such as fuel, energy, and CO_2 per passenger. [2]

1.3.2. Performance & Level of Service

The terms "Performance" and "Level of Service" (LOS) are used synonymously throughout this paper to describe the capabilities of the service being provided, with "higher" being superior to lower performance or LOS. In many cases, these terms are used in a generic sense; however, it is important to understand the various aspects that collectively define the performance.

Unlike efficiency, performance is usually considered on a smaller scale, such as for individual lines or areas, instead of the entire system as a whole. This is because performance information is typically used by planners to identify problems and adjust service to match demand on specific lines or in small areas. Overall system performance is often of limited usefulness due to the wide variation between both geographic locations and time periods. For example, an area with inadequate service may be masked by another with excessive service. [2] Also, the data can be weighted in several ways (i.e. they could be unweighted, or they may be weighted by ridership, number of trips or hours of service) and may include all time periods or only focus on a few. [7]

Transit performance can be considered and analyzed in a number of ways, and no single value can entirely describe it. [7] Instead, a variety of performance measures are used, many of which are related to one another in some manner. For example, a line's capacity is determined by the frequency and maximum load of each vehicle while travel speed is largely dependent on the mode (regular Bus, LRT or BRT) and type of service (local, express, shuttle). [2]

An important aspect of LOS for an area is connectivity, or how well it is connected to other areas. Typically, connectivity increases with the number of routes that serve a given area, as well as number of transfer locations and the number and type of lines serving each. [3] This is discussed in more detail in part three.

Level of service can also be considered in terms of customer experience which depends on both operating characterizes and the perception of each individual. Items related to customer experience include new factors such as safety, cleanliness, stop amenities, and walking distance waiting time, as well factors related to operating characteristics such as bus crowding and travel time. [2]

Reliability is an aspect of performance that is related to both operations and customer experience. For this discussion, reliability can best be thought of as providing the service customers are expecting. This is usually thought of as buses arriving and departing according to the official schedules, but also includes bus crowding and changes in service that customers are not expecting. This description has three parts that can be addressed together or independently: the service provided, the service customers expected, and how well the two match. [1] Changing the actual service provided must usually be done over several years and is often difficult or expensive to accomplish. Customer expectations stem from official communications and previous experiences, and it is important to clearly communicate schedules. Matching these two can often be done by adjusting customer expectations to match the service actually being provided. One way of accomplishing this is by updating the bus schedules in real-time using "Smart-Bus" technologies. Another way is to better understand the factors causing delays and overloads and adjust printed schedules to reflect them. [2] However, this is not always practical as it often requires expensive data collection and analysis or significant additions of time to schedules.

1.4. Demand

An appreciation of transit demand is critical to understand the purpose and operation of a transit system. The term "demand" refers to ridership, either actual ridership on current transit service or expected ridership on future service. Most goods and services have some inherent value to consumers; however, transportation is instead required in order to accomplish a task. As such, the attractions at a location will largely determine the demand patterns to and from that location, and demand along a given line depends on the locations it provides access to. [8]

1.4.1. Variable demand

While demand is often treated as a single, constant value such as "1000 riders per day", this is actually an average over a specific area and a range of time. In reality, demand varies widely both over time and between locations. Fortunately, while demand is constantly changing, it does tend to follow regular patterns and trends.

Time variable demand

Daily Travel Patterns

Travel volumes change dramatically over the course of a single day as shown in Figure 1. The most well-known example of this is the change in demand over the course of a typical weekday. Many workers commute to and from work on similar schedules, resulting in two dramatic "peaks" in travel demand (often referred to as "Peak periods"). The AM Peak is usually shorter and steeper than the PM return trip due to after-work activities. [2]

Annual Patterns

When analyzing demand over the course of a week, month or year, other patterns emerge as shown in Figure 2. Such patterns will depend on the unique character of each municipality, but often follow seasonal trends such as reduced demand during summer when school is out and many families take vacations. [9]

Long Term Trends

When analyzing demand over a period of years, demand often follows a trend similar to the one shown in Figure 3. This trend is typically driven by population growth and physical size of the municipality along with impacts from the strength and growth of the economy. [2]

Spatially Variable Demand

Equally significant is the spatial variation in demand which is based on nearby land use and intensity, population density and demographics as well as activity patterns. Just as these characteristics vary between locations in the city, so too will the demand pattern [2]. For example, downtown sees high travel volumes due to the density of employment and services offered, while remote neighborhoods tend to see low travel volumes due to the low intensity development with limited services and employment. Figure 4 shows a portion of an ETS demand map from 2011.

1.4.2. Contributing Factors

In order to predict demand patterns it is important to understand the factors causing them. Previous sections described demand as being driven by work schedules and land use. [8] While these are the primary drivers, many other factors play important roles as well. Factors such as specific attractions, local character and road condition create localized impacts while factors like urban form and the state of the economy impact travel patterns throughout the network. Other important factors include the purpose of the trip (i.e. work, school, shopping, and recreation) and demographic group (i.e. age, gender, income, race) [10]. A more detailed discussion of this topic can be found in the Guiding Perspectives report on Factors Affecting Transit Ridership.

FIGURE 1: ETS DAILY DEMAND PATTERN

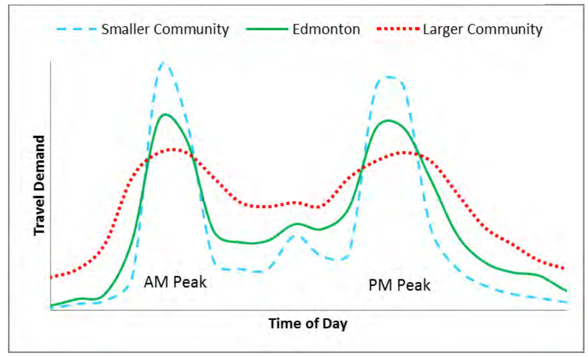


FIGURE 2: ETS ANNUAL DEMAND PATTERN

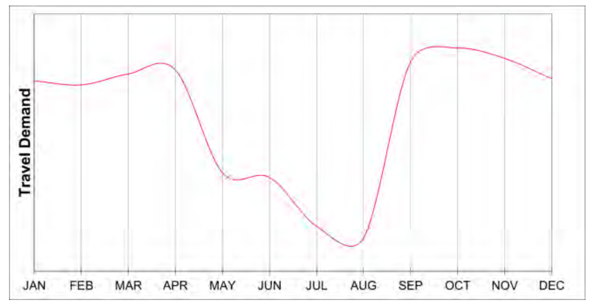


FIGURE 3: ETS LONG TERM DEMAND TREND

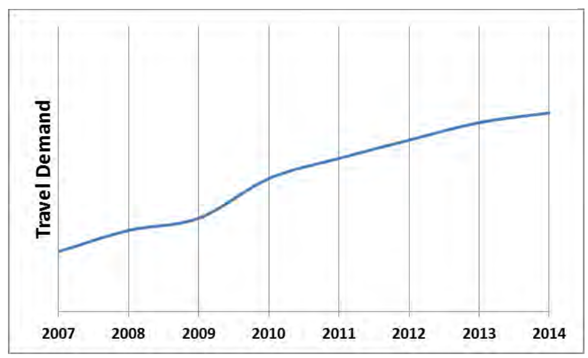
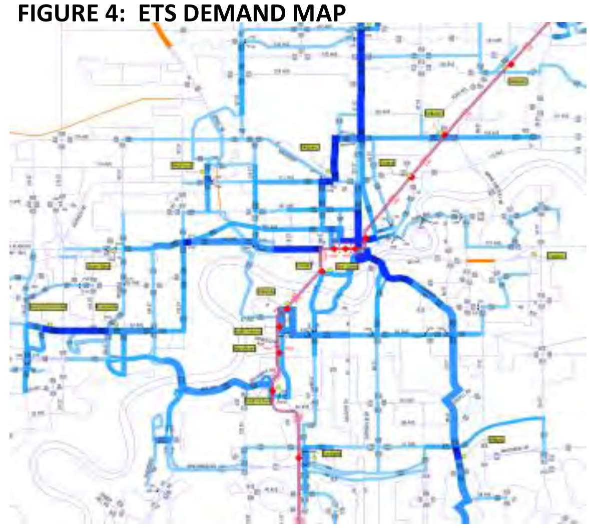


FIGURE 4: ETS DEMAND MAP



Captive and Choice Riders

Not all users have the same options available to them when deciding how to travel. Many users are unable to afford a personal vehicle or cannot operate one due to physical limitations. Users who rely on transit are often called “captive riders”, while those who can easily drive, or be driven, are called “choice riders”. [4] As expected, demand from choice riders varies considerably with the level of service, convenience and fares while demand from captive users is more consistent, so long as they are still able to use the service. Transit agencies must be careful not to focus on attracting choice riders at the expense of captive riders, as this can result in potential equity issues as discussed in the Guiding Perspectives report on Social Sustainability in Transit.

Modal competition

For choice users, transit must compete with other modes such as personal vehicles and taxis. This competition results in some travelers using transit while others use other modes, with the relative amounts of each termed “Mode-Split” or “mode-share”. In theory, the likelihood of an individual choosing one mode over another depends on the relative value of each mode. The competitiveness of transit is thus based on factors such as fares, travel time, how the service is perceived, and the relative cost of other modes. As such, transit can be made more competitive by higher LOS, as well as improving perception through branding, cleanliness and comfort. In addition, other modes can become less attractive through congestion, reduced parking availability or higher parking costs, although imposing such barriers can harm businesses and may be unpopular. [10]

1.4.3. Induced demand

While attractions in an area largely determine travel demand, the transportation system plays a role in how appealing that area is, forming a type of feedback loop called “induced demand”. For example, if the travel time to a mall is reduced, that mall is likely to become more attractive and thus see an increase in demand. This ability of improved transit LOS to increase demand must be considered when planning any transit project when estimating capacity. The magnitude of the induced demand is dependent on factors such as the change in travel time, reliability and capacity of the service as well as the current destinations along the line. [4] Public perception of the improvements created in order to induce the demand is also critical. For example, a newly constructed LRT line is likely to induce more demand than additional bus service, even if travel times and capacities are similar. [11]

2. Transit purpose and Service

2.1. The purpose of Transit

Continuing with the previous analogy of the pocket watch, the carefully designed and scheduled service lines might be thought of similar to gears and springs. Together they form a complex network that is laid out and timed to provide the appropriate level of service through a large area, during many time periods. Most transit users are only familiar with a small part of this network and so have a limited understanding of the larger intended purpose it is meant to accomplish. More than a few individuals have complained about a legitimate issue with some aspect of the service and passionately provided what they feel is an obvious solution. If their request is refused, they may become upset and question the motives and priorities of the transit agency. This common situation is not due to disinterest or lack of empathy by administration, but different understandings of what transit is supposed to do and limitations on how it must operate.

Determining the purpose(s) of transit is at the very heart of the Transit Strategy, and is no small task. As with any public service, there are a number of stakeholders with differing views that must all be considered. At the highest level, public transit has three groups of stakeholders: users, operators and the community. As shown in the Table 2, users are most often concerned with individual experiences on the current service while the community is more concerned with larger scale impacts on the long term. The agency providing the service must connect these two extremes by meeting current service standards efficiently while addressing customer concerns and supporting the long term visions of community leaders.

TABLE 2: GOALS BY GROUP

[2]	**Users**	**Operators**	**Community**
Primary Goal	Positive individual experience with a low perceived individual cost	Meet predefined requirements with available budget	Maximize benefits to community while minimizing costs to community
Time Frame	Current operations	Current operations and short term plans	Long term plans
Target Group	Personal/individual	Users, municipal leaders & management	Community as a whole
Example Indicators	● Easy to access service o Temporally o Physically ● Easy to understand and plan trips ● Predictable and reliable service ● Short door-to-door travel time ● Frequent service ● Comfortable & safe facilities ● Reasonable cost to use	● Operational requirements ● Total operating costs ● Total capital costs ● Efficiency: o Cost per passenger-trip o Cost per capita o Ridership per capita ● Km traveled per passenger-Km	● Long term public costs ● Social goals ● Equity ● Worker access ● Access for Special needs groups ● Economic goals ● Livability & sustainability ● Environmental pollution ● Urban form & density

Despite the variety of conflicting viewpoints, a clear purpose must be decided upon. Operational service standards provide the agency with critical information needed to provide current transit service, but lack the clear reasoning and long term perspectives that are also needed. Inevitably, situations will arise where difficult decisions must be made and additional guidance is needed. One common example is where service is operating below minimum ridership levels but cannot be removed due to coverage criteria. Other common scenarios include where not all warranted service can be provided, where service standards can be met in more than one way, and/or where service standards conflict with other municipal policies. Furthermore, service standards only reflect current needs, and so are of limited value when making decisions regarding future service. [12]

The purpose of transit can take a variety of forms, as shown by the variety of published objectives, goals, missions and vision statements given by various transit authorities. [2] Each of these has unique nuances that reflect the needs and values of the area served. However, in order to provide the needed guidance, they must provide clear priorities which can be presented in a variety of ways. One way to consider them is in terms of several broad questions. For this discussion, we have four such questions posed as dichotomies. In actuality, neither extreme is practical and so a balance between the two must be found.

- **Should transit benefit those who use it or the entire community it serves (including non-riders)?**
- **Should transit serve as many people as possible or should transit focus on specific groups over others?**

- **Should transit service accommodate the current needs or support long term goals of the community?**
- **Should transit performance be high or should funding be used elsewhere?**

These four questions are far from comprehensive, but taken together they can provide an adequate description of the over-arching purpose of transit.

2.1.1. Individual vs Community Benefits

Transit service provides both a benefit to the individual user, in the form of transportation, and also to the entire community in the form of reduced pollution, reduced congestion and by supporting social and economic goals. While both individual and community benefits are inevitable, their relative magnitude and how each is valued can vary dramatically. This relative value is particularly useful when determining how much of the cost of providing the service should come from public funding and how much should come from user fees.

Individual costs & benefits

Individual benefits are the ability to travel to various destinations, which is related to connectivity and LOS. However, actually monetizing such benefits to individuals is often not possible. Instead, the individual benefit is often monetized by using "The Law of Demand" which assumes that the value of a good or service is equal to the price consumers are willing to pay for it. [4] [13] The most obvious cost to a user is the fare or "out of pocket" (OOP) cost however, travel time and "hardship" should also be included resulting in a "generalized cost". [2] This general cost is most often presented as either monetary cost or travel time. It is calculated using a conversion factor called the "value of time" (VoT) or "time value of money" and represents individual preferences as well as factors such as comfort, safety, and ease of use. If the generalized cost of a given service is assumed to be constant, users will be willing to pay more for service with a shorter travel time or a service that is more comfortable (and thus has a lower VoT) [4].

Calculating the VoT and maximum OOP can be done using "demand elasticity". This relates changes in ridership to changes in parameters such as fares, travel time, comfort and safety. This method determines average values for segments of the population using data collection and calculations and is beyond the scope of this paper. A more detailed discussion of this topic can be found in "The Way We Finance - User Fees White Paper" published in March 2016. [13]

Societal (or public) costs & benefits

Public benefits of transit are more difficult to quantify as the benefits are often subtly spread across the entire area and accrued over the course of many years. For example, the impact on development patterns, urban form and economic strength is difficult to determine due to numerous larger contributing factors and abstract benefits such as "livability", "civic pride" and "social equity" which have no agreed upon measure or monetary value. [13] Even those aspects that can be estimated, such as reduced congestion on roads and pollution, usually require a large amount of data and complex modeling. While it is seldom possible to accurately calculate the precise value, it is often possible to estimate relative changes using known relationships. Several such relationships are shown in the Table 3. In general, most of the benefits are thought to increase by increasing numbers of individuals choosing to take transit instead of personal vehicles. The exception to this is the public benefit related to vulnerable populations, which is discussed in the Guiding Perspectives report on Social Sustainability in Transit and briefly discussed in the section below.

Public costs are much easier to monetize as they are typically considered to be the total operating costs of the entire system along with total capital costs to build it. Direct operating and capital costs can usually be determined, although there may be some ambiguity when considering indirect costs or those that are shared with other groups [13] [2]. Another source of cost is in the footprint of the land dedicated to transit facilities and disruption of other traffic caused by transit operations and facilities.

TABLE 3: INDIVIDUAL AND PUBLIC BENEFITS

[2]	Individual Users	Community/Public
Increase Benefits	• Reduce Travel Time. • Improve comfort and ease of use • Improve perceived safety	• Increase ridership • Reduce personal vehicle trips • Support vulnerable populations
Increase Costs	• Increase Travel time • Reduce comfort and ease of use • Reduced perceived safety	• Higher operating subsidy • Higher capital Investment • More land required

2.1.2. Greater numbers vs special needs

Individuals with special needs (such as persons with disabilities and/or limited finances) pose a unique challenge to public services such as transit. Such individuals usually make up a small minority of the population, but accommodating them requires substantial funds. [14] Furthermore, they are less able to pay for these services and are more harmed by the loss if they are taken away. Common examples include low income families who live in areas with low transit demand and those with limited mobility who can only use vehicles or facilities with special devices. [14]

This challenge can be seen from several opposing points of view. If transit is intended to provide transportation for everyone without favoritism or discrimination, the focus should be on benefiting as many individuals as possible. From this perspective, money should be spent on improving service in the most important areas, maximizing vehicle loads and improving travel speed. [2] Service to areas with low demand should be reduced or cut and fares based on market pricing. On the other hand, if transit is intended as a support for those with special needs, then passenger volumes and travel speeds are much less important. Instead, money should be spent on providing service as close as possible to as many locations as possible, regardless of the demand or travel speed. [14] Furthermore, many discounts should be offered to those with financial need and vehicles should be designed to accommodate those with limited mobility and cognitive impairments. [14] However, providing such accommodations often reduces the capacity and travel speed while increasing costs. [12] Further complicating this issues are the shared benefits from either choice. For example, service intended to maximize ridership will also be used by many individuals with special needs, and accommodations for special needs are often enjoyed by other passengers as well. Examples of the latter include kneeling buses which make boarding and alighting easier for all passengers and wheel chair accommodations that also work well for strollers. [14]

2.1.3. Current Demand vs Long Term Goals

Another important choice that must be made is between accommodating the current demand and supporting long term goals. This tradeoff is particularly important in cities that are rapidly growing or going through large scale redevelopment (Edmonton could be considered to be in both categories). Established cities, such as New York and Paris, with a mature urban form and stable travel demand often need only maintain or renew the existing infrastructure for incremental growth. In contrast, rapidly growing cities and those undergoing fundamental changes must build infrastructure intended to support a future demand pattern that is very different from the current one.

Meeting the current travel demand is essential to support the economy and growth of the municipality, as well as the viability of the transit system. Failing to meet the current needs of the city can have a range of negative impacts, which can undermine any long term plans for the city. [2] Furthermore, poor service can create negative public perception, along with a loss of political support, which can reduce funding in the short term and limit the role of transit in long term plans. [2]

However, many growing cities are shifting to a greater use of transit and a more compact urban form. Making this transition is difficult, and often results in a "chicken and egg" dilemma, where the desired urban form requires transit infrastructure be in place, but current demand does not justify construction of that infrastructure. [14] To overcome this dilemma, cities must entice land developers by constructing, permanent transit facilities, such as tracks and stations, to show a long-term commitment to high quality transit. However, taking this step is risky and difficult. In addition to financial investment, such construction often requires prolonged disruption of critical corridors used by existing transit services. [2] Because of this, supporting long term goals may reduce funding available for transit operations, while simultaneously disrupting current service and potentially increasing demand.

PERMANENCE VS FLEXIBILITY

As mentioned above, permanent transit infrastructure is often required to provide developers with the confidence they need to build compact, transit friendly projects. This topic, which is explored in detail the Guiding Perspectives report on the Land Use and Transit Connection, highlights an important aspect of the tradeoff between current needs and long term plans: flexibility and permanence.

Flexibility is often thought of as a universally desired goal of transit agencies, which seems reasonable as it allows the operator to adjust the level of service to correspond with demand. [2] However, flexibility is a broad term with several meanings, and permanence (which is the opposite of flexibility) has many advantageous characteristics as well. As shown in Table 4, flexibility can result in reduced efficiency, higher operating costs, less attractive service and limitations on the type of service which can be provided. This is not to say flexibility is universally detrimental, simply that the relative importance of flexibility and permanence must be carefully considered for each individual line. [2]

TABLE 4: DRAWBACKS TO FLEXIBILITY

Flexible Item	Examples	Drawbacks
ROW and Alignment	• Length of line • Travel path • Areas served • Stop locations • Coverage	Lack of permanence reduces attractiveness and impact on urban form Lack of permanence makes facility enhancements less practical Permanent alignment required for dedicated ROW (needed for rapid transit) Limitation on size and type of Transit Unit Likely impact on travel speed and efficiency
Operations	• Frequency • Hours of service • Type of service	Potential for change in LOS reduces the attractiveness of development. Storage areas for extra TU when service is reduced Split shifts are less cost effective and less attractive to drivers
Technology	• Transit Vehicle type • Fare acceptance • Data collection • Schedule updates	Lack of consistency will result in performance penalties Lack of consistency will increase operating costs Lack of consistency may result in more complex operation Inconsistent data collection is likely to result in reduced data quality

2.1.4. Cost vs Performance

The common cliché "You get what you pay for" is often used to imply higher performing products and services typically cost more than inferior ones. This relationship is common, particularly where there is competition, regulations and shared knowledge. Even so, the cost and chosen performance should be appropriate for the actual need. For example, frequent rail service would be inappropriate, and wasteful, for an area with very low demand potential. Usually, this forces both the agency and municipality to make tough decisions regarding actual needs and the transit service they are willing and able to pay for. [10]

Costs can be considered in several ways, with the most obvious type being financial costs related to construction, purchase of vehicles and operation of the service. [12] Capital costs generally include construction of facilities such as rights of way, stations, garages and other support buildings as well as vehicles. The amount of service provided, is often considered in terms of hours of service or kilometers traveled, which are both directly related to operating costs such as operator wages and fuel and maintenance costs. There is an important trade off that must usually be considered, in that service with a higher capital cost often has a lower operating cost when considered in terms of cost per passenger. This is due to the higher capacity and efficiency associated with dedicated rights of way and larger vehicles. [12] [2]

In addition to financial costs, land use footprint and opportunity costs must be considered. As urban density increases competition for land grows dramatically with buildings, roads, sidewalks and bike paths all vying for the diminishing open space. [10] Allocating space to transit can have a lasting impact on how the area will develop, in addition to potentially angering motorists, developers and pedestrians. This is a serious conflict, even within the concept of "transit oriented development" (TOD) where planners must often choose between placing bus transfer facilities or other large developments in the land adjacent to LRT stations.

2.1.5. Clear statement of purpose and priorities

When taken together, the choices regarding these four tradeoffs, along with other, agency specific concerns, can provide a basic structure for the intended purpose of transit. Additional tough choices and consideration are still required to develop a clear purpose with priorities. This can be accomplished in many ways, usually requiring agency management and political leaders along with the aid of technical guidance from within the agency or external consultants. [11]

2.2. Quality of Service

Once the purpose of transit is defined the agency must then determine how to provide service that accomplishes it. This also requires difficult decisions between important tradeoffs, although decisions at this level have more apparent impacts on what groups and areas will see service improvements and reductions. In general, these questions are more technical in nature and the agency can usually play a much larger role in accomplishing this with less input from civic leaders and citizens required.

As with the purposes, these decisions are presented as dichotomies for illustrative purposes, although in reality a balance between the two must be found as neither extreme is practical.

- Should transit service connect as many places as possible or focus on the most popular and important connections?
- Should transit service provide direct service without transfers or should service be standardized and well connected?

- Should transit service be focused during the busiest time periods or provide consistent service?
- Should transit service focus on operational performance or providing the best experience to customers?

These four questions are far from comprehensive, but taken together they provide an overview of the major trade-offs that must be made regarding the levels of service. In the paragraphs below, each of these is briefly described followed by a discussion of how it relates to the trade-offs related to the purpose of transit that were previously discussed.

2.2.1. Spatial coverage

One of the most challenging service tradeoffs is between providing a high degree of coverage throughout the entire service area and providing higher levels of service to the corridors and areas with highest demand. Providing service to every part of the coverage area reduces walking distances, increases potential destinations and better serves vulnerable individuals. However, this leaves fewer resources to improved service where demand is higher and so those lines may be overloaded and less reliable [1]. In addition, service to areas with low density will often see low ridership and service with many stops is often very slow, making trips to distant areas slower and more likely to require multiple transfers. Furthermore, providing the widely spread coverage often results in less frequent service and so waiting times are likely to be quite long. [1]

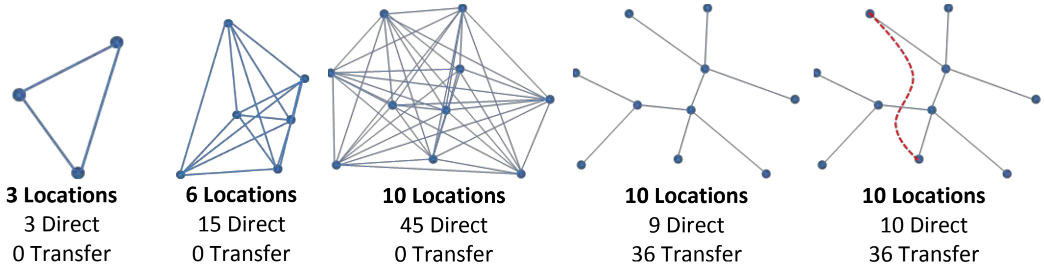
- **Individual vs Community Benefits:** High coverage tends to provide community benefits by ensuring that vulnerable individuals have access to transit service. This not only provides a higher quality of life for those individuals, it also enriches the community and supports a number of social programs. The direct effect of focused coverage is a benefit to individual passengers through less crowded vehicles and faster travel speeds. At the same time, the service may be more attractive to choice riders which can result in secondary benefits to the community through reduced congestion and pollution. [1]
- **Ridership vs Special Needs:** Clearly higher coverage will benefit those with special needs while reducing the quality of more popular service. Likewise, focusing coverage on the most critical corridors and locations will benefit a large number of riders at the expense of those in less popular areas, including any individuals with special needs. [15]
- **Current Demand vs Long Term Goals:** High coverage can support current goals, but is typically thought to provide a greater long term benefit by supporting social initiatives. Focused service can be used to support long term goals, but is often more effective at meeting current needs, and may even encourage unwanted outward growth. [10]
- **Cost vs Performance:** High coverage typically sees lower performance, lower fare revenue and higher operating costs as the service is often both slow and poorly used. In contrast, focused coverage usually sees high performance and efficiency, and increased fare revenues which together can reduce costs. [1]

2.2.2. Connection Type

Connection type, in this context, is the choice between direct service and reliance on transfers. This topic will be discussed in more detail in part three, so only a short overview of the trade-off is mentioned here. Direct service simply means that passengers can travel between locations without the need to transfer between different vehicles or lines. [2] Direct service can use fast express lines, slower local service or lines that are some combination of the two. Each line can be customized for the specific areas it serves. [2]

However, as the coverage area expands and the number of possible locations increases, direct service between all of them becomes increasingly complicated and inefficient as shown in Figure 5 below. As such a system of transfers becomes necessary to efficiently provide of service between many areas. However, this forces many passengers to transfer between lines which is inconvenient and may add delays if schedules are not properly synchronized. [2]

FIGURE 5: NETWORK CONNECTIVITY



Direct service requires many more connections as the network grows making it very complicated and inefficient. A transfer based network dramatically simplifies the network, but additional direct service may be overlaid to reduce transfers between specific locations.

Even in networks that rely on transfers, some amount of direct service is always inevitable. However, it is not uncommon for extraneous direct lines to provide convenient service between specific locations. Such service may be provided for a variety of reasons, such as heavy demand for a particular connection, or benefiting a group living in a specific area. They can improve the performance and efficiency of the network if they are heavily used, but can have the opposite effect if ridership is low. Regardless, such extraneous lines provide redundant service, increase the network complexity and only benefit individuals following a specific travel pattern. [2]

- **Individual vs Community Benefits:** Heavy use of direct service can provide benefits to the community by attracting motorists to transit and reducing loads on other lines; however, in most cases this is overshadowed by the benefit to the individual riders who use the convenient service. Transfer based service can benefit the entire community through uniform coverage and, potentially, increased efficiency. [3] [1]
- **Ridership vs Special Needs:** Direct service can increase ridership by improving service along important corridors or be provided to areas with a large concentration of individuals with special needs. [1] (Obviously, direct service provided to areas with neither vulnerable populations nor high demand supports neither equity nor equality.) Transfer based systems are, however, intended to benefit as many users as possible. [2] Furthermore, all passengers are required to transfer, regardless of need or physical ability.
- **Current Demand vs Long Term Goals:** While it is possible to use direct service to support long term goals, direct service is more often provided due to current demand, often to remote suburbs and so can actually work against long term goals. [2] Transfer based networks also support current demand although long term goals are often included when determining locations of transfer points, particularly if upgrades or amenities are provided at them. [1]
- **Cost vs Performance:** Networks relying on direct service are usually inefficient for service in large areas when compared to transfer networks. However, transfer based networks often require upgraded facilities at transfer locations which can increase capital costs. Also, extraneous direct service lines in a transfer based network can have either a positive or negative impact on cost and performance. [2]

2.2.3. Temporal Coverage

No transit system provides constant service during all time periods and all days of the year. Instead, additional service is provided during periods of high demand (such as rush hour) and service is reduced during periods of low demand (such as late at night). While running empty buses is inefficient, travel outside of peak times is still important, removing important connections during off-peak times undermines transitioning to an auto-free lifestyle and degrades the quality of life for those dependent on transit. [16]

- **Individual vs Community Benefits:** Providing service during periods of lower ridership clearly benefits those riders who do use the service, but also benefits the community by providing transportation options, supporting vulnerable populations and encouraging transit friendly lifestyles. Which benefit is greater is a matter of debate. Limiting service to only times when demand is high allows for higher levels of service during those time periods which can encourage ridership and improve efficiency, although the community benefit is largely offset by limiting off-peak travel to personal vehicles. [13] [11]
- **Ridership vs Special Needs:** While service during low demand time periods can increase overall ridership, individuals with special needs who are reliant on transit usually benefit more than the general public [14]. Focusing on peak service promotes ridership [1] while limiting travel options for vulnerable populations who are dependent on transit. [17]
- **Current Demand vs Long Term Goals:** As expanded off-peak service is not warranted by the current demand, such service is thought to support long term goals, particularly those related to social equity and mode shift. [11] [2] Focused service is reactive to the current demand and can work working against long term goals such as auto-free lifestyles. [14] [2]
- **Cost vs Performance:** Providing unwarranted off-peak service can be expensive to provide and often sees low performance [2]. On the other hand, focused service typically improves performance while also generating revenue and thus potentially reducing operating costs. [2]

2.2.4. Customer Experience

Customer experience is critical to attracting passengers and maintaining support for transit. While operational performance is important to customer experience, so too are, amenities, cosmetic appearance, information provided and perceptions of safety. Unfortunately, such items are often expensive to provide and have little impact on operation. [2] This is particularly true for large bus networks with thousands of stops and service spanning well into the night.

While customer experience cannot be entirely neglected, it is possible to provide only a minimum level that most passengers are willing to tolerate [2]. Doing so can reduce costs, but often results in passengers developing a negative impression of transit service and the agency providing it [2]. A poor public image can work against almost every purpose transit might have by reducing ridership, public funding and consideration in long term plans.

- **Individual vs Community Benefits:** Improving customer experience provides an obvious benefit to the individual customers, while reducing costs can provide funding for other programs or improving service quality. On the other hand, public perception of the transit system is important because it influences many of the subtle impacts of transit such as municipal pride and livability [8].
- **Ridership vs Special Needs:** Amenities, comfort and security are clearly important to retaining ridership, although the impact on attracting riders is unclear. However, many amenities are particularly beneficial to individuals with special needs [14].
- **Current Demand vs Long Term Goals:** Improving customer experience clearly benefits current users while positive experience and amenities also support long term plans. On the other hand, cost cutting can provide funding for other programs or service improvements, in both the current and long term. [2]

- **Cost vs Performance:** Improving customer experience increases costs substantially while having little to no impact on performance. In fact, some amenities can actually reduce operational performance by reducing passenger space on vehicles and slowing movement [2].

2.3. Integration with the Service area

Another important factor of transit service is how it is integrated with the land uses and the urban form. While factors discussed in the previous section were tradeoffs between aspects of transit service, this section summarizes several of the ways that urban form impacts transit. Unlike previous sections, these challenges are difficult to refine into a simple dichotomy and so are instead presented as open ended questions:

- How should transit support a city growing up, in and out?
- How should transportation infrastructure be allocated between personal vehicles and transit?
- To what extent should the road pattern and urban form be designed to support transit?

2.3.1. Growing cities and transit

The growth pattern in Edmonton can best be described by the title of the 2014 Annual Growth Monitoring Report "Our Growing City: Up, In and Out". While this growth pattern may be appropriate, it poses challenges for transit. Inward and upward growth will increase transit demand and traffic congestion, both of which tend to reduce bus travel speeds, and thus the level of transit service provided. Furthermore, these new developments are adding new destinations that must be served while competing with transit for land and imposing new restrictions on road usage. This is compounded by outward growth which expands the coverage area and increases travel times. Travel by transit, particularly on shared roadways, often takes longer than personal vehicles, particularly if transfers are required and so transit becomes less competitive as trip lengths increase. This often results in more vehicles on the roads and thus further increases congestion.

Providing transit service in rapidly growing cities, such as Edmonton, is a challenge with few easy answers. Capacity and speed of transit service must increase and the network must accommodate changing travel patterns. The challenge is how to provide these improvements. Expanding service into new areas can be served by either extending current lines or adding new ones and travel times can be improved by overlaying express lines or adding buses to reduce overloads [2]. However, each additional bus increases traffic congestion, each new area served puts additional strain on the current infrastructure, and none of this comes without additional cost.

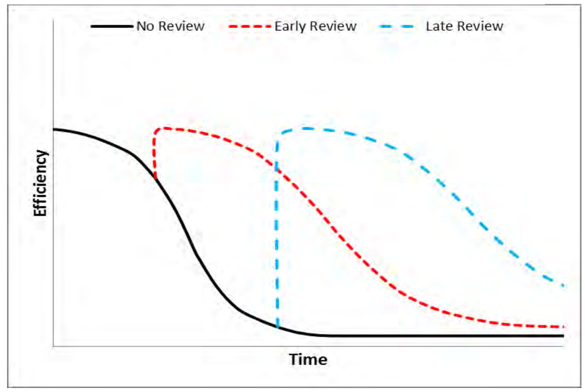
In order to reduce the impact of congestion and improve bus performance, infrastructure upgrades are required. Possibly the least expensive upgrade is to use traffic signals that provide priority to busses [10]. More dramatic performance improvements require dedicating lanes to bus use only, although this reduces the lanes available for other traffic. Even higher performance can be achieved by constructing separate rights of way such as busways or rail tracks, although this option is expensive and usually takes years to complete. [10]

Potentially more difficult is deciding when and where to provide the upgraded service. Should new areas be given service to encourage early adoption of transit, or should existing service be improved as an incentive for infill? Providing high quality transit to suburbs encourages sprawl [8] but lower land values often push vulnerable populations to those areas [14]. If existing service is not upgraded, development within the core areas is less likely and existing service may become overloaded with "downstream" riders transferring from service in the suburbs [2]. Unfortunately, providing upgraded rights of way in core areas is often very difficult and expensive as existing home, businesses and utilities must be removed [10]. Even within developed areas, it is often difficult to decide where to upgrade service as any potential choice will benefit some to the detriment of others.

Evolution of transit networks

Just as cities evolve as they grow, so do transit networks. When first created, transit networks are typically optimized for the current demand. Over time, priorities change, growth often deviates from projections and budgetary constraints limit network expansion. As such, the transit network will tend to grow in a piecemeal fashion based on the current needs and short term projections, sometimes referred to as the "shortsightedness tragedy". [2]

FIGURE 6: TRANSIT NETWORK LIFE CYCLE



This piecemeal growth often results in decreased efficiency and inconsistent service levels favoring some areas at the expense of other as shown in Figure 6. Over time, travelers become accustomed to the routes and schedules, which make it difficult for transit authorities to make changes. Hopefully, the transit authority will periodically conduct network wide revisions and develop long term network growth plans. Such revisions typically cause confusion among passengers and can have substantial costs, and should not be made too frequently. However, the severity of the confusion and magnitude of the cost tend to increase if the time between reviews is excessive. [2] [17]

2.3.2. Transit Priority on Shared Roadways

Municipal transportation systems must accommodate transit, personal vehicles and commercial vehicle traffic. However, municipalities are able to change the relative priority of each when designing the road networks and urban form. This difficult decision is critical to almost every aspect of land use and urban planning. [10] This topic is discussed in detail the Guiding Perspective report on Land Use and Transit Connection, and the only aspects to be discussed here are the amount of land allocated to transit, and the interaction between transit and other traffic on roadways.

Pedestrians, cyclists, personal vehicles, commercial trucks and transit all compete for the same limited space allocated to transportation. Use of a dedicated right of way is considered to be the single most effective way to improve transit performance, which becomes increasingly important as a city grows. However, allocating space for transit is often opposed by motorists and other concerned parties. [10]

FIGURE 7: BUS UNABLE TO LEAVE TURNOUT

Image courtesy of Google Street View

Some amount of on-street service is inevitable in any transit system and so the planners must resolve the relative priority where the two conflict. Vehicular traffic can be given priority through road design, particularly the use of bus bay or turnout at stops. Such bays reduce the impact that stopping buses have on other traffic, however re-entering traffic can be difficult for busses during peak times as seen in Figure 7. More commonly, transit is simply given no priority and so is subject to congestion and traffic signals as other vehicles. This usually results in buses traveling considerably slower than other traffic due to frequent stops and slower acceleration. [10]

Transit can be given priority over other traffic in a variety of ways, the least intrusive of which is often called “Transit Signal Priority” (TSP), were traffic signals give buses a slight advantage over other traffic. These may simply allow buses to leave a few seconds before other traffic, or may detect approaching buses and adjust the signal timing. Prioritizing transit beyond this typically requires limiting other traffic in roadway lanes. For example, transit may use a lane that is otherwise limited to specific uses, such as a turning lane that might be dedicated to transit only during specific times and otherwise used for traffic or parking. Another method is to provide short segments of separated busway at problematic locations, such as transit centres near busy intersections. [10] Beyond this, separated rights of way are typically required.

2.3.3. Road Patterns and Transit

The relative importance of transit is also critical when designing road patterns of areas buses will serve. The alignment of roads in an area and their intersections form a network that provides connections between locations within that area as well as areas beyond it. Roads can have a variety of shapes and orientations and when taken together, the network can take on a wide range of patterns which can be grouped in a variety of ways. This discussion will be limited to only four broad categories of road patterns: Rectilinear, Curvilinear, Hybrid and Organic which are shown in Figure 8.

Types:

Grid/Rectilinear

Rectilinear or “Grid” road patterns are made up of regularly spaced, linear roads, which run parallel and perpendicular to one another to form a grid of uniform rectangular blocks and four way intersections. These often have a less distinct road hierarchy as local, collector and arterial roads are typically part of the same pattern with only slight increases in width between local and arterial. The linearity and uniformity of this road pattern typically allows for highly efficient traffic flow, which as well as short-cutting on local streets when larger roads become congested. [2] [17] [4]

Curvilinear

In contrast to rectilinear, curvilinear road patterns are made of meandering roads that often change direction as they branch from one another at 3 and 4 way intersections. In many cases, curvilinear roads produce inward facing neighborhoods with internal collector loops and use of cul-de-sacs is not uncommon. The non-linear roads reduce speeding and a strict road hierarchy is often used to limit access into the neighborhood, and thus reduces shortcutting. However, the curving roads result in longer routes and reduced bus speeds while the irregular street spacing increases overlapping coverage and reduces stop spacing. [2] [17] [4]

Hybrid/Compromise

A spectrum of hybrid patterns is used that include aspects of both rectilinear and curvilinear patterns. Transit service in such areas varies with the specific pattern used. Two examples are shown in Figure 8 below.

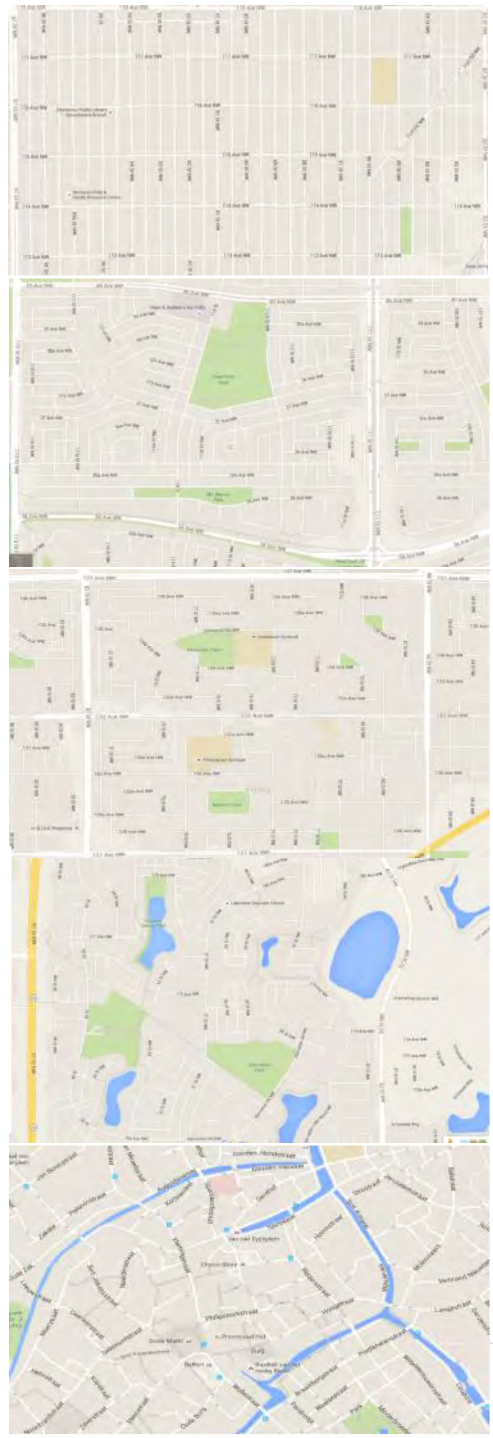
Organic/No Intentional Pattern

Even curvilinear patterns were designed with some form of overarching structure to accommodate traffic flows. Where roads were built with no such structure in mind, the result is an erratic network of streets termed "organic". Such patterns are more common in older cities or small communities that grew quickly without planners or engineers designing the road network. Instead each road, or even each block, was constructed based on the needs of the land owners and community at that time without the use of a long term plan [2] [17] [4].

IMPACT ON TRANSIT

Curvilinear road patterns are often chosen to reduce speeding and short-cutting through neighborhoods. Unfortunately, the impact on bus service is often much greater than personal vehicle traffic. Because of their size, winding roads require that buses travel slower and irregular patterns often result in more closely spaced stops. Furthermore, most routes are traversed by many buses and so the impact of even a few minutes per bus can accumulate to several hours of lost productivity each day.

FIGURE 8: ROAD PATTERNS



Grid/Rectangular/Rectilinear:

- Roads are linear
- Most parallel or perpendicular to one another
- Spacing is regular
- Most intersections 4-way at right angles

Compromise/Hybrid:

- Roads linear or slightly curved
- Many adjacent roads parallel
- Parallel roads follow regular spacing
- 3 or 4-way intersections, usually at right angles

Compromise/Hybrid:

- Roads curved or linear
- Some parallel roads
- Irregular spacing
- 3 or 4-way intersections, usually at right angles

Curvilinear:

- Most roads curved by at least 90^{o}
- Minimal use of parallel roads
- Spacing intentionally irregular
- Most intersections at approximately right angles. Emphasis on 3-way intersection other than between local roads only

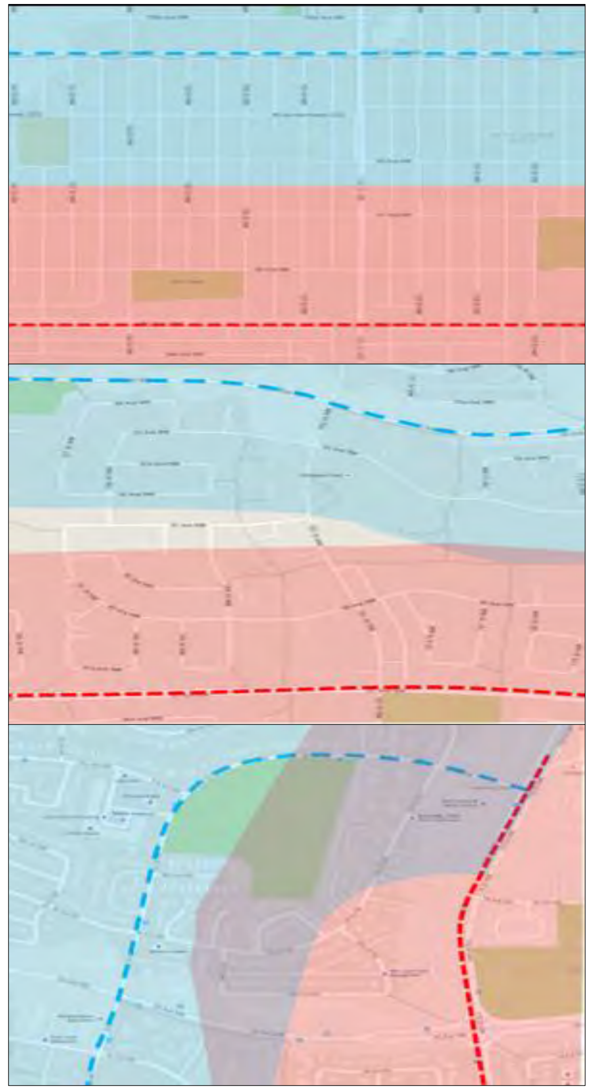
Organic/No pattern:

- Roads curved erratically
- Few if any parallel roads
- Random spacing with no clear pattern
- 3, 4 & 5-way intersections at random angles

Overlapping Coverage

An important aspect of efficiency is the amount of overlapping coverage, where a given location is within the coverage area of routes running along multiple roads. Coverage can be represented by an offset distance on either side of the road that service travels along, typically the maximum walking distance. In Edmonton, this offset distance is 400 meters in residential areas. Overlapping coverage can be used to improve service when warranted by demand. In many cases it is simply an unintended consequence of the road alignment and intersections.

FIGURE 9: OVERLAPPING COVERAGE



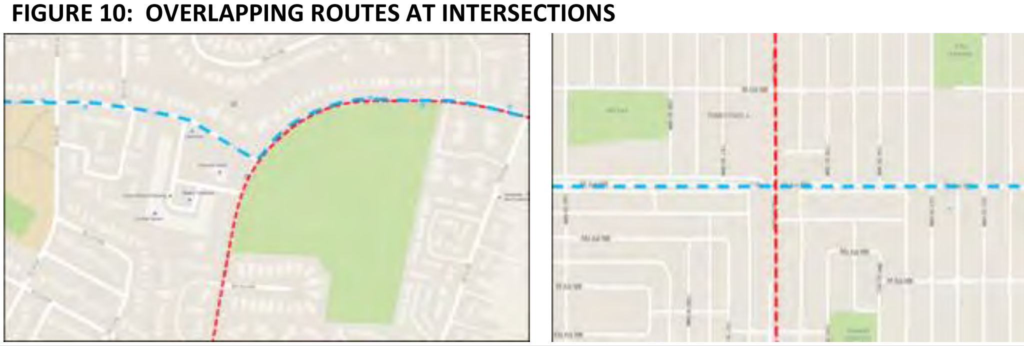
Road alignments

While the curvature of a road has only a minor impact on the distance to travel along it, buses must slow down considerably to prevent standing passengers from losing balance. Furthermore, curvilinear roads are less likely to be paralleled or regularly spaced, which often increases the amount of overlapping service. Figure 9 shows the coverage of two lines along roads with three types of alignments. As is shown in the first panel, parallel, linear roads allow for service with virtually no overlapping coverage. The second panel shows that overlap or gaps in coverage are likely result of roads with differing curvatures, with similar alignments. The third panel shows overlap when providing service along roads with very different alignments and curvature.

Intersections

Intersections between roads used by bus service provide important connections, along with overlapping coverage near the intersection. Typically, it is desirable to increase connectivity while reducing overlaps [2]. One way to accomplish this is through 4-way intersections, as 3-way intersections (or any odd number of legs) require that both lines must travel along one of the legs, thus dramatically increasing overlap (see Figure 10). In addition, the number of times any two routes intersect can be reduced, as very little advantage is gained by the same two routes intersecting repeatedly.

FIGURE 10: OVERLAPPING ROUTES AT INTERSECTIONS

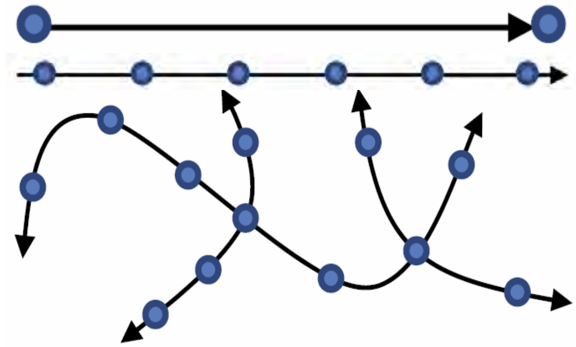


3. Transit Operation

3.1. Service Lines

3.1.1. Shuttles, Corridors & Areas

FIGURE 11: SHUTTLE, CORRIDOR AND AREA [6]



In order to understand the functioning of routes, it is important to understand the type of coverage that they provide. This can be thought of as being one of three types: a "Shuttle" service connecting only two specific points, a service that connects to multiple points along a 1 dimensional "Corridor", and a service that provides a variety of locations in a 2 dimensional "Area" (shown in Figure 11). Elevation is typically only a concern in very complex networks with multiple levels of subways and elevated facilities, and so 3 dimensional networks will not be discussed. [6]

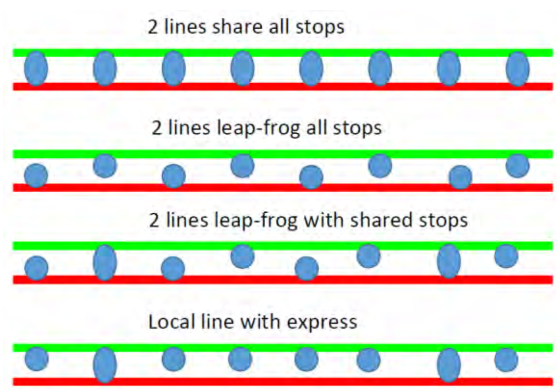
Shuttles: Simplest routes

The simplest line, called a "shuttle" has only two stops, one at each end. This is the most limited type of service as it only connects two specific areas. Such shuttles of limited use in a transit network but are often used for connecting a remote airport to the rest of the network, park & ride at special events and specialized services provided for seniors or secondary students. [6]

One Dimensional "Corridors"

Access to more than two locations is usually required for the line to be useful. Adding stops along the way provides more opportunities for passengers to access the line and provides access to more potential destinations. However, as the number of stops increase, reliability and average speed decrease, thereby resulting in longer travel times. For this reason, transit planners must choose between providing high speed service between only a small number of locations or providing slow access between many points. [6]

FIGURE 12: CORRIDOR CONFIGURATIONS



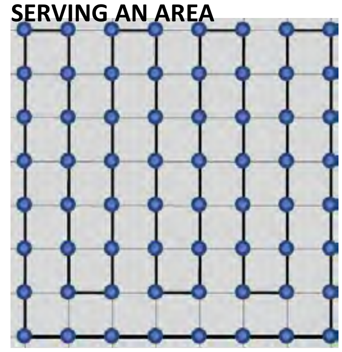
A common solution to this problem is to run a second line along the corridor. This allows for several potential scenarios as shown in Figure 12. The simplest is for both to stop at all the same stops, which is functionally identical to increasing the frequency on the original line. Another configuration, often called "leap-frogging", is for the two lines to each serve only every other stop. This improves travel speed on both lines but passengers on either of the lines have a 50% chance of the bus not using the stop closest to their destination. This problem can be alleviated by allowing transfers at a few shared stops, although transfer delays can increase travel times. [6] [2]

If many passengers are traveling to distant locations, using an express service may improve performance further. In this scenario, one line serves all of the stops while the other only uses the shared stops. This results in a slow moving local service line and a much faster express line. Passengers can then transfer between the two at shared stops. Unfortunately, express service often requires transfers, complicates schedules and does not improve service between nearby locations. [6]

Two Dimensional "Areas"

More often, transit must provide service throughout a large 2-dimensional area. A simple grid is the easiest way to visualize the challenges this poses. Transit service must still connect every location, although as the area grows in 2 directions (up/down & left/right), the number of locations increases exponentially. To provide service with a single line, travel time and distance will also increase exponentially. In order for travel times to remain tolerable some type of transfer scheme must be used [6]. Such schemes are discussed in section 3.5.

FIGURE 13: A SINGLE ROUTE SERVING AN AREA



3.1.2. FUNCTIONAL HIERARCHY OF LINES

Transit has two goals of access and mobility that must be provided by all lines. In order to best accomplish both, it is advantageous to use a hierarchy of lines which focus on each goal to differing degrees. This hierarchy can be described in several ways with a similar pattern. At one end of the spectrum are many slow moving lines that travel along smaller collector roads, stop frequently and, collectively cover a wide area. These are intended to connect nearby locations and connect riders to upgraded transfer points. Such lines may be referred to as "feeders", "collectors" or "local service". At the other end of the spectrum are lines that are intended to provide fast connections between distant locations. Such lines, often called "trunk lines", "connectors", or "express lines" typically use faster roadways or dedicated ROW and have fewer stops with greater distance between them [2].

Between these two extremes are lines representing some type of compromise. The most common examples are "Trunk and Branch", or "Express Extensions" that function as a slow feeder for part of the route and as a high speed trunk in other segments. Another compromise can be seen in lines that provide the same connections to distant locations as trunk lines, but travel more slowly and provide local service throughout.

3.1.3. GEOMETRY

In addition to the functional classification, lines can be described by their geometric appearance and the locations they connect. The geometry of each line will largely depend on the right of way it runs along, as well as the type of service it is intended to provide.

CHARACTERISTICS

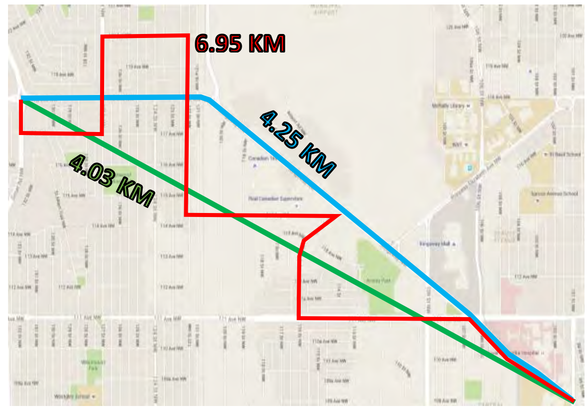
Each line has many geometric characteristics, each of which plays a large role in how the line will operate. For the purposes of this discussion, these will be limited to length, circuity, turns and orientation.

Length:

The total length, or travel distance between the terminals, plays a critical role in how the line will operate with both longer and shorter lines providing a number of potential benefits. Longer lines are preferable in terms of connectivity. They can serve a larger coverage area, provide direct service between more locations and provide transfer opportunities between other lines that are intersected. Furthermore, use of longer lines can simplify the network by reducing the quantity of service lines on maps and schedules. [1] [10]

Use of many short lines often allows greater flexibility and superior reliability. Because shorter lines have smaller service areas, the service can be more precisely matched with the demand for that area without impacting the service in other areas. This can be particularly useful in new areas where neighborhoods are developing at different rates as service to each can be adjusted as population and demand grow. Shorter lines travel through fewer areas which tends to make travel time easier to predict, and there are fewer opportunities for transfers to be be missed. As such, shorter lines tend to be more reliable [2]

FIGURE 14: CIRCUITY



Circuity:
Closely related to length is circuity of the line as it has a significant impact on both the length of the line and travel speed. Circuity is a ratio of the total line length to the straight line distance between the two terminals. For example, in Figure 14, the blue line is fairly linear with a circuity of $\frac{4.25}{4.03} = 1.05$ while the red line has a circuity of$\frac{6.95}{4.03} = 1.72$. [18]

Turns:
Closely related to circuity is how the line changes in direction due to turns at intersections and curves in roads. Any time the direction of travel changes, the vehicle will have to slow some amount and thus increase travel time. The impact is difficult to model, but usually increases with the angle of the direction change. In Figure 14, the blue line passes through a slight curve in the road resulting in two 5^{o} turns and then makes a turn of approximately 45^{o}, while the red line passes through the same curve but then makes 6 turns of approximately 90^{o}, one of approximately 45^{o} and one of approximately 135^{o}. Due to the number and severity of turns on the red line, the average travel speed would be expected to be slower than along the blue line.

Orientation and Areas Served:
Another geometric characteristic is the overall orientation, or direction of travel, of the line. While this may not provide insight into performance, it is often useful for descriptive purposes and for travel pattern analysis. For linear service, the orientation is usually obvious, but can be difficult to describe for lines that are highly circuitous. One way this can be described is in terms of the orientation of the start and end point. For example, all three lines in Figure 14 would be oriented roughly northwest. Another way of considering orientation is in terms of the areas that the line serves. These two are often similar but may be quite different when considering highly circuitous lines, express lines that do not serve all areas they travel through, or large or irregular shaped areas.

TYPES

While each line has unique characteristics, they are often grouped into broad categories based on their geometric shape. The most commonly used geometric categories are: radial, diametric, circumferential and tangential lines, which are shown in Figure 15.

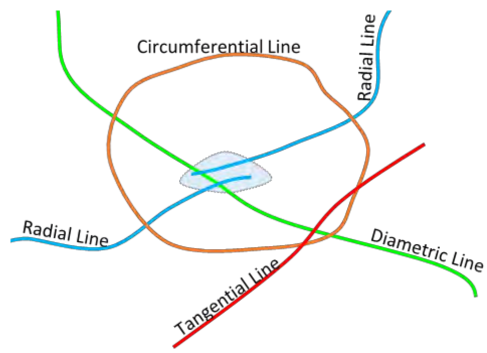
Radial & Diametric Lines
Radial and diametric lines provide connections travel from peripheral locations to the city centre. Radial lines have one termination point in or near the center and the other in a periphery area, and diametric lines are essentially two radial lines joined at a common location in the city center. As such, diametric lines may appear as a straight line, or may change direction in the center to form an “L” or “U” shape [2].

Both types have relative benefits and drawbacks. Diametric lines are essentially two radial lines joined in the city center, and thus provide direct connections between many more locations. By merit of being a single line, both operation and infrastructure can often be simplified. However, a single diametric line will usually be longer and thus likely to have lower reliability and the entire line must operate at the same frequency. Radial lines are shorter and offer greater flexibility as the service along each can be adjusted independently. As such, radial lines are potentially more efficient when serving areas with different levels of demand [2].

Circumferential Rings & Tangent Lines

Circumferential lines form a ring around the core while tangential lines connect peripheral areas without passing through the core area or forming any specific shape. These two types of lines provide connections between periphery locations without approaching the city centre, and are likely to connect to radial or diametric lines to increase access to the city center. [2]

FIGURE 14: LINE TYPES BY GEOMETRY



Circumferential rings, however, provide for many more potential connections. Such rings can be a single, continuous line or several lines with shared transfer stations. Just as with the choice between two radial or one diametric line, each configuration has advantages and drawbacks. In addition to having considerable length with uniform service, a ring formed by a single line has no termination points. Such termination stations are a source of added delay, but are also ideal locations to add or remove a transit unit from service. On the other hand, tangential lines are usually shorter and offer fewer connections but can be more reliable and service levels can be closely matched with the demand independently of other lines and areas. [2]

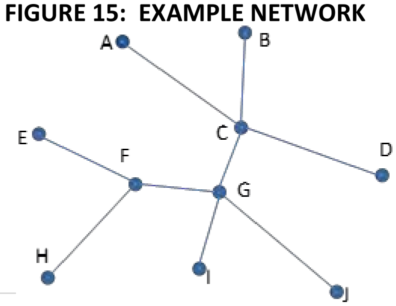
3.2. NETWORKS

Transfers between lines become increasingly important as the service area grows. Each line must be considered individually and as part of a larger network. Such network analysis is much more complicated than individual lines as any change to any line can impact transfers between each line it intersects with. As the network grows, so does the complexity and a logical over-arching pattern is needed for travelers to use and for planners to operate. Such patterns are often described in terms of connectivity, focal points and overall geometric shape.

3.2.1. CONNECTIVITY

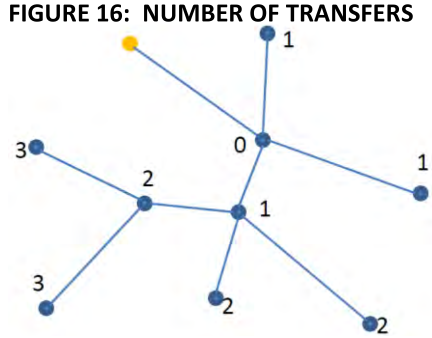
Connectivity can be described in a variety of ways; however this discussion will be limited to two basic types. At the most basic level, each line provides direct connections between all of the locations it stops at, and each stop is connected to all of the lines that use it. As such, each stop can be thought of as being connected to every stop that is connected to any of those lines. For example, in Figure 16, point C is connected to A, B, D and G, by separate routes. When transfers are considered, all points in the network are usually connected and so another indicator, such as travel time or number of transfers, is more useful [19]. Figure 17 shows the number of transfers required to travel from A to each other point.

FIGURE 15: EXAMPLE NETWORK



Network wide connectivity can be measured in a variety of ways. One way is to average the connectivity of each location, or this may be modified to weight locations by how important they are. In addition, connectivity may be thought of in terms of redundancy which allows passengers more than one way between destinations. Redundancy is also a measure of network resilience as it allows travelers a travel path during service outages or when lines become overloaded [19].

FIGURE 16: NUMBER OF TRANSFERS



Number of transfers required from origin (Orange) to each destination (Dark Blue)

Most measures of connectivity increase with the number of service lines, and the number of locations served by each. This is particularly true where additional service is provided to transfer locations. In addition, connectivity can often be increased by providing different types of service, such as local and express service, as this reduces travel times [19].

Connectivity is not without drawbacks, however. First and foremost is that the complexity of the network increases as the connectivity does. In addition, efficiency is often reduced by redundant connections, unless existing service is overloaded.

3.2.2. Focal Points

Arguably the most important characteristic of any transit network is number of focal points. A focal point is a location or area with extremely high demand and connectivity compared to other locations within the service area. The network is usually built specifically to serve these focal points and their number and location has a tremendous impact on its shape and operation. When describing networks in terms of the number of focal points, they are typically categorized as either unifocal, multifocal or ubiquitous. [2]

Unifocal networks are those with a single focal point, typically near the centre of the network. In such networks, the focal point is assumed to be the primary destination and the network is shaped to optimize travel to the focal point from other areas. This pattern can also be referred to as “many to one”, that tends to minimize travel times between locations and the centre. These high speed lines may also allow for reduced travel time between some periphery locations, although a transfer may be required at the focal point. While the network only has a single focal point, travel between other areas must still be provided, but often at a much lower level of service. Such networks are appropriate for municipalities with a highly centralized economy, typically in a downtown or central business district. [6]

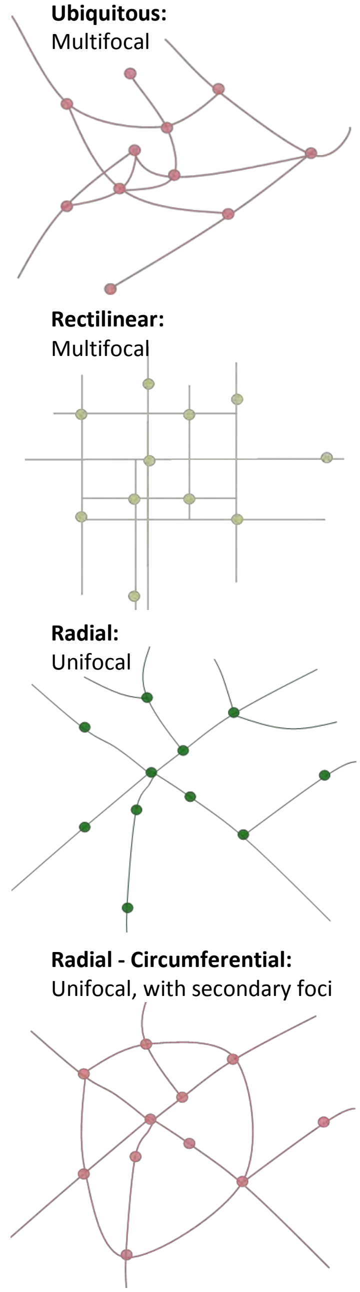
In contrast, ubiquitous networks have no clear focal point and instead either provide a homogeneous service or have a large number of focal points spread throughout the area. Such a network can be described as a “many to many” network. This lack of focus tends to provide better connections to nearby areas, but at the expense of travel times to more distant ones. Such a network is most appropriate where the economic activity is spread throughout the municipality, and not concentrated in a small number of areas. [2]

A network that is not at one of these extremes is termed a “multifocal” network, which can be described as “few to few”, or “many to few”. Such networks can take a variety of forms. Some resemble a modified unifocal network or several unifocal networks connected together while others appear nearly ubiquitous with only subtle focal points. [2]

3.2.3. Geometric Shapes

While each network is unique to the area it serves, they tend to take one of four common geometric forms: ubiquitous, grid, radial or radial circumferential as shown in Figure 18.

FIGURE 17: NETWORK GEOMETRIC SHAPES



Ubiquitous

Ubiquitous networks lack any apparent focal points, or consistent patterns. Such networks may be the result of organic growth, combinations of other network forms or simply the result of geography and land use. They are best suited to municipalities with highly decentralized economy where developments are scattered throughout. The lack of focal points tends to provide fairly homogenous service, even in outlying areas. However, traveling between locations within the network may be very confusing and is likely to require a circuitous route with multiple transfers. In addition, if a small number of areas has significantly higher demand, lines to those areas may become overloaded [2].

Grid / Rectilinear

Rectilinear networks, which are also commonly called “grid networks”, appear as a regular rectangular pattern of lines. Most connector lines have a linear routing and intersections are often at approximately right angles to one another. Much like ubiquitous networks, rectilinear networks have transfer points scattered throughout the service area, although the linear service and more regular spacing often makes these easier to use and more efficient to operate. Grid networks tend to perform well in cities with multiple focal points, particularly if the existing roads follow a regular grid pattern. [1] [2]

There are several concerns related to the grid system. First is the limited ability to adjust the level of service to specific portion of the area. If demand is inconsistent, there may be overloads in some areas and low ridership in others. Second is the large number of transfer points. Scheduling each line to support transfers at each location is impractical, so instead lines must maintain a fairly high frequency (typically 15 minutes or less). Due to this minimum frequency, areas with low demand will result in low utilization. There are also concerns regarding the cost to provide amenities and safety monitoring at a large number of transfer locations [2].

Radial & Radial-Circumferential

The final two forms, Radial and Radial-Circumferential, are closely related with the primary difference being the addition of a circumferential ring to the latter. Radial networks are dominated by lines that appear to “radiate” outward from a single focal point. These radial, trunk lines often have numerous branches as they travel to the outskirts of the city. Circumferential rings are often added to provide connections between periphery locations and expand the coverage area. Due to being focused toward a single “center” location, radial networks are best suited for unifocal networks, although circumferential rings can allow the network to accommodate smaller, secondary focal points at their intersections. [2]

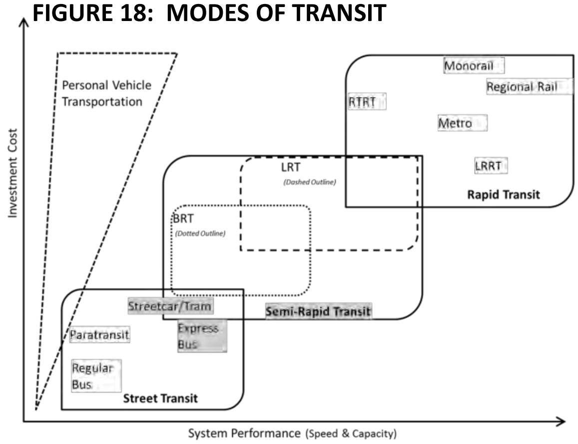
3.3. VEHICLES, ROW & MODE

Regular bus, LRT and BRT are examples of transit "modes", which are each described by the type of vehicle and right of way used as well as the type of service provided. No one mode is universally superior; each has advantages and drawbacks that make it ideal for some uses but not others. Cost and required performance are the two most critical components when deciding the appropriate mode for a given line. Figure 19 shows the relationship between these two items for many common modes. [2]

3.3.1. VEHICLE/TRANSIT UNIT (TU)

Possibly the most visible aspect of any mode is the vehicle used as it plays a large role in determining the image of the line, along with other important operating parameters. Each type of vehicle has limitations and no one is universally superior to the others. For example, the typical 40 ft. city bus, while slower, smaller and less glamorous than an LRT is able to provide service along small collector roadways where LRT would not be feasible. Typical types of vehicles and some of their primary operating characteristics and limitations are in Table 5.

FIGURE 18: MODES OF TRANSIT



3.3.2. RIGHT OF WAY (ROW)

The right of way (ROW) is closely related to the choice of vehicle and plays an equally important role in the operating characteristics and definition of the mode. There are three categories of ROW: shared, partially shared use and dedicated.

A shared ROW is used by vehicles other than transit. This is most typically a collector or arterial roadway that is used by both transit and personal vehicles, but can also refer to lanes dedicated to HOV or taxis as well as transit. This type of ROW is able to use the existing roadways used by personal vehicles and so only stops must be constructed. However, sharing the ROW with personal vehicles has the potential for numerous conflicts and delays due to congestion and traffic control. [2]

A dedicated (or separated) ROW is one that is entirely physically separated from other traffic at all locations and need not follow the alignment of other roads. Dedicated ROW not only removes any potential conflict from other vehicles, but can be designed specifically for the transit vehicles expected to use it. This allows for much higher travel speeds, reduced operating costs, and potentially automation or specialized transit unit like some BRT vehicles. [2]

This high performance comes with several significant drawbacks, the most obvious being the construction time and expense, which typically limits it to major corridors. Also, a dedicated ROW cannot efficiently provide service as close to individual homes and buildings as collector roadways. A dedicated ROW also represents a significant opportunity cost as the same space could have been used for other developments or additional traffic lanes. This use of land can pose equity concerns as the new service may benefit some areas or groups at the expense of others [2].

Between shared and dedicated is the aptly named "partially shared" category of Right of Ways, which has some characteristics of each. Such rights of way are separated from other vehicles traffic, although typically follows the same alignment as other roads and the vehicle often must stop at some or all of crossing intersections. While a partially shared ROW typically has much higher performance than a shared ROW, the limitations on alignment and conflicts at intersections result in much slower service than fully separated. In the case of busways, motorists seeing apparently empty lanes often put increasing pressure to allow HOV and other vehicles into these lanes, which quickly degrades the benefits. [2]

3.3.3. Modes

A variety of transit "modes" exist, which are typically categorized by their performance, type of ROW and support. These are summarized in Table 5 and briefly described in the following sections:

TABLE 5: MODES OF TRANSIT

Mode	**Type**	**ROW**	**Support**	**Cars per Vehicle**	**Line Capacity** [2] **(SPPHPD)**
Local Bus	Street Transit	Shared	Pneumatic rubber tires	1	3,000-6,000
Express Bus	Street Transit	Shared	Pneumatic rubber tires	1	3,000-6,000
Tram/streetcar	Street Transit	Shared or Partial	Rail (low floor)	1-3	10,000-20,000
BRT	Semi-rapid Transit	Partial	Pneumatic rubber tires	1	6,000-24,000
LRT	Semi-rapid Transit	Partial	Rail	1-4	10,000-24,000
Metro	Rapid Transit	Dedicated	Rail (heavy)	4-10	40,000-70,000
Regional rail	Rapid Transit	Dedicated	Rail (heavy)	1-10	25,000-40,000
Paratransit	Special	Shared	Pneumatic rubber tires	1	Less than 100

Local and express bus modes are used by nearly all transit agencies; both typically use 40-60 foot buses, but may also use double-decker buses or smaller vehicles. They travel along existing collector and arterial roads, and use platforms that are usually at ground level or elevated less than 1 foot. Express buses typically operate at frequencies similar to local buses and so while travel times are shorter, hourly capacity is much the same.

Trams are a form of low floor rail that run along tracks that are typically embedded in the street or use partially separated "tramways". By running on rails, trams are able to use much longer, higher capacity vehicles, but are limited to running only streets which have the rails installed. The travel speeds along tramways are largely dependent on the amount of shared ROW and the degree to which signalized intersections favor the tram over other traffic. [2] Shared tramways which also allow bus operation, such as in Istanbul, must have additional buffer spacing to ensure safety [10].

Bus Rapid Transit (BRT) is a type of semi-rapid transit representing a compromise between LRT and regular bus service. Unfortunately, this term is often used incorrectly to describe express bus lines with any amount of bus lane or HOV/bus lane. In order for a line to truly be "BRT", it must meet a number of criteria including: [2]

- ROW that is predominantly "Partially separated" with only limited sections of shared ROW
- Off-vehicle fare collection only
- Platform level boarding at all stations
- Preferential treatment at all intersections
- Clearly identifiable busway and branding to identify it separate from regular bus service
- Reliable, regular service during all day hours

This represents the minimum for a BRT line, most also use intelligent transportation technology (ITS) and upgraded stations with room for multiple buses and which connect to other modes such as LRT and regular bus. Other systems more closely resemble LRT lines and may even use specially designed vehicles resembling rubber tired trains [2].

BRT outperforms regular bus service, and often requires a smaller capital investment than trams or LRT. Due to the many similarities with LRT and lower capital cost, BRT seems a logical interim solution when regular bus service is insufficient but capital is not available for LRT construction. However, such schemes are problematic due to upgrade costs and consequences of shutting down established BRT corridor for years while upgrading to LRT [2]. Nonetheless, such interim service may be practical in some instances and so this topic is discussed in much detail in the Guiding Perspectives report on Bus Rapid Transit and Modal Integration.

Two limitations of BRT are worth noting. The first is that the level of performance is closely tied to the amount of infrastructure investment. While the simplest BRT systems require only a fraction of the capital cost of an LRT, the highest performing systems require investments similar to, or greater than, an equivalent rail system. The second is that maintenance and operating costs of BRT are typically higher than comparable LRT systems. This is due to pavement maintenance, and using smaller, diesel fueled vehicles. The smaller vehicles area able to hold fewer passengers and so more are required (and thus more drivers are required) to move a given number of passengers. Diesel fuel is nearly always more expensive than electricity and diesel powered engines usually cost more to maintain than those using electricity. [2]

Light Rail Transit (LRT) is a very common form of urban electric rail that blends elements of older metro rail lines and tram lines. These range from short trams using dedicated right of ways (ROW) to high speed trains. Typically stations are less than 1 km apart, with transit units holding 100-750 passengers, and traveling at speeds up to about 30 km/h. The primary difference between LRT and trams is the fully separated ROW, although signalized guidance, greater stop spacing and platform-level boarding are also common. Compared to metro rail lines, LRT uses smaller cars, travels slower and has much closer stop spacing. These characteristics make LRT lines well suited for urban demand patterns and integration into the urban landscape [2].

Rail Rapid Transit (RRT), also called “Metro” is designed to maximize capacity and travel speeds. To accomplish this, fare payment is always off-vehicle, platform-level boarding is universal, and stations are typically several kilometers apart. Travel speeds are further increased by full signalized guidance (instead of line of sight) and capacity is increased by use of longer cars and more cars per transit unit. The resulting ROW is highly directional, with wide turning radii that cannot be easily integrated into the urban landscape. Instead such lines typically have large, walled off ROW and are elevated or out of sight in tunnels. [2]

Regional rail lines are primarily designed for those traveling much longer distances throughout the region instead of within the municipality. The vehicles are typically larger and more comfortable with stops many kilometers apart. Such lines often use older technology and vehicles have slower acceleration and may travel at slower speeds than metro lines.

The final mode is often termed “paratransit”, and is a very specialized form of transit. Such systems typically do not follow regular schedules or routes but instead provide door to door service for individuals or small groups with limited mobility or other special needs. This is often termed “demand responsive transit”, and provides the highest level of access to passengers but does so with low passenger volumes and at a much higher cost per passenger than would be tolerated in any other mode.

3.4. Scheduling

3.4.1. General

In order to provide predictable service, each line must follow a published schedule that travelers rely on. Each schedule must accomplish several critical goals such as providing appropriate frequency of service and minimizing waiting times for important transfers and accurately. In addition, schedules must reflect the actual travel time along each segment and must allow for efficient utilization of the available vehicles.

3.4.2. Frequency

The most obvious aspect of the schedule is the frequency, which is often set by either a service standard or by demand. This can quickly become very difficult due to variations in spatial and temporal demand along the route. For example, a line may see overloads only in a small portion of the route. If the frequency was increased, the overloads would be alleviated, but vehicles will be underutilized in areas where the additional vehicles were not needed.

3.4.3. Travel times and reliability

In order for schedules to be useful, the transit unit must actually arrive and depart at the scheduled time. If a transit unit departs ahead of schedule, travelers may miss transfers to that line, and if it arrives behind schedule, they may miss transfers to other lines. In order to prevent this, the schedule must reflect the actual travel time throughout each portion of the route. Unfortunately, travel times fluctuate unpredictably, particularly during peak times, which inevitably results in delays. Such delays tend to accumulate with the longer lines being more severely impacted.

This can be alleviated by breaking each line into multiple segments with “timing points” at each end. The travel time between each timing point is estimated, and some amount of “recovery” is added to provide a buffer to offset delays. However, this scheme is only effective if the delays are less than the allocated recovery time. As such, additional recovery time is often needed to maintain reliability on segments with particularly volatile travel times. Unfortunately, any excess recovery time for a specific run will force the vehicle to wait at the timing point to avoid being ahead of schedule, which increases the minimum travel time.

For example, imagine a line with 2 segments (A & B), each of which typically takes 15 minutes to traverse (30 minutes total). Segment A is has sporadic congestion, where travel times may be as long as 25 minutes (40 minutes total). Table 6 shows travel, wait and delay times for these 3 scenarios with 0, 5 and 10 minutes of recovery time added.

TABLE 6: EXAMPLE RECOVERY

Scenario	Recovery Time	No congestion			5 minutes congestion			10 minutes congestion		
		Total Trip Time	Wait Time	Delay	Total Trip Time	Wait Time	Delay	Total Trip Time	Wait Time	Delay
1	0 min	30	0 min	On time	35	0 min	- 5 min	40	0 min	-10 min
2	5 min	35	5 min	On time	35	0 min	On time	40	0 min	-5 min
3	10 min	40	10 min	On time	40	5 min	On time	40	0 min	On time

The appropriate amount of recovery time will depend on how often such delays occur. If 10 minute delays occur regularly, then scenario 3 may be appropriate, whereas most delays are only 1-5 minutes then scenario 2 may be appropriate.

3.4.4. Transfer timing

The time spent waiting for the transfer is largely dependent on the relative arrival time of each line, as demonstrated by the following example:

Two Lines (A & B) with 30 minute headways use a single stop. Each stays at the stop only long enough to allow passengers to board and alight (~1 minute).

- Scenario 1 (Directional Transfer): Line A arrives at 5:00 and Line B arrives at 5:01. Passengers transferring from A to B have no waiting time, however passengers transferring from B will have to wait 30 minutes (until 5:30) when the next Route A bus arrives.
- Scenario 2 (Offset Schedules): Line A arrives at 5:00 and Route B arrives at 5:15. All passengers transferring between these two lines will have to wait 15 minutes.
- Scenario 3 (Synchronized Schedules): Both route A and B arrive at 5:00 and both wait for passengers to transfer before leaving. Passengers transferring in either direction have no waiting time.

Which of these scenarios is preferable will depend on the situation and priorities of the transit authority. Scenario 3 minimizes the transfer time, but requires a stop large enough for 2 buses, and each bus will have to wait longer.

This example is only a very simple transfer point. Most lines have several important transfer points, and many transfer points are used by an assortment of lines, with inconsistent frequencies and hours of operation. Also, the number of passengers transferring between these lines might vary over the course of the day or might not be known at all.

3.4.5. Efficient utilization of transit units

Another challenge of scheduling, which is less visible to the public, involves efficiently utilizing the transit unit. Ideally, each transit unit would simply remain on a given line running repeatedly, with adequate recovery in the schedule. However, service frequencies often change over the course of each day, drivers need to rest occasionally and transit units eventually need to be removed from service for maintenance and, possibly, refueling. In addition, during peak periods, the frequency along lines is often higher in one direction than the other due to the demand.

As such, transit units are often required to switch between lines or travel between maintenance facilities many times over the course of the day. To avoid confusing passengers, transit units must go “out of service” during this process. Such traveling while out of service is called “deadhead”, which should be minimized to improve efficiency. Unfortunately, due to the directional nature of demand deadhead substantial deadheading is often required. Just as with time spent in service, deadhead travel times are often unpredictable and so deadhead time is also a potential source of delays.

3.5. Network Operating Schemes

Efficient transfers are a critical to the efficient operation of the transit network and must be accounted for by both the shape of the network and the schedule. The location of transfer points and synchronization of schedules can be done according to several “schemes”. Common schemes include “Random Transfer Networks”, “Timed Transfer Networks” (TTN) and “Frequent Transit Networks” (FTN).

3.5.1. Random Transfer Networks

The term "Random Transfer" indicates that while transfers are possible, routes are not specifically designed to accommodate them. Random transfers occur at many locations in all networks as it is not possible to synchronize every possible transfer location along all routes. However, those using TTN and FTN schemes are scheduled to minimize transfer times at a number of important locations. Random transfer networks lack any such scheme, and instead lines are each scheduled independently of one another without special consideration for transfers. This may be appropriate in systems relying on direct service, or where service frequency is very high. This scheme often occurs in ubiquitous networks, systems with few routes, or may be the result of external factors such as inefficient streets, patchwork growth or multiple systems being combined. In any case traveling through such a network can be frustrating if transfers are required, particularly to those not familiar with it [2].

3.5.2. Timed Transfer Network (TTN)

This system concentrates the most important transfers at a small number of "transit centres" located in various areas throughout the city where many lines are scheduled to present simultaneously. This allows for transfers between many lines without passengers needing to wait, and so reduces transfer times. Operation is often called a "pulse" as many buses arrive and then depart within a few minutes, abruptly followed by a longer period of inactivity. In addition to efficient transfers, the TTN scheme allows for the synchronized routes to use less frequent service, so long as capacity is not an issue. In addition, the small number of transit centres can be affordably upgraded with amenities and safety enhancements to improve passenger experience [3] [2].

However this scheduling regime also comes with several drawbacks. One critical drawback is that each schedule should align with the pulse, which is typically every 15 or 30 minutes. For example, if a given route would take 20 minutes, it will arrive between pulses, so the route must either be changed or the transit unit must wait 10 minutes. Another major concern is when passengers arrive just after a pulse has ended, which might happen due to a delay or transferring from a line that runs more frequently, such as the LRT. In this case, travelers will have to wait for the next pulse, which is often 15-30 minutes. Other potential operating and safety concerns may arise from the large number of buses converging and then departing at the pulse [2].

3.5.3. Frequent transit Network (FTN)

Another scheme is to provide high frequency service, in a scheme given the name "Frequent Transit Network" by Jarrett Walker. As the name implies, this scheme requires that designated corridors have high frequency service (15 minutes or less) and transfers take place at a variety of locations along these corridors (which may include transit centres). [1]

Schedules need not be synchronized for transfers, as the high frequency of service ensures short waiting times. By no longer requiring synchronized schedules, this scheme can simplify scheduling and avoid operational concerns from pulses. Several cities that have adopted this scheme have found that the large number of transfer points can improve connectivity and the high frequency can reduce travel times. Proponents also claim such schemes can increase ridership due to the improved LOS [2].

Unfortunately, there are several potential drawbacks to the FTN scheme. The first is that the high frequency of service must be maintained regardless of demand, which can result in low utilization during off-peak times. Another concern is that this high frequency service competes for resources with service to areas with low demand, or which are distant from the FTN corridors. Removing service entirely to such areas will harm those who depend on it, resulting in potential equity concerns such as claims of discrimination. In addition, the high frequency may result in buses clustering due to congestion. [15]

4. EDMONTON'S TRANSIT NETWORK

This section provides only a brief overview of the transit network and service provided by Edmonton Transit as a more comprehensive discussion is the topic of the Guiding Perspectives report on Measuring Performance of Transit.

FIGURE 20: ETS NETWORK

4.1. OVERALL FORM

At the largest, or network scale, the ETS network appears primarily as a unifocal, radial network. This means that the network is centered around a primary focal point (the downtown core) with several much smaller, secondary, foci at Mill Woods, University of Alberta, West Edmonton Mall (WEM) and Northgate. Service radiates from the downtown until it is bounded by the outer ring road (Anthony Henday Drive). This pattern can be seen in the red and dark blue lines in Figure 20. Fifty-one separate lines travel to or through the downtown area and radial service consists of LRT lines, express bus routes (such as the 100 and 15) and local routes (such as the 1 and 8).

It is common to provide better connectivity to periphery locations and secondary foci, using circumferential or orbital lines. Edmonton however, does not have any true circumferential ring, although a partial ring could is formed by routes 127 or 137 with routes 23 or 33 (with a transfer at WEM). Instead of circumferential rings, connections are provided by numerous routes in a pattern that resembles a ubiquitous grid network.

4.2. NEIGHBORHOOD LEVEL

At the neighborhood scale, the network is primarily a collection of "hub and spoke" networks, where numerous lines in each area converge at hubs throughout the city. The ETS system uses 25 transit centres, many of which are adjacent to LRT stations as the hub locations. These hubs are then connected to form a network of express routes and LRT lines. In several areas, this hub and spoke network is supplemented by direct service to the University and/or downtown via routes that do not stop at the local hub. Examples of this are local routes with express connections such as 104, 105, 138, 139, and 160. Also, many mature and established areas are located between two or more transit centres resulting in overlapping "spokes". This is particularly apparent in the area around Castle Downs, which is served by three transit centres in very close proximity.

4.3. TIMED TRANSFER NETWORK (TTN)

In order to improve transfer times, Edmonton has a history of using a Timed Transfer Network (TTN) dating back several decades, and which is firmly established in the design of all transit centres. This means that many lines converge on transit centres within a few minutes of one another in what is termed a "pulse". The large number of routes simultaneously present allow for very efficient transfers between many routes with connections throughout the city.

While this operation has been very successful for many years, performance is declining due to three limitations. The first is that TTN are very sensitive to delays, particularly if they result in buses missing the pulse. While occasional delays are not problematic, ETS is currently reporting delays on ~35% of arrivals. The second is that the size of the transit centre limits the number of routes that can use it, and thus the connections available to customers. Most transit centres are operating at or near the maximum number of routes they can support, thus preventing some routes from using them (as well as seeing congested operation). The third limitation is modal conflicts between buses, LRT and other traffic at crossings. Edmonton has very few bus lanes and congestion in many areas is becoming very problematic. In addition, buses are delayed at rail crossings, along with all other traffic, which is problematic at crossings near transit centres. Currently only a small number of locations have transit priority signals, however they are being considered for expanded use in the near future.

5. SUMMARY

This paper is intended only as a brief summary of the key concepts of transit service models. The first sections of this paper discussed the fundamental trade off in transportation between access and mobility, which is a part of nearly every other tradeoff discussed. This was followed by clarification of several common terms, such as route, performance and demand that have some ambiguity in common usage.

In part two, purpose and service quality were discussed as a number of inter-related tradeoffs. The most appropriate balance between each of these will vary between transit systems and even within areas of the transit system. This section also included a limited discussion of how transit service can be integrated with the transportation system in the area it serves. It was discussed that transit is in competition with many other potential uses for both land and funding and the relative priority transit is given will play a tremendous role in the quality of the service it provides.

Part three of this paper discussed several aspects of how transit service actually operates. This section began with operation of individual lines, and then the operation of various types of networks that arise from their connections and hierarchy. Following this was a discussion of transit modes and scheduling. The final aspect covered was various types of operating schemes which synchronize lines to improve transfers.

Part IV was a very brief summary of the current transit network in Edmonton. As was discussed, the ETS network appears similar to a radial network with the downtown core as the focal point, but also has several secondary focal points and appears to provide additional connections by an overlaid grid network. Locations within Edmonton usually resemble a hub and spoke network plan with the transit centres forming the hubs. These hubs are then synchronized for transfers using a timed transfer network. Paper 1 includes a much more thorough discussion of the ETS network.

6. Works Cited

[1] J. Walker, Human Transit How Clearer Thinking about Public Transit can Enrich our Communities and Our Lives, Washington DC: Island Press, 2011.

[2] V. R. Vuchic, Urban Transit Operations, Planning, and Economics, Hoboken, New Jersey: John Wiley & Sons Inc, 2005.

[3] Kittelson & Associates, Inc., "TCRP Report 165 Transit Capacity and Quality of Service Manual Third Edition," Transportation Research Board, Washinton DC, 2013.

[4] Dan Boyle & Associates, "TCRP Synthesis 66 Fixed-Route Transit Ridership Forecasting and Service Planning Methods," Transportation Research Board, Washington, D.C., 2006.

[5] Dan Boyle & Associates, "TCRP Synthesis 110 Commonsense Approaches for Improving Transit Bus Speeds," Transportation Research Board, Washington, D.C., 2013.

[6] C. F. Daganzo, "Public Transportation Systems: Basic Principles of System Design, Operations Planning and Real-time Control," Institute of Transportation Studies University of California at Berkeley, Berkeley, California, 2010.

[7] The Canadian Urban Transit Association (CUTA), "Canadian Transit Fact Book," Canadian Transit Fact Book, Toronto, Ontario, 2014.

[8] APTA Sustainable Urban Design Standards Working Group and Climate Change Working Group, "Transit Sustainability Guidelines Framework for Approaching Sustainability and Overview of Best Practices," American Public Transportation Association, Washington, D.C., 2011.

[9] Edmonton transit, "ETS Statistics," City of Edmonton, [Online]. Available: http://www.edmonton.ca/ets/ets-statistics.aspx. [Accessed 20 February 2016].

[10] Ryus, P., K. Laustsen, K. Blume, S. Beaird, and S. Landgon., "TCRP Report 183 A Guidebook on Transit-Supportive Roadway Strategies (Pre-Publication Draft)," Transportation Research Board, Washington, D.C, 2015.

[11] IBI Group, "Strategies for Sustainable Transportation Planning: A review of Practices and Options," Transport Canada, 2005.

[12] Edmonton Transit, "2016-2018 Business Plan," Edmonton, Alberta, 2015.

[13] City of Edmonton, "THE WAY WE FINANCE USER FEES WHITE PAPER," City of Edmonton, Edmonton, Alberta, 2016.

[14] K. a. S. M. Bennett, "Social Sustainability of Public Transit: An overview of th Lierature and Findings from Expert Interviews," City of Edmonton, Transit Strategy, Edmonton, Alberta, 2016.

[15] Stantec Consulting with Jarrett Walker & Associates, "Edmonton Transit System Comprehensive Review," Edmonton, 2013.

[16] R. C. &. K. L. Hiroakis Suzuki, "Transforming Cities with transit," International Bank for Reconstruction and Development / The World Bank, Washington, D.C., 2013.

[17] Stantec with Jarrett Walker & Associates , "Edmonton Transit Steps Towards a Decision on Strategic Planning Version 5.0," Edmonton, 2014.

[18] j. a. L. D. huan, "Circuity in Urban Transit Networks," *Journal of Transport Geography,* vol. 48, September 2015.

[19] J. e. a. Zhan, "Characteristics on hub networks of urban rail transit networks," *Physica A,* no. 447 (2016) , pp. 502-507, December 2015.

[20] Asian Development Bank, "Comprehensive Mobility Plans (CMPs): Preparation Toolkit," 2008. [Online]. Available: https://sti-india-uttoolkit.adb.org/mod1/se6/010.html. [Accessed April 2016].

[21] Edmonton Transit, "Maps, Routes, Schedules and Service," [Online]. Available: http://www.edmonton.ca/ets/route-brochures-in-pdf.aspx. [Accessed 20 February 2016].