PART I

TRANSIT SYSTEMS OPERATIONS AND NETWORKS

1

TRANSIT OPERATIONS AND SERVICE SCHEDULING

List of Symbols

Symbol	Unit	Definition
a, $a(s)$	prs/km/h	Number of alighting passengers per kilometer of line per hour
A, B	prs/h	Cumulative number of alighting and boarding passengers per hour, respectively
b, $b(s)$	prs/km/h	Number of boarding passengers per kilometer of line per hour
c	veh/h	Vehicle line capacity
C	sps/h	Offered transit line capacity
C_o	sps/h	Scheduled line capacity
C_s	sps/h	Capacity of a line with stations
C_v, C_{TU}	sps/veh, sps/TU	Vehicle and TU capacity, respectively
C_w	sps/h	Capacity of a line section without stations
f	TU/h	Frequency of service
h, h'	min/TU	Headway
h_p	min/TU	Policy headway
h_s, h_w	min/TU	Station and way headway, respectively
ℓ	km	Passenger travel distance on line
L	km	Line length
L_p	km	Total distance from depot to a line
MLS	—	Station spacing with maximum passenger volume
n	veh/TU	Number of vehicles per TU
N, N'	veh	Number of vehicles operating on line
N_f	veh	Total number of vehicles—fleet size
N_{TU}	TU	Number of TUs
p, $p(s)$	prs/h	Passenger volume variable
P	prs/h	Utilized capacity
P_{15}	prs/15 min	Maximum 15-min passenger volume
P_{av}	prs/h	Average passenger volume
P_c	sps-km/h^2	Productive capacity
P_d	prs/h	Design passenger volume
P_{max}	prs/h	Maximum passenger volume (usually on MLS)
P_t	prs/h	Total number of passengers on line
PHC	—	Peak hour coefficient
S, S_i	km	Station spacing
S_{od}	km	Origin-destination distance
t	s	Time variable
t_a, t_b	min	Acceleration and braking time, respectively
t_d	min	Deadhead time
t_f	min	Transfer time
t_o	min	Passenger on-line travel time
t_p	min	Crew paid time
t_r	min	Running time
t_s	min	Station standing (dwell) time
t_t	min	Terminal time
t_w	min	Waiting time
T, T'	min	Cycle time
T_ℓ	s	Time lost per stopping at one station
T_o	min	Operating or travel time
T_{od}	min	Origin-destination travel time
T_p	min	Platform time
T_s	min	Station-to-station travel time
v, V	m/s, km/h	Speed variable
V_a	km/h	Access speed
V_c	km/h	Cycle speed
V_d, V_g, V_ℓ	km/h	Line design, programmed and legal speeds, respectively
V_{max}	km/h	Maximum technical speed

List of Symbols (*continued*)

Symbol	Unit	Definition	Symbol	Unit	Definition
V_o	km/h	Operating or travel speed	η	—	Operating personnel efficiency coefficient
V_{od}	km/h	Origin-destination speed			
V_p	km/h	Platform speed	η_a	—	Personnel attendence coefficient
V_r, V_s	km/h	Running and station-to-station speeds, respectively	η_s, η_t	—	Coefficient of run-cutting and schedule efficiency, respectively
w	sps-km/h	Transportation work			
w_p	prs-km/h	Utilized work			
α	prs/sps	Load factor—capacity utilization coefficient	η_x	—	Coefficient of passenger exchange
γ	%	Terminal time coefficient	φ	—	Fleet utilization factor
δ	—	Scheduled line capacity utilization coefficient			

This chapter covers the basic elements, operations, and functioning of transit systems. Transportation offered to passengers is usually referred to as *service,* while *operations* covers system management, scheduling, and functioning from the operating agency's point of view. First, the basic elements of operations and passenger flows are presented; since terminology and concepts in this area often vary, special attention is paid to precise definitions. The next section gives a review of data-collection methods and surveys; it is followed by a description of passenger demand characteristics. Methodology of scheduling and run-cutting is presented in the last section.

Transit operations, scheduling, measurements of efficiency, and related transit line planning are covered in a variety of publications. The classical textbooks and handbooks with comprehensive definitions of concepts and operational procedures that have permanent value include the following, listed by order of publication: Rainville (1947); Lehner (1950, 1957, and 1978) (in German), Rüger (1974) (in German), Groche and Thiemer (1980) (in German), Banković (1982, 1984) (in Serbian), Levinson 1992, CUTA (1993), Molinero and Arellano (1997) (in Spanish), and Kittelson & Associates (2003). An excellent collection of exercises on this topic is Krstanoski (2000) (in Macedonian). Among many publications on specific issues in transit operations the reader may also find useful Furth (1980), Bowman and Turnquist (1981), Ceder and Stern (1981), and Ceder (1987).

1.1 BASIC OPERATING ELEMENTS

The vast majority of transit services are performed by vehicles or trains traveling along fixed lines according to predetermined schedules. Thus, a transit line represents the basic component of transit system operation. Its elements are defined here.

1.1.1 Line, Network, Stop, and Station

A *transit line* is the infrastructure and service provided on a fixed alignment by vehicles or trains operating on a predetermined schedule. The infrastructure may vary from simple stop designations along a street to a grade-separated, fully controlled right-of-way with stations. A *transit route* is often synonymous with transit line, but it usually designates street transit, often overlapping lines, rather than major metro or regional rail lines. A *transit network* is a set of transit lines that connect with or cross each other and that are coordinated for efficient operation and provision of integrated services in an area for the convenience of passengers and efficiency of operations.

Line length, usually expressed in kilometers (miles), is the one-way distance between the two terminals along the line alignment. *Network length* is the total length of all alignments served by one or more lines. *Total line* (*route*) *length* is the sum of all line lengths, regardless of whether they operate alone or overlap with other lines. Figure 1.1 illustrates these concepts.

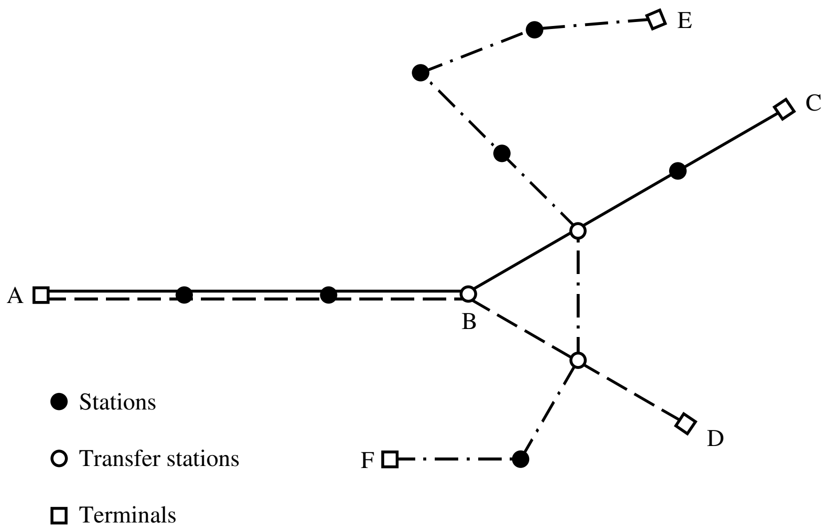


Figure 1.1 *Transit line, network, and station concepts. Assuming that each spacing between stations is 1 km long, the values in this network are:*

Line lengths: L_{AC} = *6 km,* L_{AD} = *5 km,* L_{EF} = *7 km*

Network length: $L_{AC} + L_{BD} + L_{EF}$ = *15 km*

Total line or route length: $L_{AC} + L_{AD} + L_{EF}$ = *18 km*

Transit right-of-way (*ROW*) is the strip of land on which a transit line operates; the term usually refers to the facility used exclusively by transit vehicles, but, in a broader sense, also refers to any of the numerous physical forms a transit line may follow, from street to fully controlled aerial (elevated) structure or tunnel. Among the different ROW features, the most important for transit operations, performance, and cost characteristics is the degree of its separation from other vehicles and pedestrians. With respect to this feature, transit rights-of-way are classified into three categories: *C*, *B*, and *A*.

Category C is a surface street with *mixed traffic;* transit vehicles share street space with other vehicles and pedestrians; they may have preferential treatment through visual devices (pavement markings, signs, or signals), but no physical separation from other traffic. Modes with this ROW category, such as buses, trolleybuses, and streetcars (tramways), represent the *street transit* generic class.

Category B includes a variety of ROW types that are longitudinally physically separated (by curbs, barriers, green strips, etc.) from other traffic, but with at-grade crossings for vehicles and pedestrians, including regular street intersections. This ROW category defines the generic class of *semirapid transit modes,* and its most common representative is light rail transit (LRT).

Category A is a fully controlled ROW without grade crossings or any legal access by other vehicles or persons. Also called grade separated, private, or exclusive, this ROW can be in a tunnel or a cut, on the surface, on an embankment, or on aerial structures. This ROW category is the main element defining the *rapid transit* generic class of modes, among which metro or rail rapid transit is the dominant representative. All automated guided transit (AGT) systems must have exclusive ROW, while regional rail often has some grade crossings with full signal override. This is still considered to be category A, since such crossings usually have no impact on line performance.

Capacity, reliability, speed, and other performance elements of transit lines on ROW category C and, to a much lesser degree, on category B are dependent on traffic conditions along the line, while the lines with ROW category A are fully controlled and therefore have highly reliable performance.[1]

A *transit stop* is a location along a line at which transit vehicles stop to pick up or drop off passengers; its equipment may include signs, information, a bench, and shelter. A *transit station* is a special structure and facility for passenger boarding/alighting, waiting, and transfer. It may have facilities for passengers, such as boarding platforms, mezzanines, stairways, and fare-collection equipment, and for vehicle operation, such as turnback and storage tracks. For the purpose of transit line operation analysis, the terms *stop* and *station* will be used interchangeably and will encompass all physical forms, from bus stops at curbs to major rail transit stations.

Transfer stations are joint stations for two or more lines at which passengers can transfer between lines. *Terminals* are, strictly defined, end stations on a transit line, but sometimes the term is also used for major transfer stations.

Different types of stations in a network are shown in Figure 1.1. Detailed description and analysis of transit lines and stations is given in Chapters 4 and 5, respectively.

1.1.2 Vehicles, Transit Units, and Fleet Size

Highway transit modes operate single vehicles only, with very rare exceptions (bus and trolleybus trailers are operated in Switzerland, Russia, and a few other countries). Guided modes, on the other hand, in most cases utilize their capability to couple vehicles into trains to achieve greater capacity and economy of operations. The scheduling of line operations is based on the concept of *transit unit* (TU)[2], which is defined as a set of n vehicles traveling physically coupled together. For single-vehicle operations such as buses, $n = 1$; for train operation, $n > 1$ (in rail transit terminology this n is also known as *train consist*). Thus, TU is the common concept for single vehicles and trains, whichever is used on a transit line. Frequency of operation on a line f is expressed in TU/h, so that the number of vehicles past a fixed point during one hour is $f \cdot n$.

Transit vehicles, bus and rail, are referred to collectively as *fleet*. For rail modes the specialized terms *rail cars* and *rolling stock* are often used.

Fleet size N_f is the total number of vehicles needed for operation of a line, or of an entire network. The vehicle fleet consists of the vehicles required for regular service N (determined by the peak hour operation), vehicles needed for reserve N_r, and vehicles on maintenance and repair N_m:

$$N_f = N + N_r + N_m \qquad (1.1)$$

The utilization of a fleet, which depends on its physical condition and schedule efficiency, is measured by the fleet utilization factor φ, defined as percent of fleet available for service:

$$\varphi = \frac{N + N_r}{N_f} \qquad \left| \frac{\varphi}{-} \middle| \frac{N}{\text{veh}} \right| \qquad (1.2)$$

1.1.3 Usage of Service: Passenger Flow and Volume

Transit service must be based on the demand for travel by transit along the line, as well as the required level of offered service. The demand is expressed as passenger flow (prs/time) or passenger volume p (prs/h). These two concepts represent the rate of passenger travel, or the number of passengers traveling past a fixed point in one direction per unit time. However, flow refers more to the continuous process of passenger travel in general, while volume designates the number of passengers traveling during a specified time interval, usually one hour.

Passenger travel on a transit line can be shown by a sequence of theoretical diagrams. First, assuming that

---

[1]For a detailed definition and description of ROW types and other physical components of transit systems, see Vuchic (1981), particularly Chapter 2.

[2]For definitions and descriptions of TUs see also Vuchic (1981), Sections 2.1.3 and 7.2.1.

passengers can board and alight TUs at any point along the line (distance s), boardings are plotted as a continuous density function $b(s)$, and alightings as $a(s)$; both are in persons per unit distance (usually kilometers, km) per unit time (hour, h). These functions are shown in Figure 1.2(a). From this diagram the cumulative (distribution) functions of boarding and alighting, $B(s)$ and $A(s)$, respectively, are derived and plotted in Figure 1.2(b). Naturally, $B(L) = A(L)$.

$$B(s) = \int_0^s b(s)\,ds \quad \text{and} \quad A(s) = \int_0^s a(s)\,ds$$

$$\left|\frac{B,\,A}{\text{prs/h}}\right|\frac{s}{\text{km}}\left|\frac{b,\,a}{\text{prs/km/h}}\right| \tag{1.3}$$

A passenger volume diagram along the line is then derived mathematically as:

$$P(s) = \int_0^s b(s)\,ds - \int_0^s a(s)\,ds = B(s) - A(s)$$

$$\left|\frac{P}{\text{prs/h}}\right|\frac{s}{\text{km}}\left|\frac{b,\,a}{\text{prs/km/h}}\right|\frac{B,\,A}{\text{prs/h}}\right| \tag{1.4}$$

and it is shown graphically, as in Figure 1.2(b) and (c). The extremes (maximum and minimum points) of the function $p(s)$ occur wherever

$$\frac{dp}{ds} = 0, \text{ so that } b(s) = a(s) \tag{1.5}$$

i.e., wherever the boarding and alighting functions intersect. If there are several such points, as in Figure 1.2(c), the values of all maximum points must be compared to find the maximum load for the line $P_{\max}$, which is of critical importance for operations planning and scheduling.

Translating this theoretical approach to practice, in most cases passengers can board and alight only at transit stops or stations, i.e., at discrete points along the line. Therefore, passenger demand is computed and plotted for a set of existing (or planned) stations. A set of three diagrams with such discrete values, corresponding to the diagrams in Figure 1.2, is shown in Figure 1.3. Boarding and alighting passenger volumes at any station i are now discrete values shown as b_i and a_i bars, respectively; their cumulative functions are B_i and A_i, and the passenger volume (actual or forecasted) on a given section k is

$$P_k = B_k - A_k = \sum_{i=1}^{k} b_i - \sum_{i=2}^{k} a_i$$

$$\left|\frac{P}{\text{prs/h}}\right|\frac{B,\,A}{\text{prs/h}}\left|\frac{b,\,a}{\text{prs/h}}\right| \tag{1.6}$$

as shown in Figure 1.3(b) and (c). Naturally, total numbers of boardings, alightings and passengers using the line per unit time must be equal: $B_L = A_L = P_L$.

The maximum passenger volume for any section along the line, $P_{\max}$, is the most important factor for planning the required transporting capacity on the line. The section on which that volume is found is called the *maximum load section* (MLS).

Temporal variations of demand and their handling in the scheduling process will be discussed in Section 1.3.

1.1.4 Operating Elements: Headway and Frequency

The basic process of transportation can be defined as the movement of objects u over a distance s during an interval of time t. As described in Vuchic (1981), various relationships of these three elements constitute most of the basic operating elements of transportation systems. The diagram in Figure 1.4 shows these relationships.

In transit operations, the most commonly used among these elements are *headway, frequency, capacity, work, productivity, travel time* and *speed*. Headway and frequency, which define movement of TUs on transit line, are defined here; the other elements are defined in the following sections.

Headway h is the time interval between the moments two successive TUs pass a fixed point on a transit line in the same direction. In scheduling it is expressed in minutes, while in capacity analysis seconds are often used.

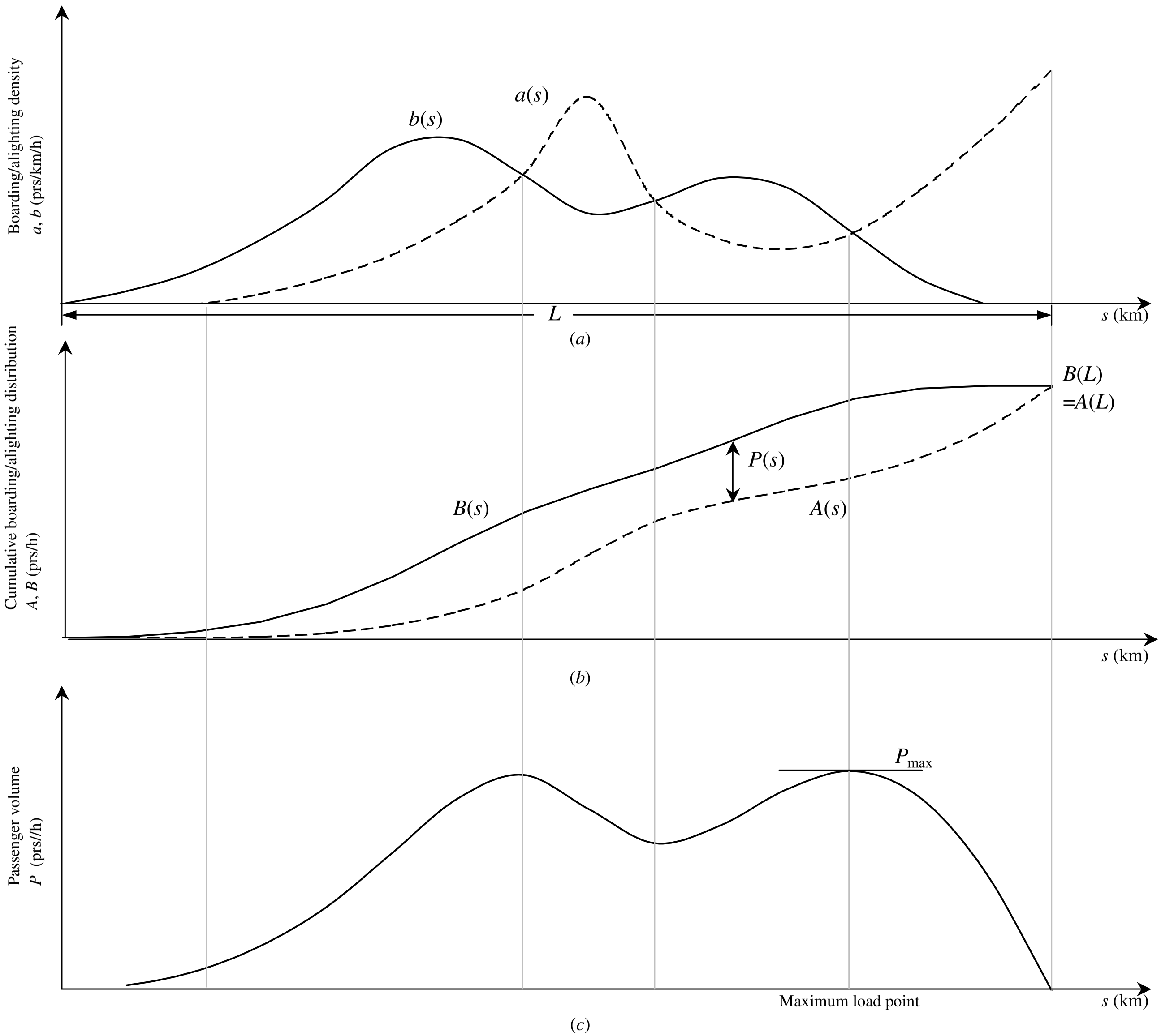


Figure 1.2 *Passenger boarding, alighting, and volume along an idealized transit line: (a) boarding and alighting distributions; (b) cumulative functions of boarding and alighting; (c) passenger volume profile.*

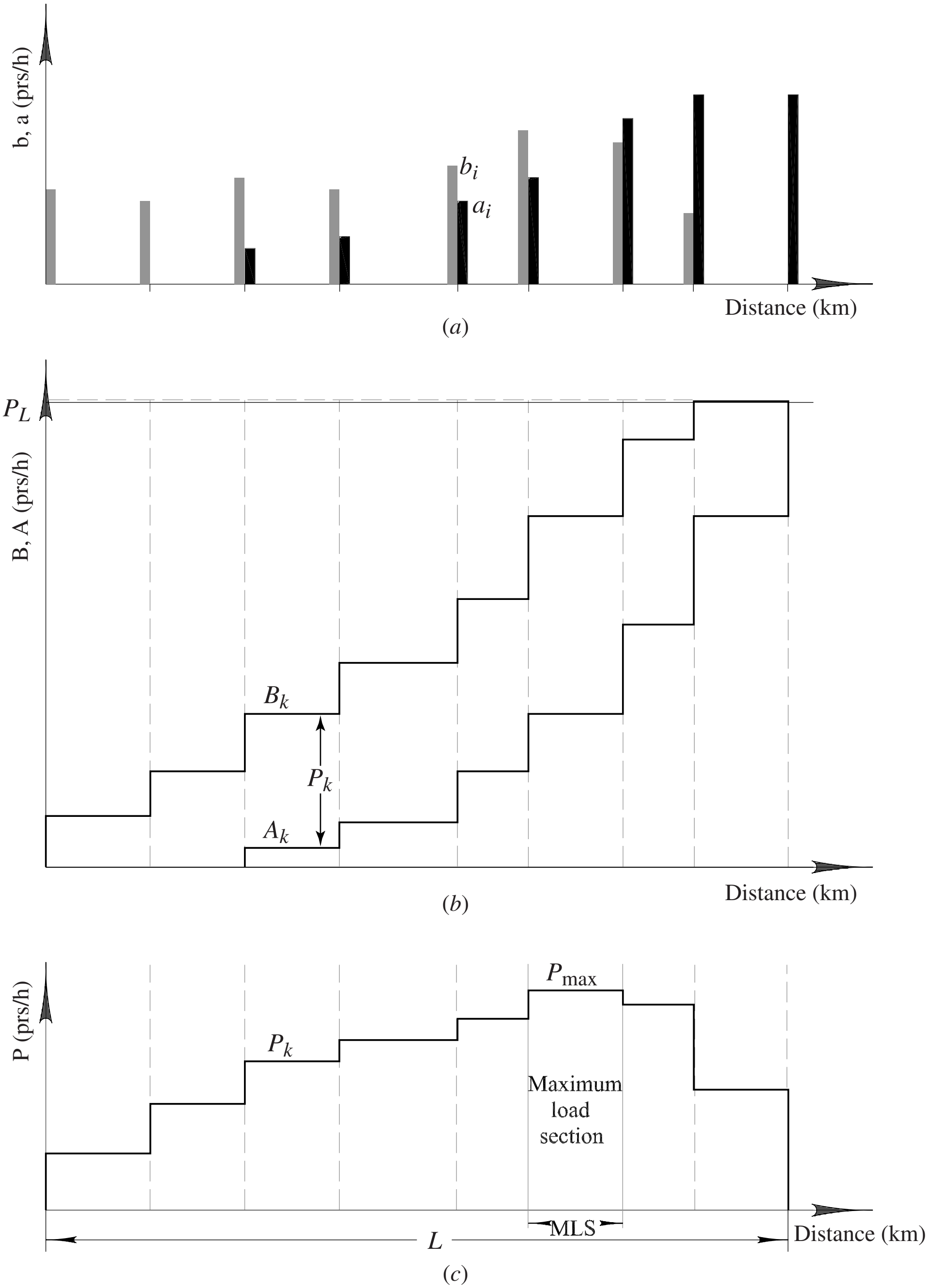


Figure 1.3 *Passenger boarding, alighting, and volume diagram for a transit line:* (*a*) *boarding and alighting volumes at stations;* (*b*) *cumulative boardings and alightings along the line;* (*c*) *passenger volumes transported along the line.*

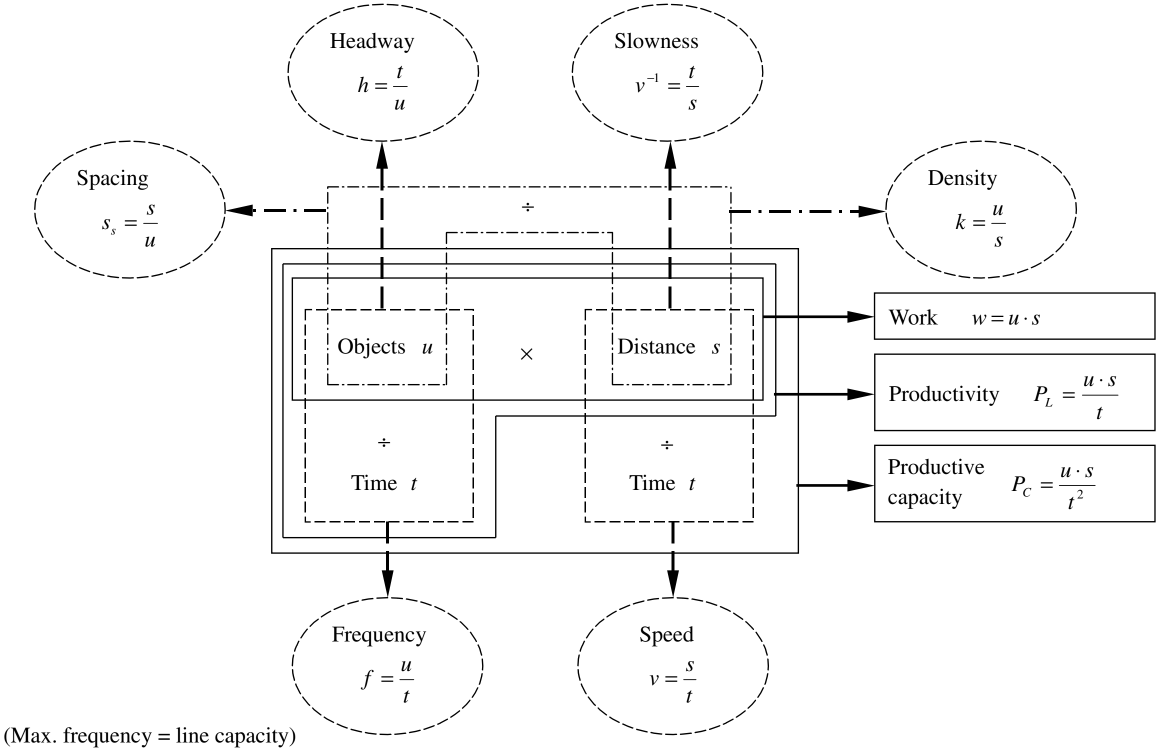


Figure 1.4 *Basic transit system operating elements and performance measures.*

Passengers are interested in having short headways to minimize their waiting time. However, since for any given passenger volume it is cheaper to operate a small number of large TUs than a large number of small TUs, a transit agency is usually interested in operating long headways. Consequently, headways are usually determined as a compromise between passenger travel time and convenience, and the cost of operation.

Whenever passenger demand is reasonably stable, uniform (constant) headways provide the most efficient operation (even loading of TUs and maximum schedule stability), and they are most attractive for passengers (simple, reliable, waiting time minimized). When headways are longer than 6 minutes, it is desirable to use only values divisible into 60 (7.5, 10, 12, 15, 20, 30 or 60 minutes), known as *clock headways,* because with them departure times at any one stop fall on the same minutes in each hour, so that passengers can easily memorize the schedule.

The longest headway scheduled for a line is determined by the minimum level of service considered acceptable for that line; it is therefore called *policy headway* h_p.

The *minimum headway* $h_{\min}$ on a line is, on the other hand, determined by the physical characteristics of the system (technology, method of driving and control, degree of safety required) and station operations (rate of boarding/alighting, departure control, etc.). The former factors influence the shortest headway achievable on open line between stops, *way headway* $h_{w\,\min}$, and the latter determine the shortest headway that can be operated at stations, *station headway* $h_{s\,\min}$. The greater of the two represents the minimum headway for the line:

Table 1.1 Capacity, work, and utilization concepts related to transit line operation

Category	*Definition*	*Derivation Formula*	*Units*		*Utilization Coefficient*
			Offered	*Utilized*	
Static capacity	Vehicle capacity (seats only or seats + standing spaces)	C_v (given)	C_v (sps/veh)	P_v (prs)	$\alpha_v = P_v/C_v$
Dynamic capacity	Maximum frequency	f_{max}	f_{max} (TU/h)	f (TU/h)	$\delta_f = f/f_{max}$
	Vehicle line capacity	$f_{max} \cdot n$	c (veh/h)	$f \cdot n$ (veh/h)	$\delta_v = f \cdot n/c$
	Line capacity, min (C_w, C_s)*	$f_{max} \cdot n \cdot C_v$	C (sps/h)	P (prs/h)	$\alpha = P/C$
	Scheduled capacity	$f \cdot n \cdot C_v$	C_o (sps/h)	P (prs/h)	$\alpha = P/C$ $\delta = C_o/C$
Transportation work	Work on a line during one hour	$f \cdot n \cdot C_v \cdot L$	w (sp-km/h)	w_p (prs-km/h)	$\overline{\alpha} = w_p/w$
Productive capacity	Product of capacity and speed	$f \cdot n \cdot C_v \cdot V_o$	P_c (sp-km/h^2)	(not used)	(not used)

* Way capacity C_w and station capacity C_s are defined in Section 2.1.

$$h_{min} = \text{Max}\ (h_{w\,min}, h_{s\,min}) \qquad (1.7)$$

In the vast majority of cases, $h_{s\,min} >> h_{w\,min}$, and the minimum headway at the station with the heaviest exchange (boarding and alighting) of passengers, i.e., the longest minimum headway among all the stations on the line, represents the minimum headway that can be achieved on the line.

The number of TUs passing a point on a transit line in one direction during one hour (or some other time interval) represents *frequency of service f*. Thus, frequency is inverse of the headway:

$$f = \frac{60}{h} \qquad \left|\frac{f}{\text{TU/h}}\right|\frac{h}{\text{min/TU}}\right| \qquad (1.8)$$

and it is one of the components of quantity of offered service.

The concepts of headway and frequency are often confused not only by the lay public, but by transit operators. For example, many schedules state that a transit line has a "Frequency: 10 minutes." Actually, 10 minutes is the headway. Frequency is its inverse, and in this case it is six departures per hour.

1.1.5 Capacity, Work, and Utilization

Capacity of a system or facility in a broad sense refers to its maximum ability to perform under prevailing conditions. For public transport systems two different capacities are particularly important: *vehicle capacity,* expressed in spaces per vehicle, and *transit line capacity,* with the dimension of spaces per hour. With these dimensions they represent *offered capacities C* (sps/h). Maximum number of passengers, or flow of passengers per hour that are actually transported, represents the demand or *utilized capacity P* (prs/h). The ratio of utilized to offered capacity is the *capacity utilization coefficient* α, also known as *load factor,* with the dimension of persons transported divided by spaces offered:

$$\alpha = \frac{P}{C} \qquad \left|\frac{\alpha}{\text{prs/sp}}\right|\frac{P}{\text{prs/h}}\left|\frac{C}{\text{sps/h}}\right| \qquad (1.9)$$

These different capacity concepts, as well as transportation work and productive capacity of a transit line, are listed with their formulae, units, and utilization coefficients in Table 1.1. Each will be briefly defined here in sequence.

1.1.5.1 Vehicle Capacity. Vehicle capacity C_v is the maximum number of spaces for passengers a vehicle (or TU) can accommodate. This "static capacity" is expressed in spaces, and it can be computed in three different ways: *seats plus standing spaces, seats only,* and *ratio of seats to standing spaces*.

Seats plus standing spaces is the capacity definition used for high-volume rail and bus services. This capacity depends on the standard used for floor area per standee (for detailed discussion of passenger level of service as a function of area per standee, see Fruin, 1971). For capacity computations in different transit systems and adopted comfort levels, any one of the values shown in Table 1.2 can be used. In most cities in industrialized countries, the standard of 0.25 or 0.20 m^2/standee is used. In a few very large cities (Mexico, Moscow, New York, Tokyo) and in most cities in developing countries, heavy crowding of vehicles is common, so that the value of 0.15 m^2/standee might be more applicable.

With this definition of capacity, the load factor α cannot exceed 1.00 (unless assumed standing comfort standards are exceeded), and its value clearly shows how close the volume is to capacity.

Seats only; this definition, known as an exact number, is sometimes used for transit services designed to provide seats for all passengers, such as many regional rail and high-quality bus systems. However, if these systems often carry loads exceeding seating capacity, this definition yields load factors greater than 1, which give no indication how full the vehicles actually are and how much reserve capacity is available. For example, a load factor of 2.0 on the San Francisco Bay Area Rapid Transit (BART) cars, which have large standing areas, results in a higher comfort and has more reserve capacity than load factor of 1.5 on the regional rail cars with high-density seating, such as the Regional Rail cars in Philadelphia.

The ratio of seats to standing spaces (e.g., 40:60) is sometimes used as a goal for comfort standard in initial design of a vehicle interior. This measure is inconvenient because it does not describe the comfort (density) of standees and cannot be controlled in operations.

In summary, on transit systems which sometimes carry standees, total vehicle capacity rather than seats only should be used.

Average number of passengers per vehicle is used as a measure of the capacity of semipublic and private modes (such as van pools and private automobiles, respectively) and for the public mode used individually—the taxi. Utilization of these vehicles is determined by their owners or users and is independent of demand for travel on a given facility in the short run. Thus, their physical capacity is often much greater than the number of persons they carry, but the empty seats cannot be utilized in private transportation modes.

TU capacity C_{TU}, is simply $n \times C_v$, where C_v is the respective vehicle capacity from the three definitions given above.

1.1.5.2 Line Capacity.

Line (offered) capacity is the maximum number of spaces that can be transported past a fixed point in one direction during one hour. This "dynamic capacity" is expressed in units per hour, and several different units can be used:

Transit units per hour represents the *maximum frequency* f_{max} that can be physically achieved on a line under given conditions (speed, safety, station operations, signal system, etc.).

Table 1.2 Passenger comfort as a function of floor area per standee

Density of Persons prs/m²	*Area per Standee m²/prs*	*Area per Standee ft²/prs*	*Standing Passengers' Condition*
<1	>1.00	>10.8	Independent standing, easy circulation
2–3	0.50–0.33	5.4–3.6	Some body contacts, circulation disturbing others
4	0.25	2.7	Extensive body contacts, difficult movements
5	0.20	2.2	Pressed standing, extremely difficult movements
6.7	0.15	1.6	Crash loads, possible injuries, forced movements

Area conversion: 1 m^2 = 10.76 ft^2.

Vehicles per hour, designated *vehicle line capacity* c, defines the maximum number of vehicles that can pass a fixed point on a line in one direction during one hour, regardless of whether as single units or in trains: $c = n \times f_{\max}$. Obviously, for single-vehicle operation $c = f_{\max}$, but for train operation $n > 1$, so that $c > f_{\max}$.

Spaces per hour is the line capacity C, most frequently used in transit mode description, planning, and scheduling. It represents the actual capacity offered to passengers in spaces transported past a fixed point in one direction during one hour.

All three capacities are interrelated and are functions of a line's minimum headway:

$$C = C_v \cdot c = C_v \cdot n \cdot f_{\max} = \frac{60 C_v \cdot n}{h_{\min}}$$
$$\left|\frac{C}{\text{sps/h}}\right|\frac{C_v}{\text{sps/veh}}\left|\frac{c}{\text{veh/h}}\right|\frac{n}{\text{veh/TU}}$$
$$\left|\frac{f}{\text{TU/h}}\right|\frac{h}{\text{min/TU}}\right| \tag{1.10}$$

Scheduled line capacity C_o is the number of spaces transported past a fixed point in one direction during one hour under a given operating schedule. Ratios of scheduled to absolute capacities of a line represent utilization coefficients of service frequency δ_f, vehicle line capacity δ_v, and scheduled line capacity δ, respectively:

$$\delta_f = f/f_{\max};\ \delta_v = n \cdot f/c; \quad \text{and} \quad \delta = C_o/C$$
$$\left|\frac{\delta}{-}\right|\frac{f}{\text{TU/h}}\left|\frac{n}{\text{veh/TU}}\right|\frac{c}{\text{veh/h}}\left|\frac{C}{\text{sps/h}}\right| \tag{1.11}$$

These coefficients are particularly useful for analyses of heavily used transit lines.

Transportation work w performed on a transit line represents its output or quantity of offered or utilized service. The offered work w_o can be expressed in train-, vehicle-, or space-km. When all TUs operate on the entire length L of a line, the work performed during one hour is:

$$w_o = C \cdot L = f \cdot n \cdot C_v \cdot L \quad \left|\frac{w}{\text{sp-km(/h)}}\right|$$
$$\left|\frac{C}{\text{sps/h}}\right|\frac{L}{\text{km}}\left|\frac{f}{\text{TU}/h}\right|\frac{n}{\text{veh/TU}}\left|\frac{C_v}{\text{sps/veh}}\right| \tag{1.12}$$

Passenger-km traveled on the line represent *utilized work* w_p; it is computed for one hour as:

$$w_p = \sum_i p_i \cdot S_i \quad \left|\frac{w_p}{\text{prs-km(/h)}}\right|\frac{p}{\text{prs/h}}\left|\frac{S}{\text{km}}\right| \tag{1.13}$$

where p_i is passenger volume on section i of the line, and S_i length of that section.

Work utilization coefficient $\bar{\alpha}$ is the ratio of utilized to offered work:

$$\bar{\alpha} = \frac{w_p}{w_o} = \frac{\sum_i p_i \cdot S_i}{C \cdot L} \quad \left|\frac{\bar{\alpha}}{\text{prs-km/sp-km}}\right.$$
$$\left|\frac{p}{\text{prs/h}}\right|\frac{S}{\text{km}}\left|\frac{C}{\text{sps/h}}\right|\frac{L}{\text{km}}\right| \tag{1.14}$$

This ratio represents the average utilization of offered capacity along the line, or the average value of the load factor weighted by the passenger volume.

Figure 1.5 shows some of the above-defined concepts: a transit line with passenger volume (as shaded area) and offered capacity. The load factor α varies along the line, reaching its maximum value on the maximum load section (MLS), where it is $P_{\max}/C$. The coefficient $\bar{\alpha}$ cannot be plotted numerically, but only visually: it represents the ratio of the shaded area to the area under the C line.

Productive capacity P_c, the product of line capacity and its operating speed, is one of the most useful quantitative indicators of line performance, since it incorporates capacity, which concerns the operator, and speed, which affects both passengers and the operator. Its detailed description and use for comparison of

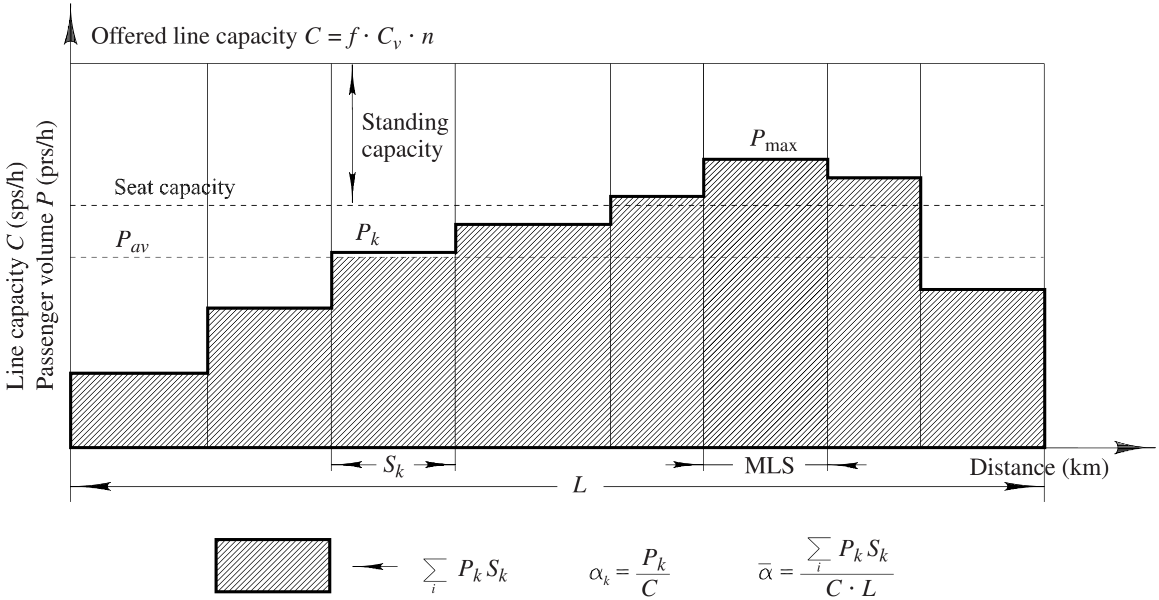


Figure 1.5 *Line capacity, passenger volume, and utilization factors.*

different modes is given in Vuchic (1981), Sections 2.3 and 7.6.

1.1.6 Travel Times

The durations of individual time intervals in transit system operation or in passenger travel are referred to as *travel times*. Many different intervals can be defined. Individual travel time components will be designated by t, the intervals consisting of several elements by T, while subscripts will indicate specific intervals. Transit travel times are usually expressed in minutes.

1.1.6.1 On-Line Travel Times. Several different travel times of TUs on a line can be defined.

Running time t_r is the time interval between a TU's starting from one station and stopping at the next one, i.e., the net travel time between stations.

Station standing (*or dwell*) *time* t_s is the duration of a TU's standing at a station for the purpose of boarding and alighting of passengers. Definitions of elements influencing t_s and its values for different modes are given in Vuchic (1981), Section 7.4.4.2.

Station-to-station travel time T_s is the time interval between a TU's departures from two adjacent stations, i.e., it is equal to running time t_r plus station standing time t_s on any spacing i:

$$T_{si} = t_{ri} + t_{si} \qquad \left| \frac{T,\ t}{\text{min or s}} \right| \tag{1.15}$$

This time interval is thus the basic module of TU travel along the line.

On lines where there are no geometric limitations or interferences by other traffic, such as on some rapid transit lines, station-to-station travel time is expressed by two different equations (derived in Vuchic, 1981, Section 3.6.3). For station spacings S', so short that TUs cannot reach their maximum speed, travel consists of only acceleration, braking and standing intervals t_a, t_b, and t_s, respectively. Travel time T_s' is expressed by the following formula:

$$T'_s = t_a + t_b + t_s = \sqrt{\frac{2(\overline{\alpha} + \overline{b})\, S'}{\overline{a}\overline{b}}} + t_s$$

$$\left|\frac{T,\ t}{\mathrm{s}}\right|\frac{a,\ b}{\mathrm{m/s^2}}\left|\frac{S}{\mathrm{m}}\right| \tag{1.16}$$

where $\overline{a}$ and $\overline{b}$ are average acceleration and deceleration rates, respectively. For spacings on which the maximum speed of TUs, $V_{\max}$, can be reached, the equation is:

$$T_s = \frac{3.6S}{V_{\max}} + T_\ell \qquad \left|\frac{T}{\mathrm{s}}\right|\frac{S}{\mathrm{m}}\left|\frac{V}{\mathrm{km/h}}\right| \tag{1.17}$$

where T_ℓ is the incremental time loss per stopping at one station, computed as:

$$T_\ell = \frac{V_{\max}}{7.2}\left(\frac{1}{\overline{a}} + \frac{1}{\overline{b}}\right) + t_s \qquad \left|\frac{T,\ t}{\mathrm{s}}\right|\frac{V}{\mathrm{km/h}}\left|\frac{a,\ b}{\mathrm{m/s^2}}\right| \tag{1.18}$$

using the approximation that both acceleration and deceleration rates are constant.

Operating (*or travel*) *time* T_o is the scheduled time interval between departure of a TU from one terminal and its arrival at the other terminal on the line. The operating time is therefore the sum of station-to-station travel times for all i interstation spacings between terminals:

$$T_o = \sum_i T_{si} = \sum_i (t_{ri} + t_{si}) \qquad \left|\frac{T,\ t}{\mathrm{min\ or\ s}}\right| \tag{1.19}$$

Strictly speaking, there should be $(i - 1)$ t_s time intervals if standing at terminals is included in terminal times (see below). However, the definition given by Eq. (1.19) is simpler, and yet correct, if it is assumed that a portion of terminal times at each terminal equivalent to $t_s/2$ is included in operating time. Since terminal times are usually much longer than t_s (several minutes versus 20–30 seconds), they are practically unaffected by this definition. In the cases where passenger boarding (or alighting) time at a terminal is very long, these definitions may have to be modified.

Terminal time t_t is the time a TU spends at a line terminal (strictly, after a t_s time or two $t_s/2$ times is subtracted). This time is provided for some or all of the following purposes:

- Vehicle turning or change of driver's cab
- Resting of the crew
- Adjustment in schedule (e.g., to maintain uniform headway)
- Recovery of delays incurred in travel

Transit agency work standards, and usually labor contracts, stipulate that a certain percentage of working time must be provided for a driver's rest or meal break. The need for schedule recovery is also related to line length, since the probability of delays and their propagation increases with the length. Total terminal time (at both terminals, $t'_t + t''_t$) is therefore usually determined as a percent of operating time on the line, γ:

$$\gamma = \frac{t'_t + t''_t}{2T_o} \times 100\% \qquad \left|\frac{\gamma}{\%}\right|\frac{T,\ t}{\mathrm{min}}\right| \tag{1.20}$$

The value of γ usually varies between 10 and 30% (i.e., the sum of the two terminal times represents 10–30% of total, two-way operating time on the line), the typical average being around 15%. Exceptions to this relationship exist in the following cases.

Modes with high reliability of operation, such as rail rapid transit and regional rail, can have fixed terminal times regardless of line length. If that time is short, drivers are exchanged: each driver falls back, i.e., leaves his train and takes over a later one, after his break for rest. The extreme case of this is found on circle lines, which often operate without any terminal time.

On lines with long uniform headways, cycle times (defined below), being integer multiple of headways, must sometimes include long terminal times, resulting in $\gamma > 30\%$.

Bus and other street transit lines are often scheduled with longer terminal times during peaks than off-peaks, to provide reserve times for schedule recoveries from delays caused by frequent peak hour traffic congestion.

Sometimes terminal times are related to cycle time T, defined below, giving factor γ':

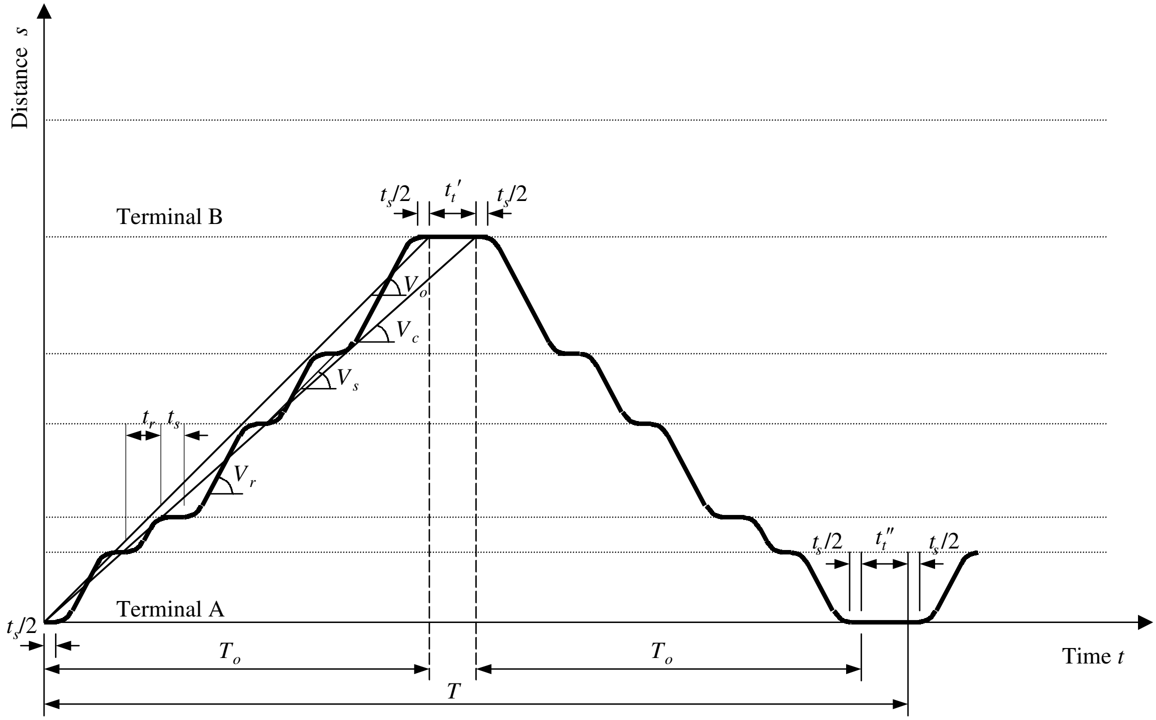


Figure 1.6 *Travel times and speeds on a transit line.*

$$\gamma' = \frac{t_t' + t_t''}{T} \times 100\% \qquad \left|\frac{\gamma}{\%}\right|\frac{T,\ t}{\min}\right| \tag{1.21}$$

Naturally, $\gamma' < \gamma$, except on lines without terminal times, such as most circle lines, where $\gamma' = \gamma = 0$.

Cycle time T is the total round trip time on a line, or the interval between the two consecutive times a TU in regular service leaves the same terminal. It consists of operating times for the two directions, T_o' and T_o'', and terminal times:

$$T = T_o' + T_o'' + t_t' + t_t'' = 2(T_o + t_t) \qquad \left|\frac{T,\ t}{\min}\right| \tag{1.22}$$

The latter expression is for the case when operating times in both directions and terminal times at both terminals are equal, respectively. Cycle time is the basic time unit for scheduling transit service, and it strongly influences investment and operating costs.

All of the preceding travel times are shown graphically on the time-distance diagram of TU travel in Figure 1.6.

Deadhead time t_d is the portion of TU travel time during which the TU is not in passenger (revenue) service. It includes travel from depot to the line and back, or between lines when a TU is reassigned. Deadhead time is not directly productive, and it is therefore desirable to minimize it.

Platform time T_p is the total time a TU is in operation. When a TU makes k round trips on a line and has both deadhead times equal, its platform time is:

$$T_p = kT + 2t_d \qquad \left|\frac{T,\ t}{\min}\right|\frac{k}{-}\right| \tag{1.23}$$

Sometimes platform time is also affected by the waiting times at terminals for schedule adjustments, or, if the TU is scheduled to operate on more than one line, by cycle times on different lines. Platform time represents the net on-duty time for computing labor requirements for transit line operation. To compute total labor time, check-in, check-out, and idle times of drivers must be added to T_p.

1.1.6.2 Passenger Travel Times. The travel of a transit passenger on an origin-destination path, including his/her approach to a transit stop or station, travel on the line, a transfer between lines, and departure from a stop to his/her destination, is shown on the diagram in Figure 1.7. This type of diagram is often used for analysis of passenger travel on individual lines and comparison of passenger travel times by alternative modes. The diagram shows that the passenger has the following time components.

Access time t_a is the time which an individual passenger requires for approach to a transit stop or departure from a stop to his destination for a given trip.

Waiting time t_w is the time between passenger arrival at a stop and the time of TU departure. For frequent transit service the average waiting time is equal to half of the headway. For longer headways (usually $h > 6$ or 10 minutes) passengers begin to use a timetable and adjust their arrivals to the scheduled TU departures, so that the average waiting time becomes somewhat shorter than for random passenger arrivals and remains approximately constant for longer headways, as the diagram in Figure 1.8 shows. An excellent analysis of t_w is given in Bowman and Turnquist (1981).

On-line travel time t_o is the duration of passenger travel in a TU for a given trip.

Transfer time t_f is the time used for transferring between different lines or modes, i.e., the interval between alighting one TU and boarding another. Transfer time depends on the walking time between the two line platforms, on headway of the second line, and on schedule coordination between the lines.

Origin-destination travel time T_{od} is the total passenger's travel time from his point of origin (o) to his

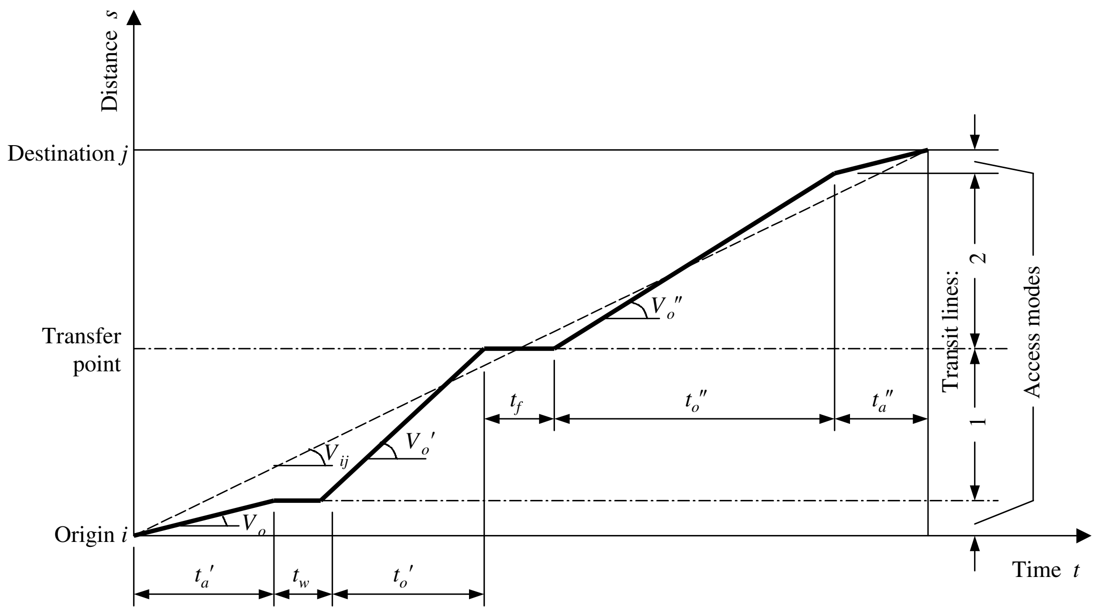


Figure 1.7 *Passenger travel times and speeds.*

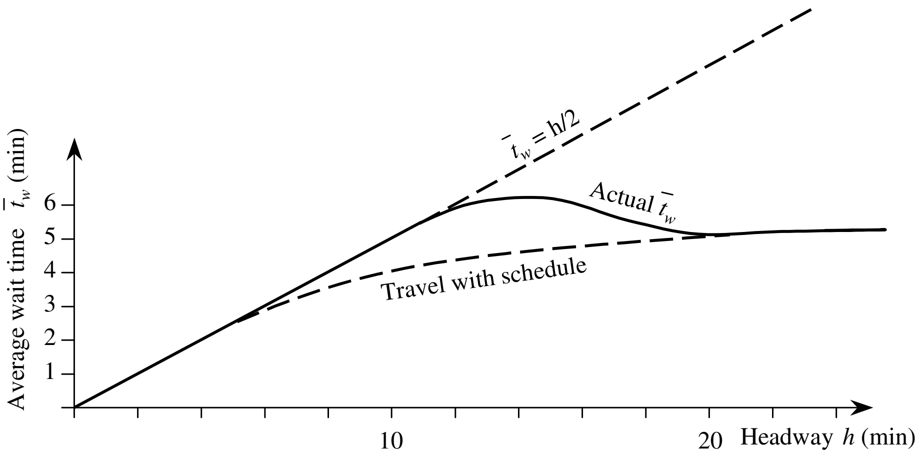


Figure 1.8 *Average passenger waiting time as a function of headway.*

point of destination (*d*). This time may consist of some or all of the following components:

$$T_{od} = t_a + t_w + t_o + t_f \qquad \left|\frac{T, t}{\text{min}}\right| \qquad (1.24)$$

1.1.7 Speeds

A number of different speed definitions are used in transit operations and analyses. They are listed in Table 1.3, classified in four categories. In the following sections each speed is defined and most of them are shown graphically on diagrams. Three types of diagrams are particularly useful for illustrating speeds:

- A time-distance diagram is used for plotting movement of a single TU along a transit line, for schedules, and for street traffic movement with signal control.
- A time-speed diagram conveniently shows vehicle performance, travel regimes, energy consumption, and travel time for station-to-station movement.
- A distance-speed diagram is best suited for plotting of speed profiles along a transit line and analyses of its different sections, braking and stopping distances, signal blocks and their locations, etc.

1.1.7.1 Vehicle Speeds. The variable of *actual vehicle speed* on a transit line is designated as V (km/h) or v (m/s).

Maximum technical speed V_{max} is the highest speed a transit vehicle is physically capable of achieving on a straight horizontal way under normal weather conditions when its maximum power is applied and acceleration has gradually ceased.

High technical speed of vehicles is usually desirable for achieving better performance of the line; however, speed of transit vehicle travel is often limited by traffic conditions and spacings between stops, so these should be carefully analyzed in determining the maximum speed vehicles should have for a given type of service.

1.1.7.2 Alignment Speeds. In connection with design and operation of transit lines, the following three speeds can be defined.

Line design speed V_d is the maximum speed transit vehicles can achieve on a given section of line with adequate comfort and safety when physical conditions govern. It may vary among line sections and directions, depending on geometry of alignment, gradient and other factors. Average line design speed for an entire line is:

Table 1.3 Transit speeds and corresponding travel times

Category	*Speed Designation*	*Symbol*	*Corresponding Travel Time*
Vehicle speed	1. Actual vehicle	V, v	—
	2. Maximum technical	V_{max}	—
Alignment speeds	3. Line design	V_d	—
	4. Legal	V_ℓ	—
	5. Programmed	V_g	—
Vehicle-on-line speeds	6. Running	V_r	t_r
	7. Station-to-station	V_s	T_s
	8. Operating	V_o	T_o
	9. Cycle	V_c	T
	10. Platform	V_p	T_p
Passenger speeds	11. Access	V_a	t_a
	12. Travel on line	V_o	T_o
	13. Origin-destination	V_{od}	T_{od}

$$\overline{V}_d = \frac{\sum_i S_i}{\sum_i \frac{S_i}{V_i}} = \frac{L}{\sum_i \frac{S_i}{V_i}} \qquad \left| \frac{V}{\text{km/h}} \right| \frac{S, L}{\text{km}} \right| \tag{1.25}$$

where S_i is length of section i and V_i design speed on it.

Legal speed V_ℓ is the maximum speed transit vehicles can legally operate on a given section of line. Since the legal speed is determined on the basis of various limitations (design speed, traffic and environmental conditions, etc.), it can be equal to or smaller than the design speed. Legal speed is less permanent than design speed: it can be changed with traffic conditions, between day and night, with improved control devices, etc.

Programmed speed V_g is the speed transit vehicles can operate meeting given standards of safety, comfort, economy, and vehicle performance. Average programmed speed for a line cannot be computed as precisely as design and legal speeds because its changes among sections depend on vehicle performance and various influences, while the former ones change instantly. Programmed speed is the speed that transit vehicles can actually achieve: rapid transit systems with automatic train operation (ATO) can operate it precisely; manually driven TUs can follow it as closely as drivers' skills allow.

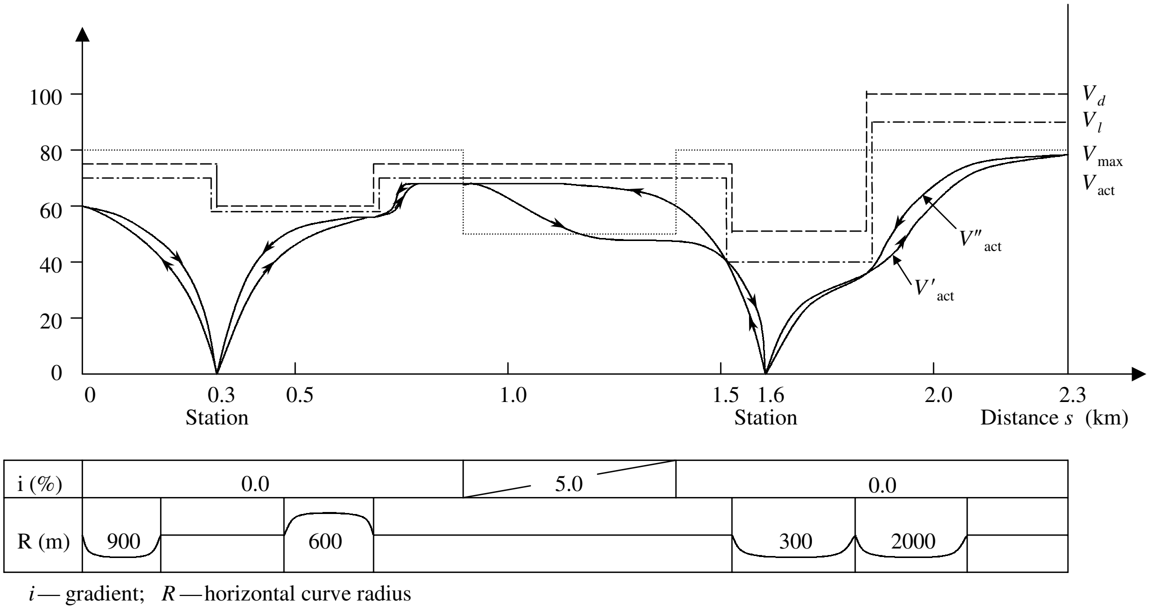
None of the speeds defined so far include transit stops. They refer to physical conditions of TUs running along the line, not considering requirements to stop and related time delays. All these speeds are shown on a distance-speed diagram in Figure 1.9, which illustrates the following facts:

- V_{max}, as an element of vehicle performance, is affected by gradient i, but not by V_d and V_ℓ.
- $V_g \leq V_\ell \leq V_d$.
- Alignment gradient, as well as different street conditions on directionally split lines, may result in different V_g's for the two directions.

1.1.7.3 Vehicle-on-Line Speeds. Actual speeds achieved on transit lines are influenced by vehicle and alignment speeds, as well as by stoppings at passenger stops and general traffic conditions (rapid transit is not affected by the latter ones). Five different speeds can be defined for transit line operations.

Running speed V_r is the average speed TUs achieve from leaving one station (stop) to arriving at the next one. On a spacing between stations S_i, this speed is:

$$V_{ri} = \frac{60 S_i}{t_{ri}} \qquad \left| \frac{V}{\text{km/h}} \right| \frac{S}{\text{km}} \left| \frac{t}{\text{min}} \right| \tag{1.26}$$



Different types of speed given in km/h

V_{max}	80				50		80			
V_d	75	60		75				50		100
V_l	70	60		70				40		90
V_{act}	70	60	60	70	70	50	40		40	80

Figure 1.9 *An illustration of relationships among vehicle and alignment speeds on a transit line.*

Thus, running speed varies among spacings and is usually analyzed for individual spacings, rather than as an average for line.

Station-to-station speed V_s is the average speed of travel between moments a TU leaves two adjacent stations; it includes running and one station dwell time:

$$V_{si} = \frac{60S_i}{t_{ri} + t_{si}} = \frac{60S_i}{T_{si}} \qquad \left| \frac{V}{\text{km/h}} \middle| \frac{S}{\text{km}} \middle| \frac{t,\, T}{\text{min}} \right| \tag{1.27}$$

Consequently, this is the speed for the basic module of travel along a line, movement along one-station spacing. When these speeds are averaged along the line, one obtains the next speed definition—operating speed.

Operating or *travel speed* V_o is the average speed of TU travel along transit line with j spacings (or on a section of it):

$$V_o = \frac{60\sum_{i=1}^{j} S_i}{\sum_{i=1}^{j} T_{si}} = \frac{60L}{T_o} = \frac{120L}{T'_o + T''_o}$$

$$\left| \frac{V}{\text{km/h}} \middle| \frac{S,\, L}{\text{km}} \middle| \frac{t,\, T}{\text{min}} \right| \tag{1.28}$$

Operating speed is the speed of travel offered to the public; it is therefore one of the basic elements of offered transit service performance.

Cycle speed V_c is the average speed of a TU for a complete round trip on a line:

$$V_c = \frac{60 \cdot 2L}{T} = \frac{120L}{T} \qquad \left| \frac{V}{\text{km/h}} \middle| \frac{L}{\text{km}} \middle| \frac{T}{\text{min}} \right| \tag{1.29}$$

This speed is the most important one for the operator, since it directly influences the number of TUs required for a given level of service, and thus the transit line's capital and operating costs.

All vehicle-on-line speeds defined here are shown on a time-distance diagram in Figure 1.6. Clearly, $V_{ri} > V_{si}$, for every spacing i, but both vary among spacings, while $V_o > V_c$ is true for all cases, except for lines without terminal times (such as circle lines), on which $V_o = V_c$.

Platform speed V_p is the average speed of TUs operating on a line from the time they leave depot (garage or yard) until they return to it, i.e., during the platform time. In the case of simple, regular scheduling, platform speed can be computed as:

$$V_p = \frac{120(k \cdot L + L_d)}{k \cdot T + 2t_d} = \frac{60\ L_p}{T_p} \qquad \left| \frac{k}{-} \middle| \frac{L}{\text{km}} \middle| \frac{T,\ t}{\text{min}} \right| \tag{1.30}$$

where k is again the number of round trips performed, L_d is deadhead distance, and L_p is the total distance the TU travels while absent from the depot. Platform speed thus includes two deadheading trips (between the depot and the line) and a number of round trips on the line, so that it is influenced by the location of depot in relation to the line and various schedule requirements. It is therefore used for measuring efficiency of vehicle deployment on individual lines.

1.1.7.4 Passenger Speeds. Transit passengers experience several different speeds while they travel.

Access speed V_a is the average speed of passenger travel to and from transit stops or stations. Computed as access distance divided by access time, this speed can vary greatly: from 4–5 km/h for walking to 30–50 km/h for access by automobile. The same trip may have different access speeds for travel to and from a transit line, depending on access mode used. Access speed is important for planning and analyses of transit network layout and spacings of stations.

Travel speed on line V_o that a passenger experiences is actually the operating speed on the line section that he/she utilizes.

Origin-destination speed V_{od} is the average speed of passenger travel along his/her path from origin to destination, including access, waiting, on-line travel, and transfers, if any. For a total distance past between origin and destination S_{od}, it is:

$$V_{od} = \frac{60\ S_{od}}{T_{od}} \qquad \left| \frac{V}{\text{km/h}} \middle| \frac{S}{\text{km}} \middle| \frac{T}{\text{min}} \right| \tag{1.31}$$

This speed is an element of mobility (or ease of travel) by transit, and it influences modal split.

With all travel times and speeds defined, it is now possible to present a convenient graphical method for computation of passenger origin-destination travel time. The diagram for this computation, presented in Figure 1.10, is obtained in the following way.

The upper right quadrant represents a distance-time diagram of transit line operation: distance is on the abscissa, time on the ordinate, and a family of lines is plotted with slopes representing decreasing operating speeds (40–20 km/h). The abscissa from the origin to the left represents travel time on the line plus waiting time. This time is plotted at the same scale as the ordinate, so that a line at a 45° slope represents travel time without waiting ($t_w = 0$). Lines representing travel with different waiting times are plotted parallel to it toward the left. Assuming that the average waiting time is half of the headway on the line for headways up to 10 minutes and then remain constant for all headways longer than that, the lines for $h = 3$, 6, and 10 minutes are plotted at a horizontal distance of 1.5, 3, and 5 minutes, respectively, from the first (45°) line.

Finally, the total travel time T_{od} is plotted on the ordinate downward from the origin, usually at a smaller scale than travel time on the abscissa. A straight line from the origin through the intersections of the same values on the abscissa and ordinate in the third quadrant is plotted for travel time without access

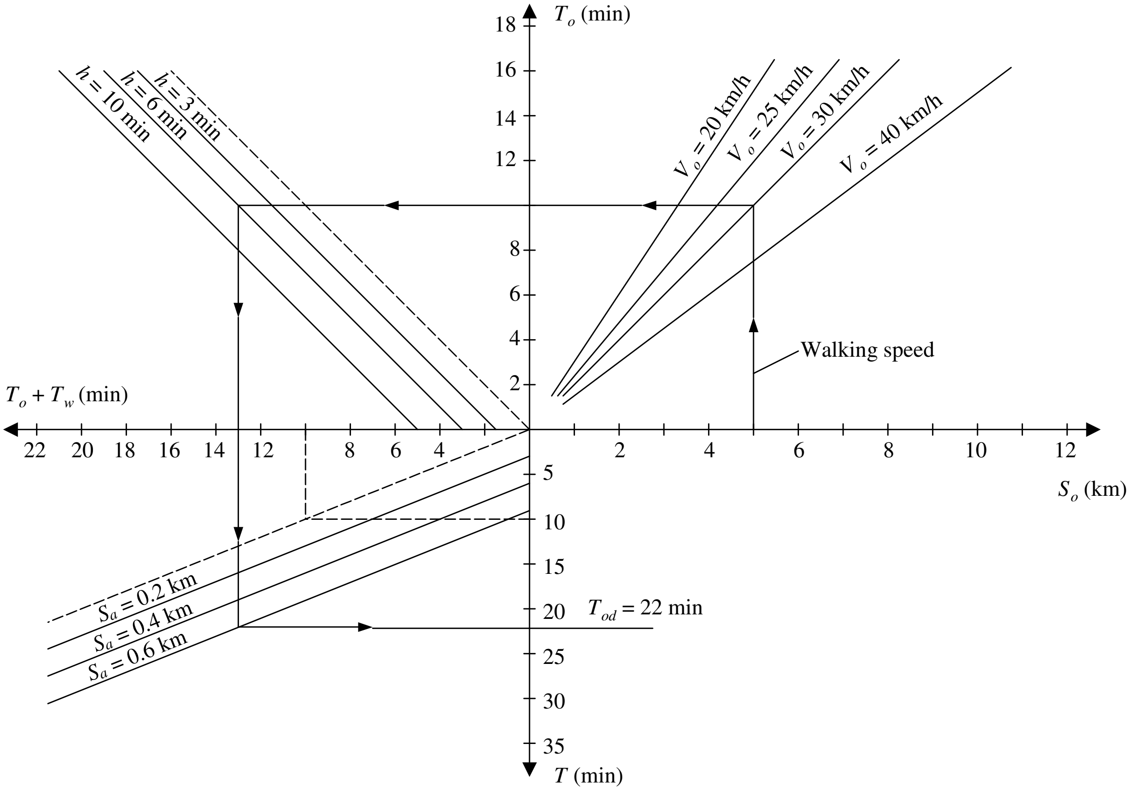


Figure 1.10 *Diagram for computation of passenger origin-destination travel time by transit.*

time ($t_a = 0$), and a family of parallel lines is plotted below it at a distance of travel time for various access distances (representing the sum of origin-to-station and station-to-destination distances) by a given mode. In Figure 1.10, the lines are plotted for access by walking speed of 4 km/h. If it is assumed that average bicycle and automobile access speeds are 20 and 40 km/h, these same lines will represent, respectively, 5 and 10 times greater distances than those shown.

The use of this diagram is very simple. As the plotted example shows, the derivation starts from the distances traveled on the transit line on the abscissa and then continues through the three quadrants for a given set of parameters, to the resulting origin-destination travel time on the downward ordinate. The values assumed in the plotted example are: $S_o = 5$ km, $V_o =$ 30 km/h, $h = 6$ min, and $S_a = 0.6$ km, giving $T_{od} =$ 22 min.

Naturally, the scales and values of parameters on the diagram may vary among different cases; the most convenient ones for a specific case should be chosen. Where appropriate, transfer times should be added to the computed times.

1.2 INFORMATION FILES AND DATA COLLECTION: SURVEYS AND COUNTS

Effective planning of transit operations, scheduling and analysis of efficiency, and sound management in general require accurate data about operations and usage of transit lines and networks. Transit agencies must therefore organize, maintain, and regularly update systematic files of information and data. These files should include:

- *Physical inventories of lines and facilities,* such as locations and equipment of bus stops on streets; designs and dimensions of rail transit rights-of-way, track layouts, signals, stations, yards and workshops, etc.
- *Vehicle data,* including body dimensions, design and performance data, age, condition, breakdowns, etc.
- *Operating conditions on lines:* traffic regulations, relationship of transit vehicles to other traffic, their speeds, standing times and reliability of service; rail transit ROW conditions, speed limits, signal operations, etc.
- *Types of services provided and schedules* for all modes and lines.
- *Usage of services:* passenger boardings, alightings and volumes along the lines, their time fluctuations and their trip lengths; trip generation of different major activity areas, such as stadiums, university campuses, medical complexes, and others.
- *Miscellaneous information on events in operations, fares, and passenger attitudes,* such as accidents, crime incidents, methods of fare collection and types of fares, train-dispatching efficiency, passenger preferences with respect to schedules, riding comfort, and other service parameters.

Much of this information (e.g., transit ROW and station facilities, vehicle characteristics) should exist on permanent plans and records. Data on variable phenomena must be obtained through field surveys: observations, counts, measurements, and interviews. Transit agencies must therefore perform a number of surveys, some on a systematic (regular) basis, others randomly or as special needs occur. The most common transit surveys are described here.

1.2.1 Organization of Surveys

Good planning and continuity in data collection are necessary to obtain and maintain the current data base. The frequency and comprehensiveness of field surveys must be determined based on a compromise between the need for accurate information and the cost of surveys. A good practice is to organize major, detailed surveys at longer intervals and supplement them by minor ones, often on a sampling basis, within these intervals. In other words, a comprehensive cross-sectional set of data, describing entire operations at one time, is combined with longitudinal or time data on key elements, such as passenger volumes on maximum load sections, the busiest (or most typical for the line) stations, or peak-hour flows into the CBD.

For example, every five years the transit agency may organize passenger counts on all routes on one weekday, and on several routes for a whole week. Selected routes are counted every month of that year to find seasonal variations of travel. This count provides data on passenger demand by line: its distribution along the line and its hourly variations. Then every year several lines selected as typical of the whole system are counted along their entire lengths or on their maximum load sections. This may be done during all hours, or during peaks only, depending on the variations that may have occurred and on desired accuracy of data.

Surveys of transit operations often focus on TU running conditions: operating speeds should be measured regularly since they must be known for the purpose of scheduling. However, analyses of traffic conditions, reasons for delays, etc. are usually done on an ad hoc basis, when a new service is planned, or on existing lines when conditions that affect transit operations change or operational improvements are contemplated.

Each type of manual field survey requires a special field sheet and at least one summary sheet. These sheets are usually specially designed for each transit system and type of survey to suit the specifics of the location and type of survey and/or analysis. Typical forms that can be used for many surveys with minor adjustments are given in most manuals for transit and traffic operations, such as Vuchic et al. (1978), Hulbert (1982), and Robertson (1994).

1.2.2 Transit Speed-and-Delay Survey

The purpose of a speed-and-delay survey on street transit lines is to find the distribution of time a TU spends in travel, classified as running, dwells at passenger stops, and several categories of delays. It is also intended to record the locations, durations, and causes of the delays. The measured travel times and delays are used to compute operating speed and reliability of service during different periods of day. This information is important for planning possible operational improvements.

A speed-and-delay survey is usually performed by an observer riding the transit vehicle in a front seat so as to observe reasons for slowdowns and stops. To obtain representative results, several runs are made on one day and repeated on other weekdays. Their exact number depends on statistical variability and the desired accuracy of results. The observer is given detailed instructions on how to define stopped times, how to classify causes of delays, and what codes to use on the field sheet. The observer's equipment may consist of the field sheet clipboard, pen, watch, and stopwatch for measuring delays (in seconds), or a portable electronic recorder with function keys for standing at passenger stops, red signals, and other delay categories. Use of such a recorder eliminates the need to count seconds for various time components.

Speed-and-delay data can also be collected by an on-board computer connected to the vehicle's power supply, drive shaft, and door-opening mechanism. It records the locations and times of each vehicle start and stop, as well as door openings. Its diskette is then given to a main computer for summary and analysis of the data. This type of survey provides more exact and detailed data with far less personnel, although with fewer personal observations about special events.

Manual speed-and-delay surveys have been replaced in many cities, such as Zürich, Oslo, and Helsinki, by equipment monitoring travel of TUs on lines: their passing of fixed points, global positioning system (GPS), or automatic vehicle location (AVL) systems. The information obtained is automatically transmitted to the control center, where computers convert it into graphical time-distance diagrams with the lines of scheduled and actual travel. Deviations from scheduled times are shown in different colors when they are within a certain time interval, such as two minutes. They are shown in different colors if the vehicle is late or if it runs ahead of schedule. This allows the controller on-line monitoring of operations and intervention when irregularities occur, thus increasing service reliability.

Finally, transit systems with automatic driving, control, or monitoring of TUs on the lines usually have automated recording systems for travel times and other operational data. This is the case with metro systems that have automated train operation (ATO) and LRT systems that have signalized lines, as well as with rail vehicles and buses that have various types of AVL systems. These systems continuously monitor locations of all vehicles, and the entire travel process is automatically recorded.

The data from speed-and-delay surveys on streets are summarized and statistically analyzed. Travel times are classified by line section, time period of day, and category: running times, dwells at stops, and standing for other reasons, such as traffic congestion, signals, parking maneuvers, turning movements, etc. Average travel times are then computed for each line section with deviations indicating the reliability of service.

Travel time breakdown on different categories can be presented numerically on a summary sheet, or, more clearly, graphically as a bar chart or a pie chart with segments showing percentage distribution. These are used for analysis of the efficiency of operations: comparison of running times with dwell times at passenger stops, and standing for other reasons. These data are essential in planning new routings, operational improvements, particularly the introduction of various transit priority measures, one-way street regulation changes, elimination of parking, etc. For such changes before-and-after speed-and-delay studies can be performed. Sometimes monitoring of individual driver's performance can also be done.

Speed-and-delay studies are usually not performed on rapid transit systems, since they do not have random delays caused by numerous external causes. However, very precise surveys are sometimes undertaken to measure station standing times and travel regimes between stations so that the extremely high level of service reliability required on rapid transit systems can be achieved.

1.2.3 Passenger Volume and Load Count

The purpose of passenger counts is to determine the passenger volumes on TUs over different sections of a line, the maximum TU load and the section on which that occurs, variations of volumes in time, analysis of service quality—i.e., all elements needed for scheduling of operations.

A detailed survey should include TU passenger load counts at several points along each line, particularly focusing on the sections with heavy loads to find the maximum load section (MLS). Minor annual surveys may be limited to MLSs and one or more additional sections on each line to verify the changes recorded on the MLS.

Personnel for the load count consist of observers at each counting location. For single vehicles operating on streets, one observer is usually sufficient at each location. If the route is heavily loaded or has simultaneous loading of several vehicles, two observers may be needed. Larger teams are needed at busy rail transit stations and bus terminals.

The observer must have adequate training in the method of counting. Very often, exact counts cannot be made because of short standing time of the TUs, so the observer must know how to make a fast estimate of number of people in each group or seating section. The observer should be trained in this through fast estimates at locations where exact counts can be made to check his/her estimates so that he/she can correct any tendency to over- or undercount. In addition, the observer must know the exact seating and total capacities of vehicles, so that he/she can accurately estimate the number of passengers in a full vehicle; or, when there are a few standees, their number can be added to the seating capacity of the vehicle.

It is often possible for the observer to count the load on the arriving vehicle as well as alighting and boarding passengers. These data then provide counts for both sections before and after that station.

Each observer must have an electronic recorder with the appropriate program, or a specially designed field sheet, clipboard, pen, and watch. The following information should be recorded:

- Data about the count: line, location, TU capacity, date and day, time period of count, weather conditions, name of observer, and remarks
- Count records in columns: run number, scheduled and actual arrival times, and passenger counts: on arriving vehicles, alighting and boarding (if possible), and on departing vehicles (this can be computed later)

Counts are summarized after the field survey on a single summary sheet for the 15- or 20-minute peak, 30- or 60-minute off-peak periods, and average TU loads for each period. These numbers are then ready for use for line scheduling and various analyses.

1.2.4 Passenger Boarding and Alighting Counts

The most detailed information on passenger volumes on a transit line is obtained through the counts of boarding and alighting passengers at each stop or station along its entire length. This count provides the data on the number of passengers using each station, as well as TU loads at all points along the line. One can also compute the distribution of passenger trip lengths and total work of the line in passenger-km for any hour or day. This represents virtually all the information needed for scheduling, analyses of operations, extending or shortening of the line, addition or abandonment of some stops, etc.

The decision on what time periods to make this survey is again dependent on the cost of the survey and the need for accurate data. Whenever the operator can afford its higher cost, a boarding/alighting survey should be undertaken instead of a passenger load count, since it provides much more detailed information.

Personnel are usually best utilized on street transit if one observer travels on each vehicle and counts passenger boardings and alightings at up to two doors. Thus, one person can cover a regular bus, but on long LRT vehicles two persons may be necessary. At very busy stops, additional observers may be placed to assist. For high-frequency service, particularly on rapid transit lines, it is more efficient to place one or more counting persons at each station.

Personnel must be given similar training to that for load counts. Boarding/alighting is usually simpler to count than vehicle occupancy, provided the observer can be positioned inside or outside the vehicle so that he/she can easily observe all doors.

The equipment required again consists of an electronic recorder or a special field sheet, clipboard, pen, and watch. It is convenient to use one sheet for each one-way or for each round-trip run. After the survey, recordings from all TUs (or stations) are summarized by time periods, usually 15 minutes for peak and 60 minutes for off-peak hours. A simplified example of a summary sheet is given in Table 1.4. These data are used for development of various passenger volume diagrams, such as shown in Figure 1.11, for computations of trip length distribution on the line, boarding density (prs/km), travel density (prs-km/km), and other characteristics of line usage.

There are a number of automatic passenger counter (APC) devices, from hand recorders and portable computers to automatic scanners and vehicle step pressure detectors, recording boardings and alightings of passengers. They can be utilized in different ways to decrease manpower required for counts, increase the volume and accuracy of data, and simplify their recording. For example, the transit agency in Oslo has developed a counter for its LRT vehicles that is installed behind the driver and connected with an axle to record its revolutions and thus compute vehicle location along the line. The counter records passage of every person entering through the doors. The disc from this counter is then taken to a main computer in the office, where its record is translated into passenger count summaries by station along the line.

The most complete and reliable automated passenger counts can be obtained on the modern rapid transit systems with fully controlled stations, i.e., with computer-controlled fare gates at which every passenger inserts a magnetic card—either a sliding or a contactless "smart card"—for entry and exit (e.g., BART, Washington Metro, many systems in Japan, Korea, and Hong Kong). These gates provide continuous counts

Table 1.4 Summary of passenger counts and computation of person-kilometers

Route no. *41* Direction *NB* (Departure from South terminal) Time period *16:30–17:30*

	Number of Passengers				
Transit Stop	*Boarded*	*Alighted*	*On Vehicle*	*Km Between Stops*	*Person-km*
1 (South terminal)	48	—	48	2.2	105.6
2	35	12	71	2.8	198.8
3	54	30	95	3.2	304.0
4	29	13	111	3.4	377.4
5	16	46	81	2.0	162.0
6 (North terminal)	—	81			
Totals	182	182	—	13.6	1147.8

Average trip length: 1147.8 : 182 = 6.3 km
Average boardings/km: 182 : 13.6 = 13.4 prs/km
Average passenger volume/km: 1147.8 : 13.6 = 84 prs-km/km

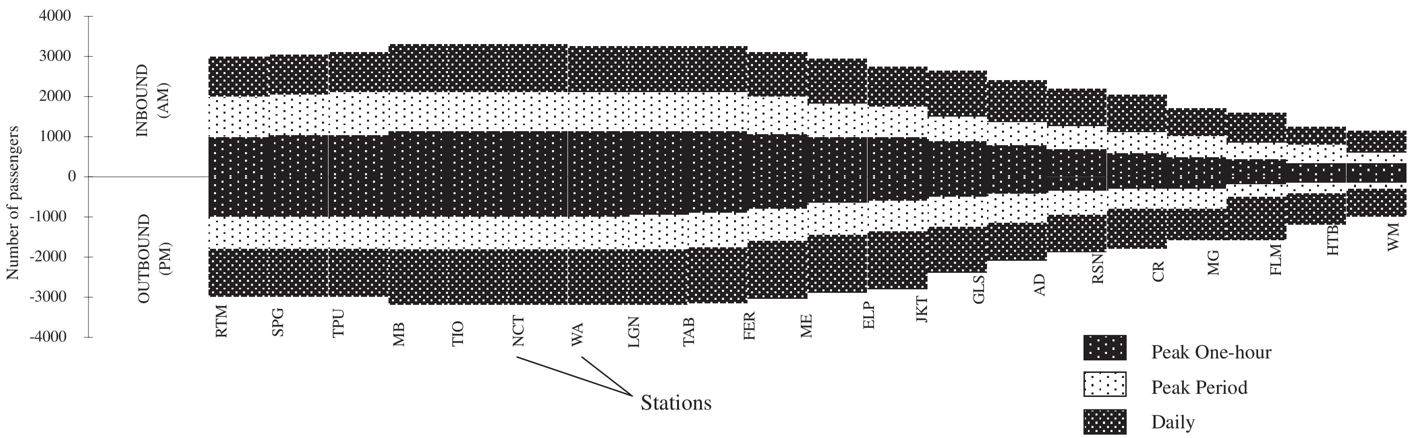


Figure 1.11 *Passenger volume profile by direction for different time periods* (*Warminster Line, Philadelphia Regional Rail System*).

of entering and exiting passenger volumes for each station, so that all data for station usage and a loading profile of the line are obtained. Slide-through magnetic tickets or smart cards that provide passenger count data are also increasingly used on street transit vehicles. When electronic fare collection is used, personnel are used only for special surveys of passenger attitudes, rates of boarding, etc.

1.2.5 Other Types of Surveys

In addition to the above-described rather standard surveys, many other surveys can be undertaken to obtain various types of data. Several examples of such surveys are described here.

- *Transfer counts*. Volumes of passengers transferring between two or more lines at transfer stations must usually be manually counted. However, on some transit systems transferring passengers are required to use magnetic or smart cards that record their path and automatically collect the statistical information about transfers. Results of such counts at major transfer stations are shown in Figure 1.12. Information about transferring is necessary to distinguish the number of passengers who use two or more lines from those traveling on individual lines only (''linked'' from ''unlinked'' trips); it is also useful for planning the types of fares and methods of their collection.
- *Fare usage*. Counts of passengers by the type of fare they pay on street transit vehicles also must sometimes be performed manually when self-service fare collection is used. This is often the case with LRT systems that have open stations and simultaneous boarding and alighting on many doors. Their results are used for analyses of fare structure and collection and control methods, and sometimes to make possible passenger counts through fare revenues. Automated fare collection on closed systems, such as metros, records all transactions and on some systems, such as the San Francisco BART, provides origin-destination records for every passenger. Thus, it is possible to obtain entire origin-destination matrices for any time period.
- *Passenger travel information*. In addition to information on passenger origin and destination stations along a transit line, it is desirable to know actual origins and destinations of their entire trips, including lengths and modes of access to/from stations, i.e., so-called submodal split. This survey usually requires interviews with passengers or distribution of questionnaires, which they complete and drop off at their destination stations or mail back.
- *Attitudinal and modal split surveys*. Preferences of passengers with respect to service parameters, such as headway length, speed, reliability, fare level, and others, and the importance of these for their choice of travel mode, are also found through interviews or questionnaires.
- *Use of timetables*. Surveys of passenger arrival time distributions in relation to scheduled TU departures can show for what services passengers use schedules and how that use is influenced by headway lengths, reliability of service, and convenience of obtaining schedule information.

As this description shows, field data collection is performed both by automatic devices and by personnel in the field. Specialized surveys, particularly for information on operational details and passenger attitudes, are usually performed manually. In all cases, handling of collected data and their storage, retrieval, and presentation are most efficiently done by use of computerized data storage. Permanent numerical and other data files allow systematic organization, storage, and fast retrieval, as well as printing of various tables or graphical displays such as diagrams, bar charts, passenger volumes along lines, or on entire networks.

1.3 TRANSIT TRAVEL CHARACTERISTICS

Transit travel volume and characteristics must be known for short-range planning of operations (scheduling), as well as for travel forecasting, which repre-

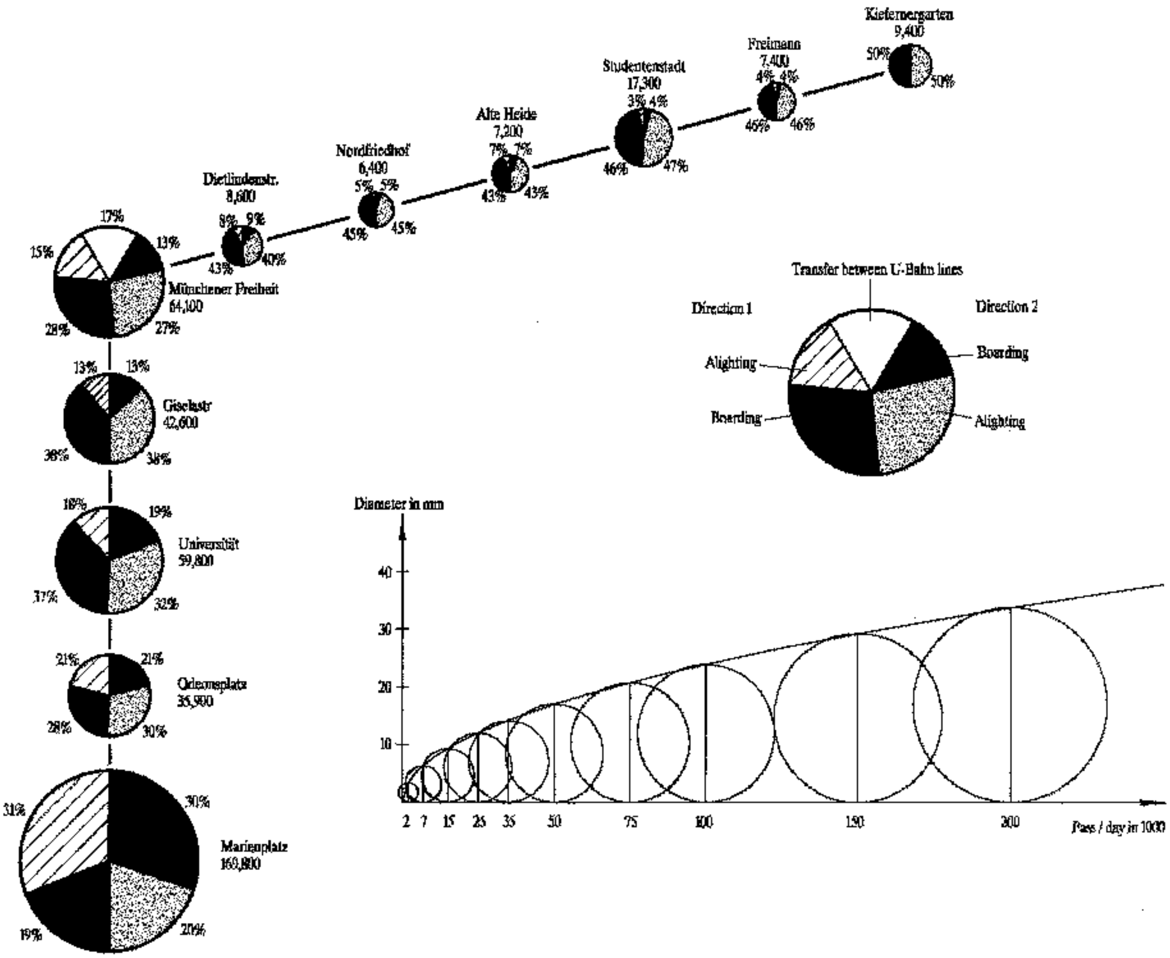


Figure 1.12 Graphical station passenger count summary: boardings, alightings, and transfers at the Munich U-Bahn stations. (Source: *Münchner Verkehrsverbund Annual Reports.*).

sents a basis for long-range planning. Although travel characteristics vary among cities, many general features are common. They are described here.

Transit travel, or use of transit services, is often referred to as *transit demand*. More precisely, this term may have two definitions. When it is defined as the number of passengers who want to use a given service and pay its price, the concept of demand is the same as passenger travel volume on a transit system. Alternatively, transit demand is defined as volume of travel that would take place when service is very good and price (fare) low or moderate. Since demand for travel is elastic with respect to level of service and its price, demand by the second definition, which should more precisely be referred to as *potential transit demand,* is at least equal to, but usually greater than, passenger volume on a given transit system. The difference between the two represents unrealized travel and is known as *latent demand*. Therefore, large-scale planning may refer to transit demand in a global sense, while in descriptions of actual passenger travel that occurs on transit systems the same concept is more precisely designated as *passenger volume,* or the number of passengers transported per hour.

Consequently, transit travel refers to actual passenger volumes under given conditions, and its magnitude varies with size, form, and character of the city, among transit modes, lines, and times of day. It is important to bear in mind that level of service (LOS) consisting of many factors affecting passengers, such as speed, reliability, comfort, and others, influences the ability of transit to attract passengers. Improved LOS, particularly if it is initially very low, can significantly increase transit travel. On the other hand, service deterioration or an increase in transit fares may decrease passenger volume.

1.3.1 Factors Influencing Transit Travel

Generally, the volume of transit passengers on a line or in a network depends on the total volume of passenger travel in a city (or in a specific area) and on the LOS/price package that transit offers as compared to the corresponding packages of competing modes, such as the private automobile, taxi, or bicycle.

The relationship between transit and auto LOS and price (or cost) of travel varies with many factors, but mostly with the concentration of potential travel demand. In cities with well-planned, coordinated intermodal transportation systems, transit is a superior mode for travel in the central city area and along heavily traveled corridors, even if auto ownership is high. In suburban areas with dispersed, low-density travel, travel by auto is more convenient, although transit systems can be organized to offer a reasonably good basic service. This is done by an integrated rail network with bus feeders, or with a timed transfer system (see Section 4.5) type of service. Transportation policies, particularly unlimited versus controlled parking, can strongly influence this balance between transit and automobile travel, referred to as *modal split*.

In cities that have concentrated all their efforts on highway construction, even major travel corridors are dominated by auto travel. Lack of acceptable transit service suppresses transit use and mobility in general. It also causes many travelers to shift to the automobile if they can own and drive one. Frequent traffic congestion then causes higher total cost and lower average quality of travel in the corridor by both modes, auto and transit.

The inherent advantage of transit over automobile in handling high passenger volumes explains the fact that transit share is highest in dense areas such as central business districts (CBDs) and other major activity centers; higher during peaks than at other times of day; and generally higher for work trips than for other purpose trips, because that travel is most concentrated in space and time. Work trips are followed by school and shopping trips, while social-recreational travel gives transit the lowest shares. Thus, in a city with good transit service, such as Munich, Paris, or San Francisco, typical shares of transit trips in central city by trip purpose are in the following ranges: work 60–90%; shopping 40–60%; school 50–70%; social-recreational 20–40%; and other purposes 10–30%.

Trip length distribution also varies among cities, depending on their density and form, types of transit and street networks, fare structure, and other factors. Several typical trip length distributions are shown in Figure 1.13. Cities with dense central areas (e.g., London, Milan, San Francisco) have very skewed distributions, due to large volumes of short (1–4 km) trips. Extensive commuting to the CBD from outlying areas (and to some extent among these areas) results in high volumes of long trips: this is found in most U.S. and Canadian and some West European cities, as exemplified by Hamburg in Figure 1.13.

Trip length distributions for individual modes depend on the role each mode plays, the structure of its network, and the type of service it provides. Street transit modes—streetcars, trolleybuses and buses—typically carry the shortest trips; average lengths are usually 4–8 km. Rapid transit systems carry longer urban trips, typically averaging 6–12 km. Only in cities that have extensive network coverage of central areas, such as Madrid, New York, and Paris, does rapid transit also serve many short trips. Suburban bus lines often carry rather long trips, unless they are local within individual areas and feed long bus or rail lines into central city. Finally, regional rail systems typically serve very long trips: average trip length on this mode for all U.S. cities is about 22 km. On the San Francisco BART, which has a network regional in character, the

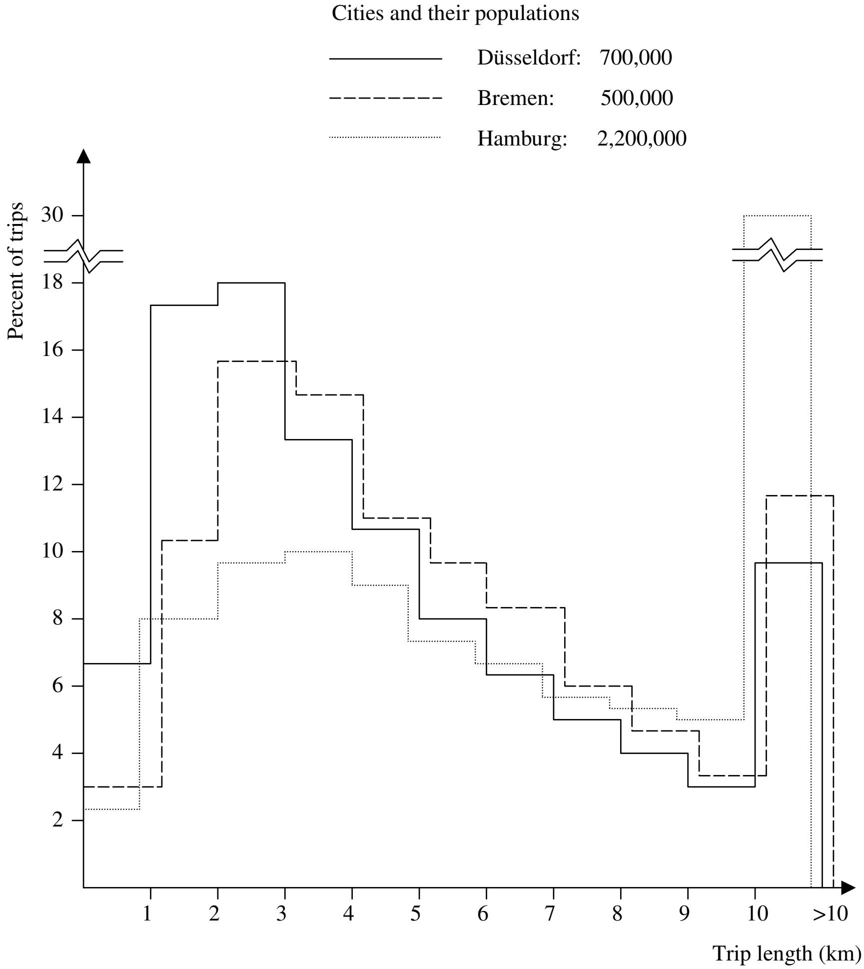


Figure 1.13 *Transit trip length distributions in several cities.*

average trip length is 21 km. A diagram of typical cumulative distributions of trip lengths for different modes is shown in Figure 1.14.

1.3.2 Spatial Distribution of Transit Travel

Distribution of travel demand by area and direction is a function of city form and land use in it. The highest travel volumes are always concentrated on radial directions toward the area of most intensive activities—the CBD. These directions therefore have the highest level of transit service. A typical transit travel volume diagram for a small city is shown in Figure 1.15.

Regional subcenters (or major activity centers) in suburban areas are also generators of large travel volumes. The transit share of these volumes depends greatly on the type of planning and design of these centers and their relationships to transportation networks. If the planning of the two is coordinated and transit services are carefully incorporated with the center's design, as has been the case in various new developments in Frankfurt, Gothenburg, Montreal, Stockholm, and many other cities, the transit share of

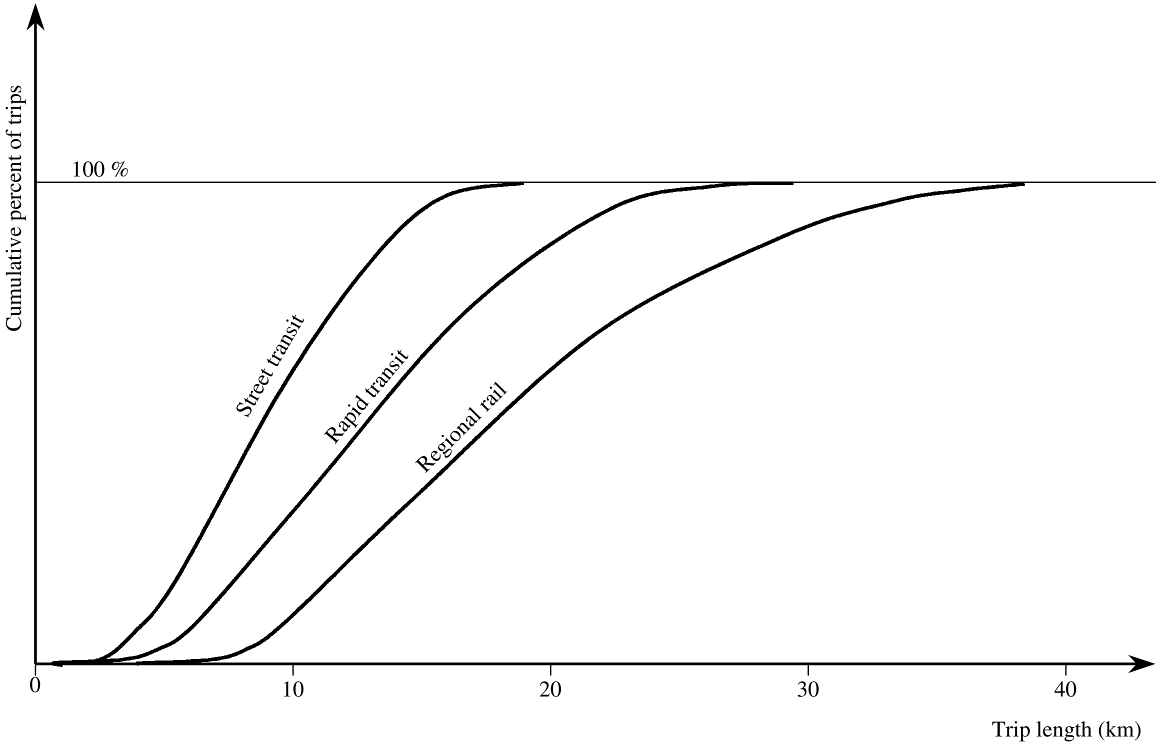


Figure 1.14 *Typical cumulative distributions of trip lengths for different modes in a large metropolitan area.*

travel is quite significant. The most important element that makes transit attractive and competitive with auto travel is the provision of separate transit ROW category B or A, which is easy during the initial design but very difficult to build later.

The other extreme, when transit is disregarded in the planning and design of major suburban activity centers (as was done in some auto-oriented cities not only in the United States but also in some countries in which transit has a very important role), it may be difficult to provide a reasonably good transit service economically. Lack of transit service may cause serious problems of excessive congestion and lack of mobility for persons who do not own or do not drive cars. Sometimes this condition leads to various citizen-organized actions, such as use of rented buses, group taxis, or other forms of semipublic paratransit. These actions can ameliorate but not solve the problems created by the absence of adequate public transit services, because semipublic modes do not offer all-day services, nor are they available to the general public.

There is sometimes a tendency to believe that if in a certain urban area or corridor there is little transit travel, there is not much demand for it. However, a major reason for low usage may be very poor transit service. The fact that people cannot use transit service if it is not offered, or that they do not want to use it if it is slow, unreliable, or expensive, should never be overlooked. Actual transit travel in such cases may be much lower than potential transit demand.

Cordon counts around CBDs represent a good indicator of the quality of transit service and its role in the city. The percentage of trips into the CBD by transit is highest during peak (commuting) hours, and it is generally higher for larger than for smaller cities. In small auto-oriented cities, transit often carries only 20–

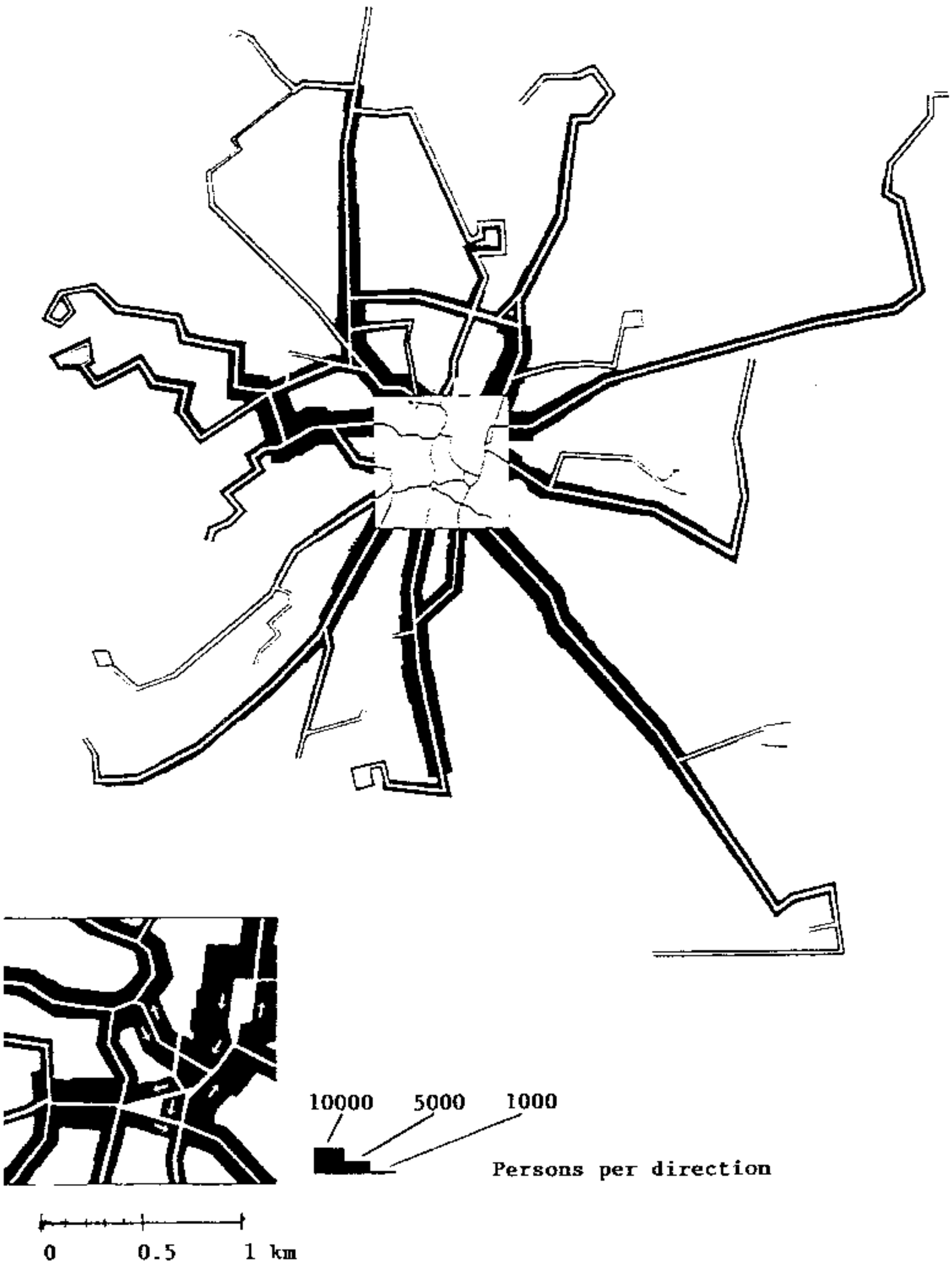


Figure 1.15 Typical transit flow pattern.

30% of trips during peak hours, and a lower percentage for the whole day. In medium-sized cities (400,000–1 million populations) this percentage is about 40–60% when transit service is good. The highest percentage is found in very large cities with extensive rapid transit systems, such as New York, Paris, and Tokyo, where transit carries up to 95% of peak-hour trips into their CBD areas.

Some cities, such as Chicago, London, and New York, have precisely defined central cities and make cordon counts regularly. Such counts show intermodal shifts of travel over time. During the 1950–1980 period, with increasing auto ownership and road construction there was a general trend of increasing use of automobiles as well as rapid and regional rail transit, mostly at the expense of buses, which were seriously affected by street congestion. This trend showed, however, the importance of the level of transit service. For example, a drastic improvement of bus speeds and service reliability achieved in Paris through the introduction of many reserved bus lanes during the 1970s resulted in an increase in the share of bus passengers; deterioration of rapid transit services in New York in the early 1980s resulted in corresponding passenger losses; subsequent service improvements were followed by increases in ridership. In the late 1990s, New York also presented an interesting demonstration of elasticity of demand to transit price. Introduction of the electronic ‘‘Metrocard’’ was used to allow free transfer between subway and buses. Compared to the preceding separate fares on each mode, this represented a reduction of fare by 50% for the transferring passengers. This resulted in a network-wide increase of passenger volumes by over 30% on buses, as well as a significant increase of subway ridership.

In general, the role of transit in CBD-oriented travel, particularly in large cities, remains very significant or dominant even with saturation-level auto ownership.

1.3.3 Temporal Variations of Transit Travel

Most of the temporal variations in transit travel are caused by the differences in travel patterns and levels of service by competing modes of transportation during different time periods. With higher travel concentrations in peak periods, the relative advantages of transit increase, so that modal split is higher for transit during peaks than for total travel throughout the day.

Transit travel varies with seasons, as its plot by month for several years in Figure 1.16(*a*) shows. The diagram shows considerable regularity: the summer period (June–August) is by far the lowest, late fall (October–December) the highest. May and September tend to be the closest to the average monthly volume.

These variations, however, are not typical for all transit systems. Variations are less pronounced in areas with diversified activities in large cities, while transit service in vacation resorts or other cities dominated by a single activity may have extremely pronounced and quite different variations. For example, transit ridership in such resort cities as Atlantic City (U.S.), Cannes (France), or Blackpool (England) is several times higher during the summer months than during the winter period. Special events, such as fairs, Olympic Games, or other sport events, create all-time peak demands, disrupting all regular patterns. Naturally, such peaks cannot be adopted as design values for permanent facilities, unless these are also justified for regular services after the event (construction of initial lines of Metro systems in Montreal, Mexico, and Munich were speeded up by the World’s Fair in 1967 and Olympic Games in 1968 and 1972, respectively).

Daily variations within a week, shown in Figure 1.16(*b*), are relatively minor among workdays (Monday and Friday being usually the highest), but they are very pronounced for Saturday and Sunday. Holidays are usually similar to Sundays. The difference in passenger volumes between workdays and weekend days increases greatly when weekend trips are shifted heavily to private automobiles. However, there are some notable exceptions to this pattern. In some cities that have attractive central areas and good transit service, such as Portland, Oregon, and Newcastle, England, the highest transit ridership occurs on Saturdays. In cities with low automobile ownership and heavy reliance on transit, weekend ridership is also much more similar to weekday volumes.

Hourly variations of demand are usually very pronounced on workdays, as Figure 1.17 illustrates, mostly because of the peaking character of commuter

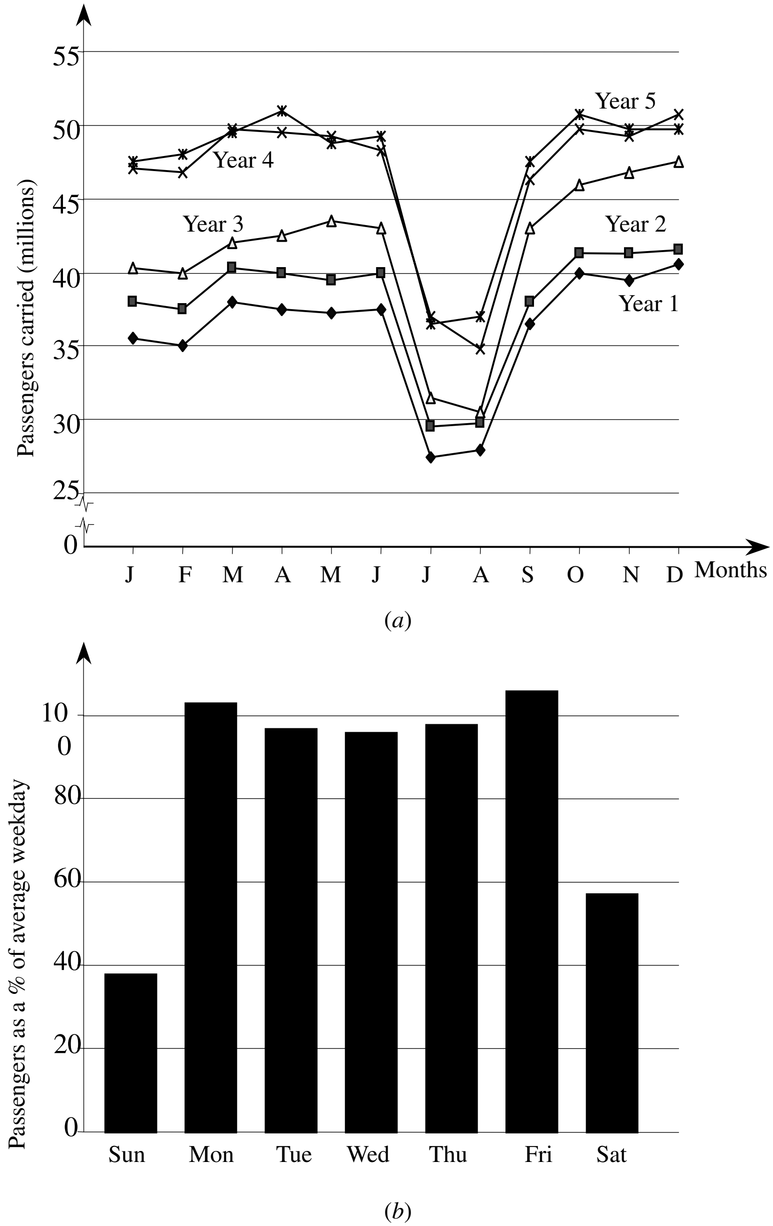


Figure 1.16 *Monthly and daily variations of transit travel:* (*a*) *monthly variations for consecutive years* (*Belgrade*) (source: *Banković, 1982*)*;* (*b*) *typical daily variations.*

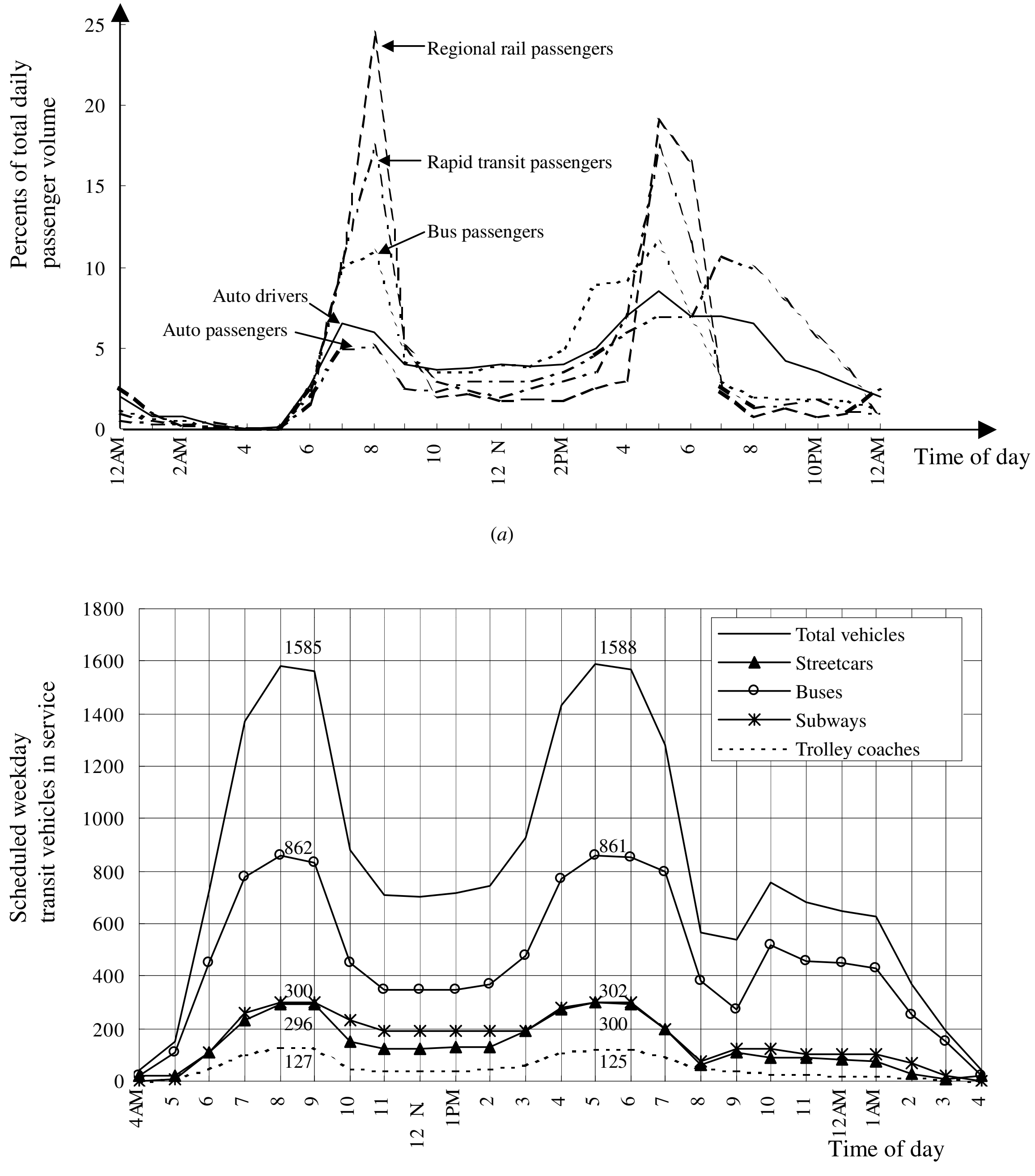


Figure 1.17 *Hourly variations in transit: peaking patterns of passengers and rolling stocks:* (*a*) *variation in passenger volume by mode in Chicago* (source: *Chicago Area Transportation Study*)*;* (*b*) *variation in fleet requirements by mode in Toronto* (Source: *Toronto Transit Commission*).

travel. Passenger counts summed up by hours are usually accurate enough for most periods of day, but not for peak hours when sharp variations may occur within an hour. Peak-hour volumes are discussed further in Section 1.4.

The peaking characteristics of passenger volume can be measured by the ratio of the highest hourly volume to the average off-peak hourly volume, such as between 9 and 12 a.m. This *peak-to-base ratio* usually varies among modes. As the diagram in Figure 1.17 shows for Chicago, the highest peaking occurs on regional rail lines, followed by rapid transit and bus passengers. Auto drivers and auto passengers are less peaked than most transit passenger volumes.

Compared to the variations and peaking of passenger volume, variations of the number of vehicles in service are less pronounced because the load factor α is greater during peaks than off-peaks. The variations are even lower for the number of TUs on guided systems, which use larger TU consists (longer trains) during peaks than off-peaks.

Variations in the required personnel to provide transit services depend on vehicle capacity, train consist, and degree of automation. Consequently, personnel requirement varies directly with the number of buses (one driver per bus); it is somewhat lower for LRT and metro systems if different train lengths are used during peaks and off-peaks. Fully automated modes, such as AGT and several metros, have even lower operating personnel ratios, which gives them an advantage in economic efficiency.

With respect to travel categories, peak-to-base ratio is the highest for radial travel dominated by commuting, which is the case on some regional rail and express bus systems. It is the lowest for services within the central city area, as well as on tangential and circumferential lines, where the diversity of trip purposes is the greatest.

Similar to daily variations of transit travel, hourly variations are also less pronounced (i.e., the peak-to-base ratio is smaller) in cities with heavy, multipurpose use of transit, which fills the troughs between the morning and evening commuting peaks. Examples of such transit use are found in Hong Kong, Mexico, Moscow, and New York.

1.3.4 Passenger Volume Analysis and Service Capacity Determination

The volume of passengers who want to travel along a transit line from a suburban terminal to the CBD terminal during one day (assuming a certain level of service) can be presented on a three-dimensional diagram with distance and time representing a horizontal plane and passenger volume the vertical topography over the plane, as Figure 1.18 shows.

A TU traveling on a line is plotted on the diagram in Figure 1.19 as a sloped line representing its distance-time path on the diagram. Since station-to-station speeds vary with distances between stations, the TU travel line has different slopes among spacings. Capacity of the TU is plotted vertically. It is constant along the line, while passenger load on it varies as a function of boardings/alightings, and it reaches a maximum value P_{max} on the MLS. Service on the line can thus be imagined as a series of slices through the topography of passenger volume, as illustrated in Figure 1.20. Each slice represents a TU with its passenger load shown as the shaded area. For clarity, variations of speed along the line are not shown on this diagram.

Naturally, if constant service were provided throughout the day, it might not be sufficient during peak hours, particularly on the MLS, while its utilization during off-peak hours would be very low. Referring to the diagram in Figure 1.20, fitting a constant-height box over the passenger volume topography could result in protruding peaks and empty space toward all side walls of the box.

To satisfy the variable demand and achieve good utilization of offered capacity, service must be tailored to cover the demand as closely as possible. This can be done in three different ways:

1. Changing headways—shortening them (higher frequency) during the peaks
2. Using different TU sizes/capacities, such as different train consists (number of cars)
3. Operating TUs on certain sections of the line only (“short-turn trains”), so that the protruding topography peaks are covered by small additional TU capacity boxes.

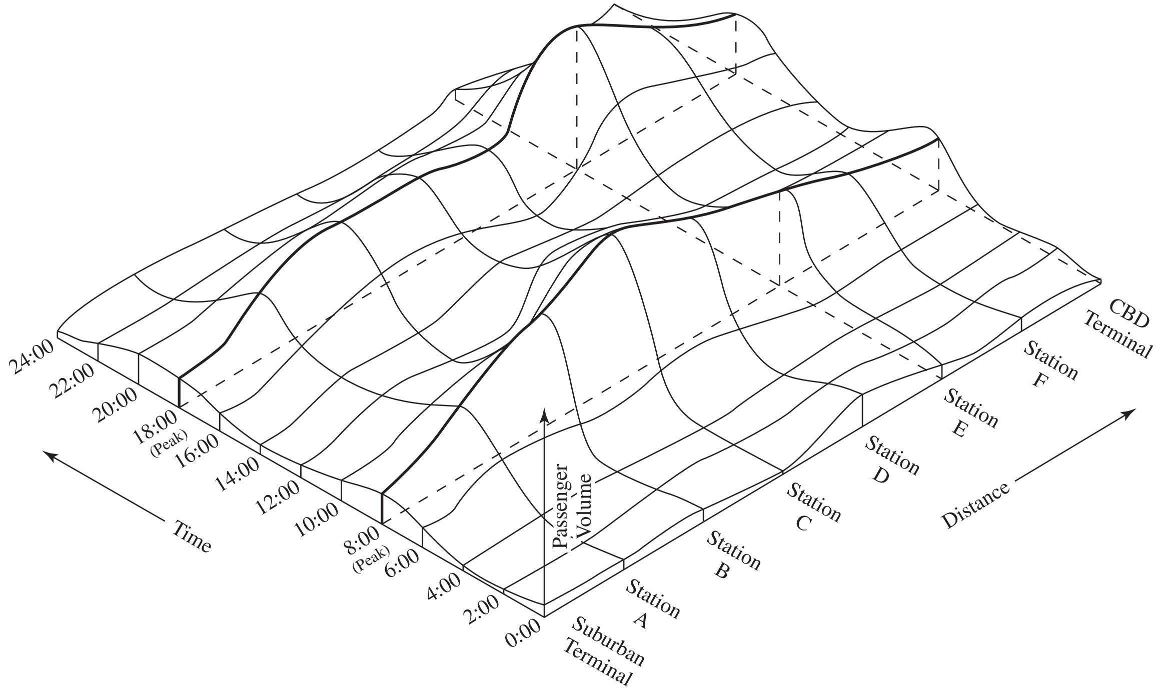


Figure 1.18 Three-dimensional presentation of passenger volume distribution in distance and time along the line from a suburban to the CBD terminal. (Design: *Thor Haatveit.*)

Another way to reduce offered capacity as the line proceeds from the CBD and volume decreases is to divide the line into two or more branches. That condition would be presented on the diagram in the same way as case 3—short-turning of trains at a point along the line.

The three capacity adjustments are illustrated in Figure 1.21.

Each of these three types of capacity adjustments is utilized in actual transit system operations, but not all are practical in every situation. The first, headway adjustments, is used in virtually all services, except when minimum headways, limited by line capacity, single track sections, etc., are already operated. The second, changing TU capacity, is used on many rail systems, since cars can be coupled or uncoupled from the train at terminals or points where storage tracks exist. With buses such adjustments (use of different buses during peak and off-peak periods) are extremely rare. The third type of capacity adjustment, addition of short-turning trains on some line sections, can be used only where intermediate turning points are available, good joint scheduling is possible, etc.

The preceding three-dimensional presentations of passenger volume and service capacity, Figures 1.18–1.20, are useful for conceptual clarity, but actual scheduling is usually performed with two-dimensional diagrams of passenger volume and capacity as functions of time and of distance, which will be presented in Section 1.4.

1.3.5 Characteristics of Travel on a Transit Line

Passenger travel on a transit line can be analyzed through various indicators. Several important ones are defined here.

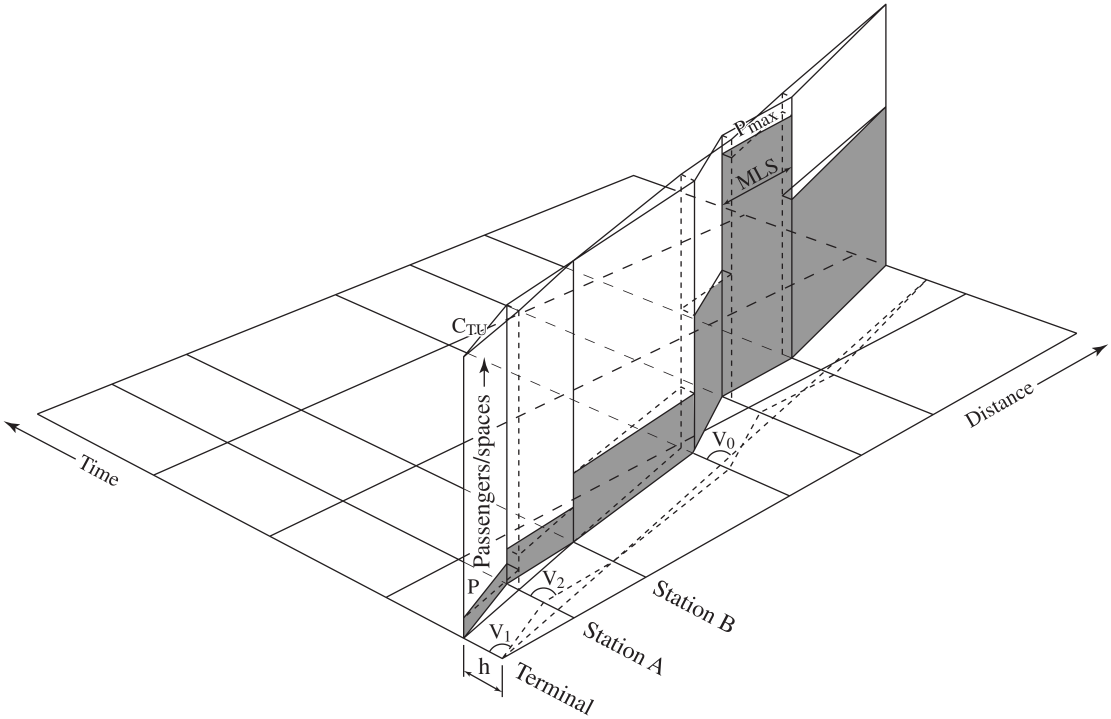


Figure 1.19 *Travel of a TU in its passenger load profile along the line.*

Average passenger trip length, or the average distance passengers travel on a line l_{av}, is obtained when the total passenger-km are divided by the number of passengers (both being for the same time period, such as hour or day):

$$l_{av} = \frac{\sum_{i=1}^{n} p_i \cdot l_i}{\sum_{i=1}^{n} b_i} = \frac{1}{p_t} \sum_{i=1}^{n} p_i \cdot l_i \qquad \left| \frac{l}{\text{km}} \right| \frac{p, P, b}{\text{prs/h}} \right| \tag{1.32}$$

where p is the number of passengers, l is interstation distance or spacing, and P is total number of passengers boarding along the line during the same time period.

Average passenger volume P_{av} is computed as the total passenger-km on the line divided by its length L:

$$P_{av} = \frac{\sum_{i=1}^{n} p_i \cdot l_i}{L} \qquad \left| \frac{P}{\text{prs/h}} \right| \frac{l, L}{\text{km}} \right| \tag{1.33}$$

Coefficient of flow variations η_f expresses the degree to which passenger volume peaks along the line. It is the ratio of the maximum volume (on MLS) P_{max} and the average volume P_{av}:

$$\eta_f = \frac{P_{max}}{P_{av}} = \frac{L \cdot P_{max}}{\sum_{i=1}^{n} p_i \cdot l_i} \qquad \left| \frac{\eta}{-} \right| \frac{P, p}{\text{prs/h}} \left| \frac{L, l}{\text{km}} \right| \tag{1.34}$$

The lowest possible value of η_f is 1, and it is found on lines with constant passenger load along their entire length, such as those which operate between two points only (for example, some CBD–airport expresses). The greater the value of η_f, the lower the average load

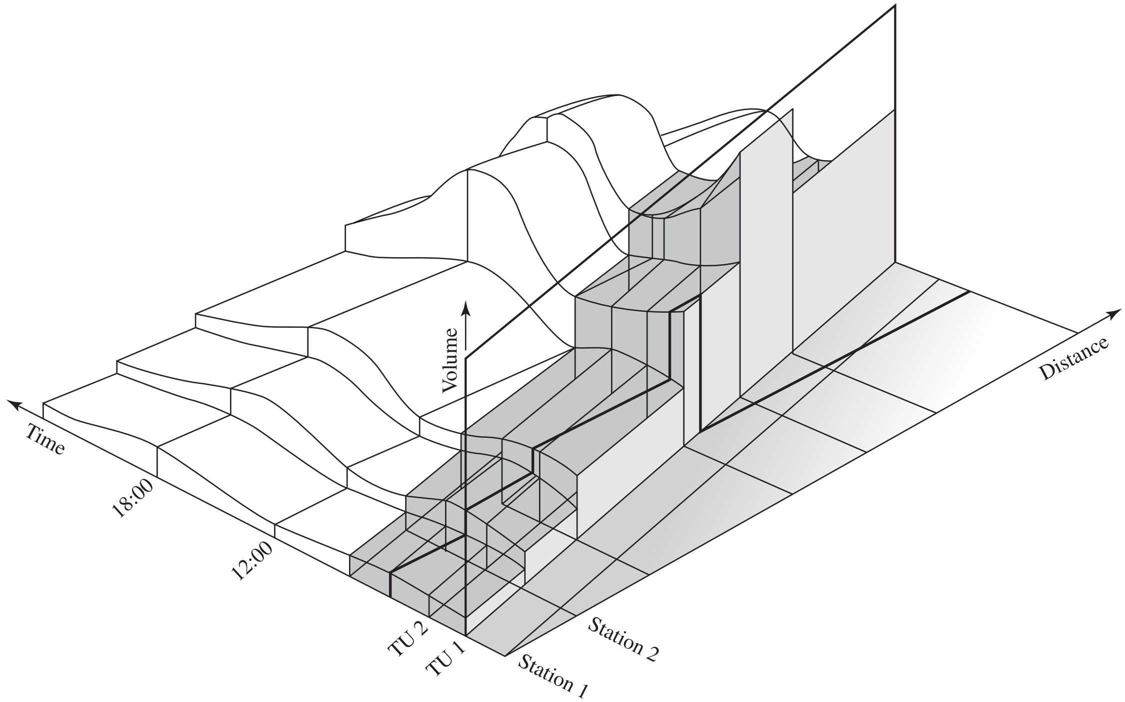


Figure 1.20 *Graphical presentation of TUs transporting passengers along a line.*

factor α_{av}, and the more desirable are adjustments of offered service to passenger volume. In the case when a line would consist of n equal station spacings and all travel would take place on one spacing only, η_f would assume its highest value:

$$\eta_f = \frac{P_{max}}{\frac{P_{max}}{n}} = n \qquad \left|\frac{\eta}{-}\right|\frac{P}{\text{prs/h}}\left|\frac{n}{-}\right| \qquad (1.35)$$

This is, of course, a hypothetical situation only.

Coefficient of passenger exchange η_x indicates what portion of passengers are exchanged along a line, i.e., their turnover rate. It is defined as the ratio of total passengers who board along a line to those who did not replace the alighting passengers. Referring to Figure 1.22(a), η_x represents the ratio of the total B_L area to the portion of B_L area that does not overlap with A_L, i.e.,

$$\eta_x = \frac{B_L}{B_L - P_x} \qquad \left|\frac{\eta}{-}\right|\frac{B, P}{\text{prs/h}}\right| \qquad (1.36)$$

where P_x is the area of overlap of B_L and A_L. Thus, η_x indicates the intensity of passenger exchange and therefore has significance for operations; it is useful to know for vehicle design, for scheduling of station dwell times, as well as in selecting fare structure.

In a general case, the boarding and alighting functions may increase and decrease several times along the line, intersecting themselves. Each intersecting point represents an extreme value in passenger flow—a local minimum or maximum. Figure 1.22(b) shows such a case, typical for a diametrical line that has two maxima, one at each end of central city.

For an actual transit line the boarding and alighting functions may follow the same general form, but with a discrete shape, as shown in Figure 1.22(c). The formula for the coefficient of passenger exchange is then:

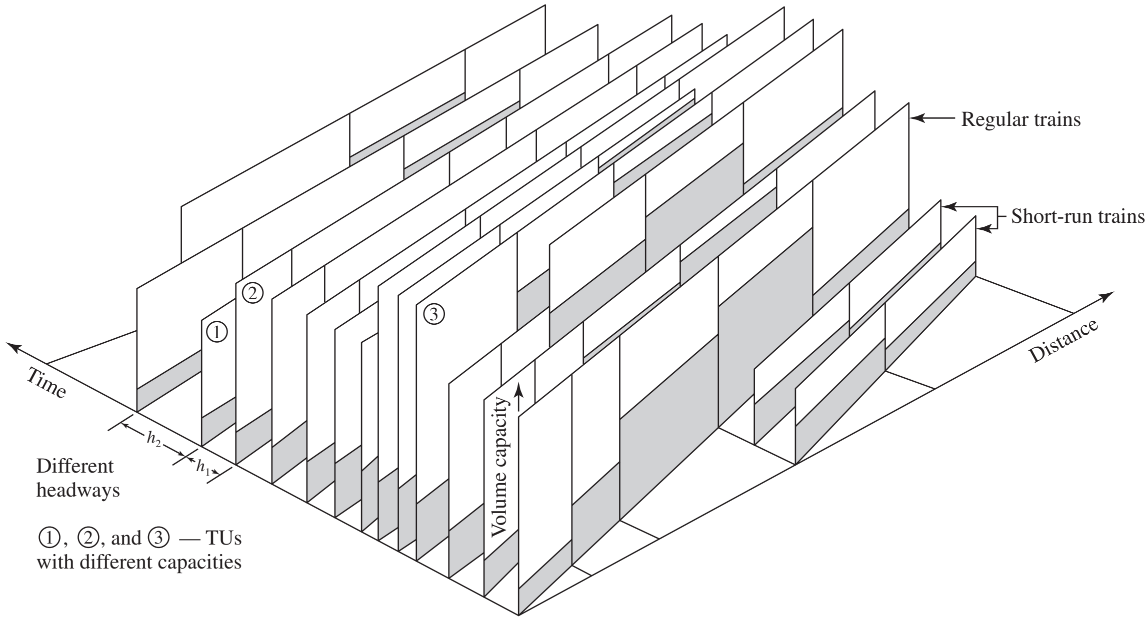


Figure 1.21 *Three methods of adjustments of service capacity to passenger volume: variation in headways, TU consists, and length of run.*

$$\eta_x = \frac{P_t}{P_t - \sum_{i=1}^{n} |b_i - a_i|} \qquad \left|\frac{\eta}{-}\right|\frac{P,\ b,\ a}{\text{prs/h}}\right| \tag{1.37}$$

Descriptively, the sum in the denominator consists of absolute values of differences between boarding and alighting volumes at all stations at which both occur.

For the lines that have only one station (k) at which the boarding and alighting curves intersect, and therefore there is only one maximum (MLS with P_{max}), as shown in Figure 1.22(d), the expression for η_x can be simplified:

$$\eta_x = \frac{\sum_{i=1}^{n-1} b_i}{\sum_{i=1}^{n-1} b_i - \left(\sum_{i=2}^{k} a_i + \sum_{k=1}^{n-1} b_i\right)} = \frac{P_t}{\sum_{i=1}^{k} (b_i - a_i)}$$
$$= \frac{P_t}{P_{max}} \qquad \left|\frac{\eta}{-}\right|\frac{b,\ a,\ P}{\text{prs/h}}\right| \tag{1.38}$$

Introducing P_{max} from Eq. (1.34) into Eq. (1.38) gives:

$$\eta_x = \frac{L \cdot P_t}{\eta_f \sum_{i=1}^{n} p_i \cdot l_i} \qquad \left|\frac{\eta}{-}\right|\frac{L,\ l}{\text{km}}\left|\frac{P,\ p}{\text{prs/h}}\right| \tag{1.39}$$

Further, replacing this sum with its relationship in Eq. (1.32) gives the coefficient of passenger exchange as a function of line length, average travel distance, and the coefficient of flow variations:

$$\eta_x = \frac{L}{l_{av} \cdot \eta_f} \qquad \left|\frac{\eta}{-}\right|\frac{L,\ l}{\text{km}}\right| \tag{1.40}$$

Equation (1.40) shows that the lowest value of η_x occurs when all passengers travel along the entire line: there is no passenger exchange, and $\eta_x = 1$. If, theoretically, all passengers are exchanged at each of the n

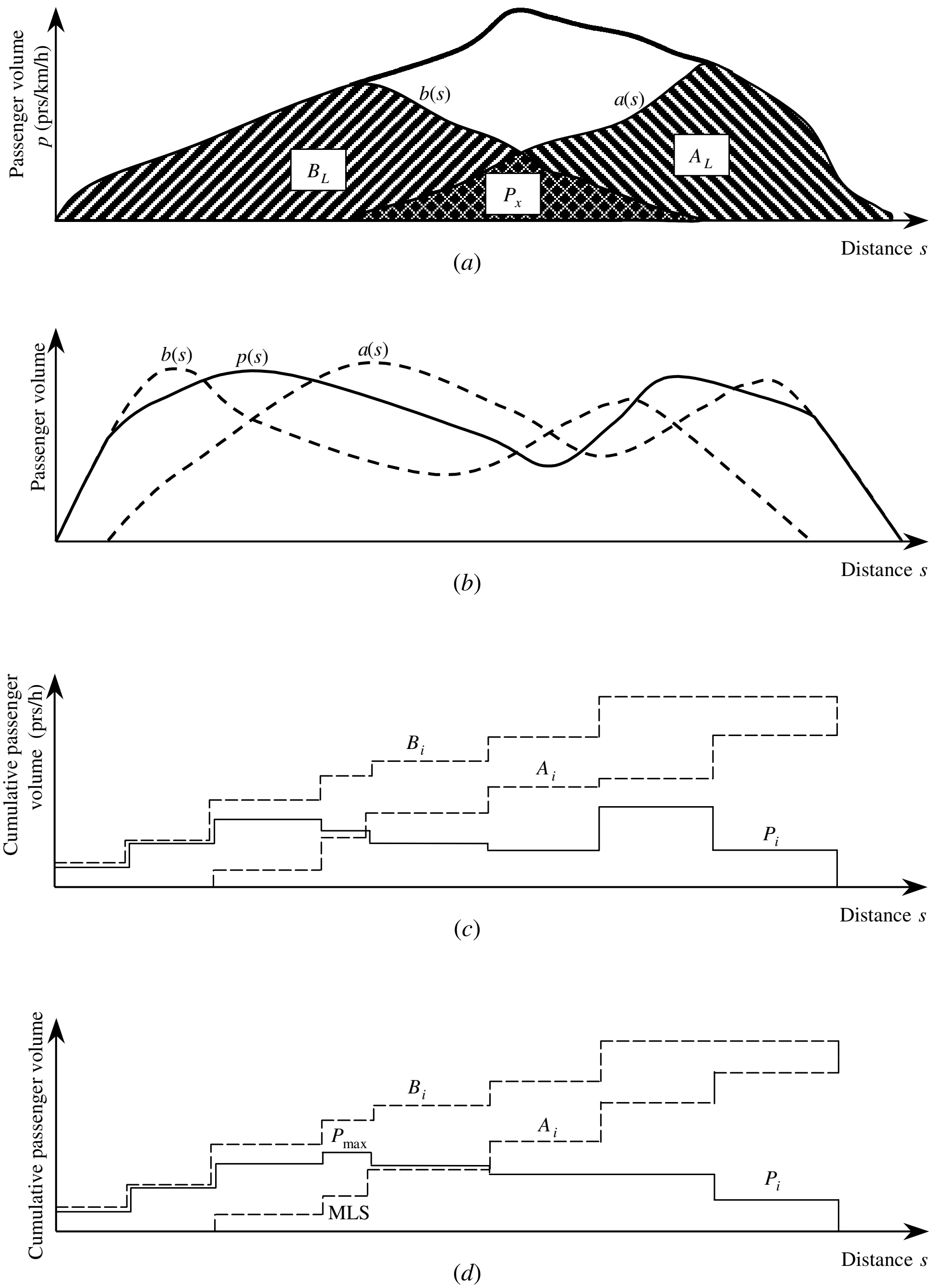


Figure 1.22 *Definition of the passenger exchange coefficient* η_x*: (a) the concept of passenger exchange coefficient; (b) general case of boarding, alighting, and flow functions; (c) discrete boarding and alighting functions; (d) discrete boarding, alighting, and load functions;* P_{max} *defines the MLS.*

stations of a line, the coefficient assumes its highest value, $\eta_x = L/\ell_{av} = n$.

1.3.6 Indicators of Transit Usage

Two indicators are most commonly used to express the absolute and relative magnitude of transit travel in a city; they reflect the role a transit system plays in the area and its relationship with other modes.

Riding habit, defined as the ratio of annual transit rides to population of the served area, indicates how much the population utilizes the transit system. Riding habit is generally greater for large than for small cities.

When transit was practically the only mode of travel for medium and long trips, riding habit in large cities was generally over 200, sometimes even close to 400 rides per capita per year (Chicago in 1926 recorded 374). After the widespread introduction of the private automobile and subsequent deterioration of transit services in U.S. cities, that number dropped to a range between 120 and 220 for large cities with rapid transit systems, frequent service, and low fare (New York, San Francisco, and New Orleans have the highest figures) and considerably below that range for smaller cities and cities that developed rapidly since 1950, generally without high-quality transit. Already in 1959 Cincinnati had only 70, Detroit 65, Minneapolis 52, and Kansas City 50 rides per capita.

This trend in transit use indicates that high-quality transit is a significant factor contributing to high transit riding habit. This correlation is corroborated by a comparison between West European and U.S. cities. Prior to the wide use of the private automobile, U.S. cities tended to have more extensive transit systems than their West European counterparts, and their riding habits were higher. In recent decades, however, as West European cities improved their transit systems to a higher level of service than most U.S. cities, riding habit in European cities has become higher: cities with populations of 600,000 to 2.5 million, such as Rotterdam, Düsseldorf and Hamburg, have about 150–200 annual rides per capita. Brussels (population 1.1 million) had 148 rides per capita in 1970; following construction of its Metro and LRT lines and other improvements, in 1984 its riding habit increased to 176. The highest riding habits are found in Oslo and Zürich: with populations of only about 500,000 but very extensive transit systems, these two cities maintained extremely high riding habits into the late 1990s: about 200 and over 300 annual rides per capita, respectively.

In medium-sized and small North American cities, reliance on the automobile is very high, so that riding habit is quite low. For example, Pennsylvania cities with populations of about 300,000, such as Wilkes-Barre and Harrisburg, have riding habits of only 15–30, while of the two large cities, Pittsburgh (1.8 million) has about 50 and Philadelphia (4.0 million) 90 annual rides per capita. These numbers are approximate because of the difficulty of defining the limits and population of the city—how many suburbs are included. This factor, very extensive dispersal of population in metropolitan areas, actually makes the riding habit numbers for U.S. cities considerably lower than in most other countries. For example, riding habits computed for transit travel in central cities of Philadelphia, San Francisco, New York, and other major cities are much higher than when they are computed for their metropolitan areas. This should be considered in comparing the computed riding habits for different cities.

It is important to note that cities with a consistent policy of transit improvement, such as was pursued in Toronto and Edmonton during certain time periods, increased their riding habits in spite of their low population densities and high auto ownership rates. The influence of good transit service on ridership attraction is also obvious from a comparison of several cities in Western Canada and the United States (1991 figures), which shows better performance of Canadian cities: Vancouver (population in thousands, 1,603), 129 annual rides per capita; Calgary (754), 103; Portland (1,172), 50; Seattle (2,348), 46; Houston (2,902), 28; and Phoenix (2,006), the most auto-oriented city, 14.

Transit travel as percent of total travel and relationship of transit to automobile travel (modal split) are indicators showing the relative role that transit plays in a city. Since these indicators vary greatly, they are usually analyzed by areas of the city, individual corridors, or travel purposes, rather than as global values for entire cities. The values given for entire cities, usually showing extremely low percentage for transit (5–15%), have been often used by laymen to downplay the importance of transit. However, these global numbers are deceptive for many reasons. First, transit does not provide ubiquitous coverage in the entire metropolitan area, as cars do. Calculating percentage of transit trips using as a base many trips that transit is not intended to serve is incorrect. Second, travel by auto and by transit is often functionally different: a person commuting by transit to work may make only two recorded trips per day; however, that person may make stops along the way or undertake, for example, another couple of walking trips during the day. To perform the same functions—get to work, go to lunch, go to a bank and a store—an auto commuter in the same city, but particularly in the suburbs, may make four auto trips in addition to the trips to/from work. Thus, while both persons have performed the same number of functions or trip purposes, the count of trips gives a false impression that auto served three times more trips and travel functions than transit.

Finally, another bias in the global modal split analysis is caused by neglect of (and difficulty in counting) walking trips. Proper land use planning and transportation policies lead to situations where walking and use of transit can serve many functions conveniently, while auto-oriented development requires a far greater number of trips and vehicle-km of travel in order to perform the same functions. Percentages of trips in an entire urbanized area is therefore an inappropriate measure of the importance of individual modes.

In some U.S. cities, another distorted statistic is also used to downplay the role of transit: the number of daily transit trips is divided by two, claiming that that figure represents the number of persons using transit (presumably, all transit riders are commuters, making two trips per day). By the same logic, if car users make, on the average, 3.2 (or a similar number) trips, traffic volume on every highway should be divided by 3.2 to show the ‘‘number of affected persons.’’ Both such figures are incorrect, and the exact number of trips should be used for each mode, rather than divided by some imaginary number of commuting trips while ignoring other trips.

1.4 SCHEDULING OF SERVICE

Transit scheduling is the process of computing the frequency of service, the number of vehicles required, the timing of their travel, and other related operating elements. The products of scheduling include graphical and numerical schedules for operators and supervisors (also known as paddles, picking lists, dispatchers’ lists, etc.), timetables for the public, as well as operating data for a line.

Public transportation services with low frequencies, such as some long-distance and suburban routes, or commuter transit, which operates during peak hours only, sometimes have variable headways, determined by demand, cycle times, crew requirements, and other constraints. For regular transit lines, however, uniform headways during each schedule period represent the optimum operation for several reasons. First, for random passenger arrivals, uniform headways minimize waiting times; second, they minimize probability of delay propagation, thus resulting in higher capacity and reliability of service. Moreover, use of clock headways allows simple information and represents an important convenience for both regular and incidental transit users. For these reasons, virtually all well-planned and operated urban transit services are scheduled with uniform clock headways for each scheduling period of the day.

1.4.1 Components of the Scheduling Process

As the flowchart in Figure 1.23 shows, the entire scheduling process can be divided into three phases.

I. *Input,* or preparation of data needed for scheduling, includes various line characteristics, schedules of lines that meet and have transfers, passenger volumes, service standards and considerations, characteristics of vehicles and train consists, operational factors and practices for each line, and work rules and standards. Clearly, these data include fixed numbers (e.g.,

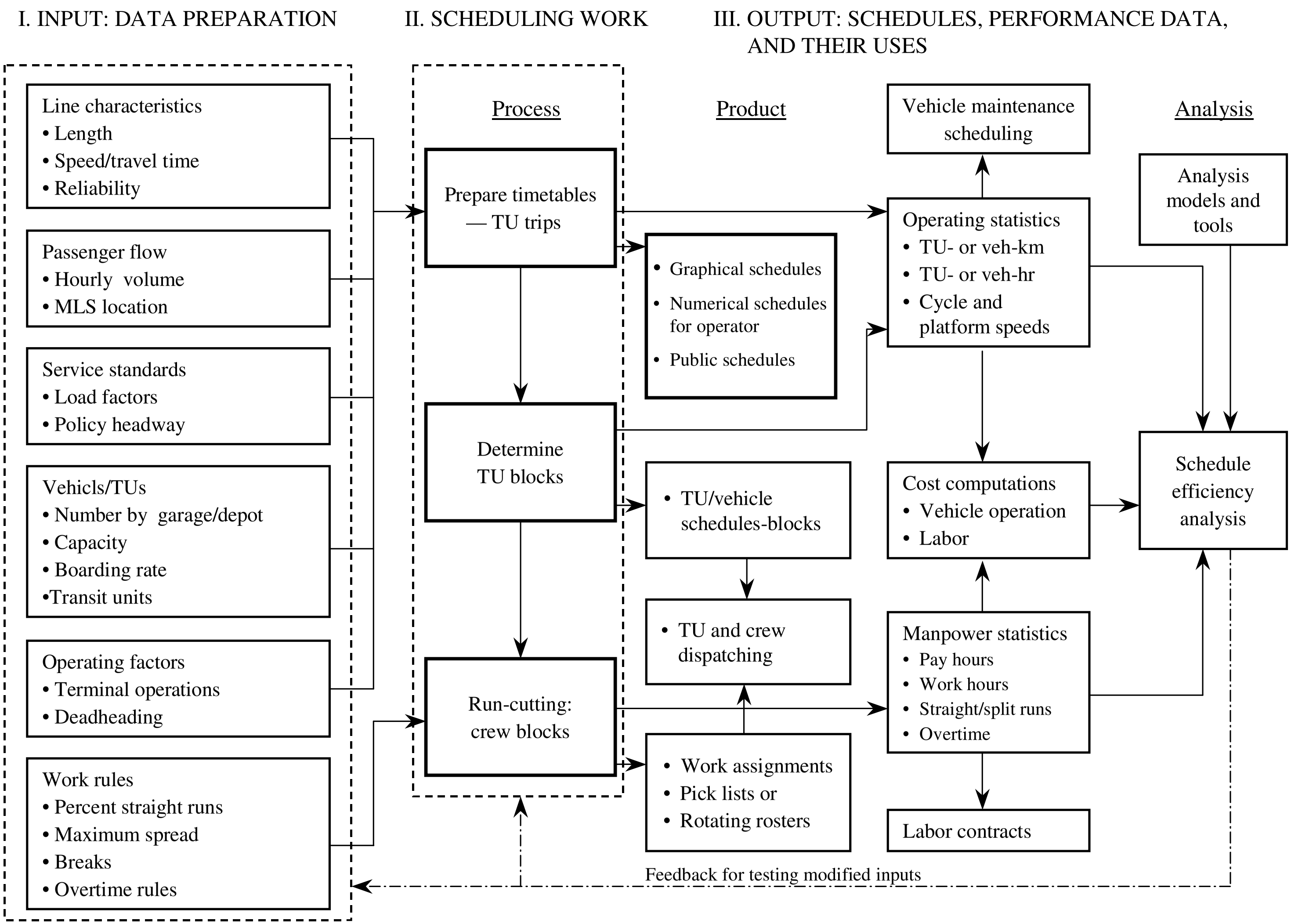


Figure 1.23 *Flowchart of transit scheduling.*

line length), data that must be periodically updated (passenger volumes), as well as various characteristics and standards, which experienced schedulers introduce in the course of their work or program preparation.

II. *Scheduling work* represents the central component of the process. In most cases it is divided into three major elements:

- *Preparation of timetables or trip building,* which determines headways, terminal times, and other elements. Its products are graphical schedules (string charts) and numerical schedules for operating personnel (headway sheets) and for the public (timetables).
- *Determination of blocks* or *block building* assigns TUs to all trips specified in the timetable. The product of this element are blocks, or work schedules for each TU for a day.
- *Run-cutting* or determination of work duties for individual drivers during the day. This process produces work assignments that are put together into pick lists or rosters from which each driver selects a specific run (which may be straight, split, include overtime, etc.).

In European transit systems, development of rotating duty rosters is sometimes treated as a separate element.

III. *Output* of the scheduling process, in addition to the direct products (schedules, blocks, runs, etc.), consists of various performance data, such as TU- and vehicle-km, pay-hours and work-hours, etc. These data are used for cost computations, various reports on transit operations, and, particularly important, analysis of schedule efficiency.

This scheduling process is shown in Figure 1.23 as a sequence of steps, from the input of information to the output of vehicle schedules and rosters. Actually, the process is often more complicated because many of the input elements, such as policy headways, types of vehicles or TU sizes, load factors, and others, may be varied to some extent, allowing testing of alternate schedules and improvements of efficiency of the initially developed schedule. The procedure therefore often has a feedback step, which allows testing of possible changes in parameters and their impact on the final schedule, so-called what-if analysis. This kind of testing is particularly common with computerized scheduling procedures because testing of many different situations is easy and fast.

It actually often happens that the last step in the process, schedule efficiency analysis, leads to the conclusion that the solutions are not satisfactory and calls for modifications in the input or scheduling process. Thus, as the dashed line in the flowchart in Figure 1.23 shows, there is a feedback from the last step, schedule efficiency analysis, to Input and Process, representing reiterative scheduling procedure.

1.4.2 Determination of Service Requirements

The schedule for every transit line must satisfy two basic requirements: it must provide adequate transporting capacity for passenger volume, and it must offer a certain minimum frequency of service (maximum acceptable headways) required from a level-of-service point of view.

During peak hours and on heavily traveled lines at all times, the former requirement is critical: the operator must provide adequate capacity on the line; if this is done, the minimum frequency is usually automatically met and exceeded. However, during low travel demand periods and on lightly traveled lines, if service is based on capacity requirements, frequency may be unacceptably low. For example, if there are 70 potential passengers per hour on MLS of a bus line, on the basis of capacity, only one or two bus departures would be provided, resulting in headways of 60 or 30 minutes. But such a schedule would be unacceptable for many travelers. Therefore headways of, say, 15 minutes should be adopted even though utilization ratio of the offered service would be rather low.

1.4.2.1 Passenger Volume Distribution in Distance and Time. The basic information needed for scheduling a transit line is its expected volume of passengers and its distribution in distance and time. Two types of diagrams are developed for this purpose.

A *passenger load profile diagram* is used to show passenger volumes P_i on each station spacing i along the line during a period of time, usually one hour, or the entire day. As shown in Figure 1.5, such a diagram shows the highest volume, $P_{\max}$, which defines the MLS. The volume on MLS is critical for determination of the required scheduled capacity.

On the diagram, three horizontal lines are plotted: average passenger volume for the entire line P_{av}, offered line capacity per hour in seats, and in total spaces C. The shaded area shows person-km traveled on the line, the P_i/C ratio represents the utilization coefficient or load factor α on that spacing (with its maximum value reached on the MLS) and the ratio of the total passenger-km traveled to the total space-km offered $\overline{\alpha}$.

This diagram is analyzed to determine the basic type of service. If the volume is much greater on one portion of line than on others, it may be appropriate to operate some short-turn services or to divide the line into two or more branches at the point where volume decreases (this is common for radial routes from city center). An example of such a profile and service with alternating short- and through-routed TUs, which reduces the offered capacity on the outer section to one half, is shown in Figure 1.24.

A *temporal variations diagram* shows detailed variations of passenger volume on the MLS of a line for each hour, half hour or sometimes, during the peaks, 15-minute intervals, covering the entire period when transit service is operated. This diagram, shown in Figure 1.25, is used to determine scheduling periods—time intervals with rather uniform travel volumes, during which a fixed schedule is operated. While in some cases one schedule may be appropriate for the entire day (e.g., a line serving many different trip purposes on a Saturday), in most cases there are three, four, or as many as six scheduling periods. The diagram in Figure 1.25 shows a typical weekday passenger demand with a.m. peak, midday, p.m. peak, evening and night (“owl”) scheduling periods.

1.4.2.2 Scheduling Periods and Design Passenger Volumes. Based on the temporal distribution of passenger volumes shown in Figure 1.25 for one line, the number and durations of different scheduling periods are determined, usually for the entire network of transit lines. Then, for each scheduling period, a diagram of passenger volume along the line, such as Figure 1.18, is developed and used for scheduling.

The general procedure of transit line scheduling will be described in the following text. However, since

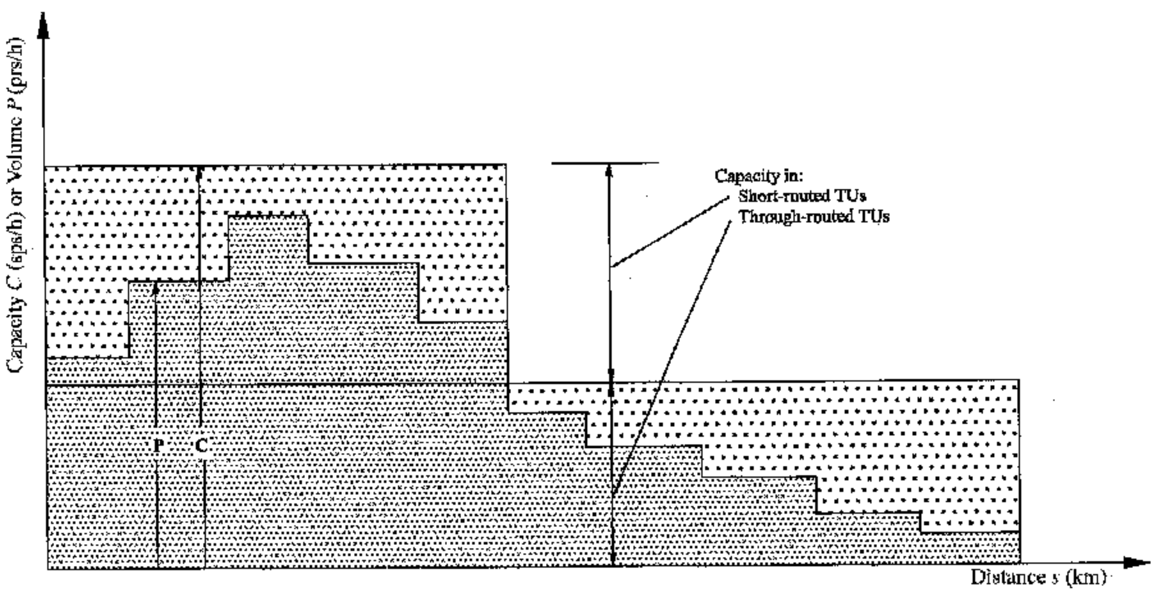


Figure 1.24 *Offered capacity and its utilization on a transit line with through- and short-turn TUs.*

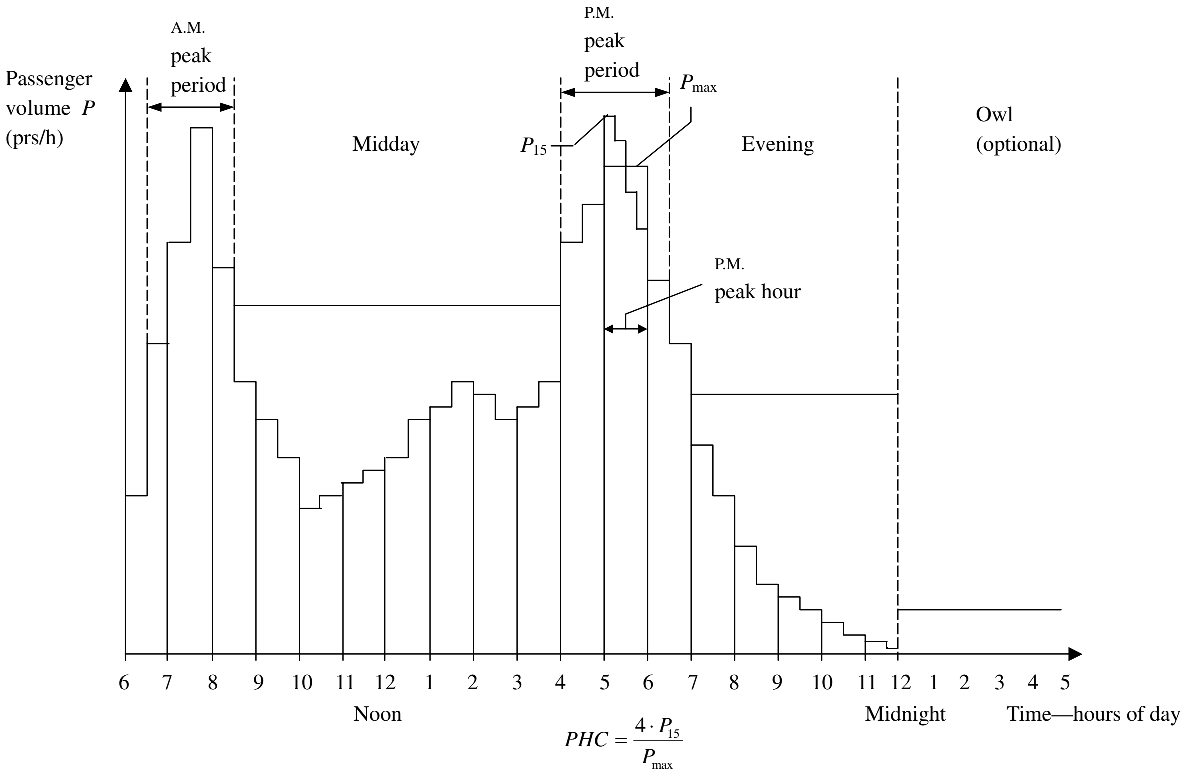


Figure 1.25 *Hourly variations of passenger volume on a line and its scheduling periods.*

scheduling is very important for providing good service on the passenger side, and achieving maximum operational efficiency on the transit operator side, rather sophisticated analyses are often performed. Two such refinements in scheduling are described here.

In some cases the MLS is not always the same station spacing during all scheduling periods. For example, a line may carry heavy commuter travel in the morning and afternoon, creating the MLS at the periphery of the CBD; midday shopper and lunch hour travel may make the MLS inside the CBD; and in the early afternoon the MLS may occur at a school in a suburban area. For such a case it is useful to make a tabulation or diagram of the line showing P_{max} for each hour by location along the line and its direction. The same table or diagram should indicate the offered capacity that is based on policy headways (see Sections 1.1.4 and 1.4.3). For every scheduling period, the higher of the two required capacities—the one determined by passenger volume and the other dictated by the policy-based level of service—should be used for developing the schedule.

Another refinement in analyzing demand is consideration of detailed variations in time. During the peak hours variations of passenger volumes are sometimes considerable, so that scheduling based on the total hourly volume may lead to both excessive crowding and oversupply of vehicles at different times within the same hour. To avoid this, passenger counts during the peaks must be made for 15-minute or even shorter time periods. Variations among these periods, as shown in Figure 1.24, are used to compute the peak hour coefficient (PHC), defined as the ratio of the highest 15-minute volume multiplied by four, and the total counted hourly volume on the MLS:

$$PHC = \frac{4\,p_{15}}{P_{max}} \qquad \left|\frac{PHC}{-}\right|\frac{p}{\text{prs/15 min}}\left|\frac{P}{\text{prs/h}}\right| \tag{1.41}$$

When the actual peak hour volume is multiplied by this coefficient, the hourly volume equivalent to the peak 15-minute volume is obtained, which should be used as the design volume:

$$P_d = P_{\max} \cdot PHC \qquad \left| \frac{P}{\text{prs/h}} \right| \frac{PHC}{-} \right| \tag{1.42}$$

Theoretically, values of PHC can vary from 1 to 4. PHC is the inverse of the peak hour factor (PHF), used for highway volume variations, which can have values between 0.25 and 1.

In some cases, particularly on heavily traveled lines operating with short headways, use of P_d obtained this way for scheduling the whole peak hour results in uneconomical operation. Instead, scheduling is made for each 15-minute (or even shorter) period based on its actual passenger volume.

Scheduling must be done for each scheduling period separately. Lengths of individual periods need not always coincide with hours. Transitions between scheduling periods should always be gradual, fitted to the change in demand. In the beginning of peak periods, capacity can be increased in two ways: by shortening headways or by increasing the capacity of TUs. A combination of the two may also be used.

The simplest way to shorten headways is to insert additional TUs between regularly scheduled ones. Thus, the headway is cut in half. This regularity is convenient for scheduling and provides simple operation, but when all TUs have equal capacities, it represents an abrupt transition, providing double capacity rather than incremental increases. Reduction of headways to other than one-half is therefore often made. For example, a basic headway of 10 minutes is reduced to 7.5 or 6 minutes during the peak. The inconvenience to passengers of having to memorize a different schedule is not serious when headways are short. For long headways (>10 minutes) such a change must be clearly indicated in timetables.

The second technique for transition to higher capacity is often used on rail systems. A rapid transit line operating with four-car trains throughout the day can be changed into an operation with six- or eight-car trains during the peak hours by coupling additional cars at terminals or intermediate stations. With buses and trolleybuses this technique cannot be employed, except in special cases when standard buses are used for regular service, but larger-capacity vehicles (e.g., articulated) are used during the peak hours. The changes of fleet involve additional deadheading, however.

1.4.2.3 The Required Capacity. On lightly traveled lines and during off-peak hours, the capacity offered on a line is usually dictated by the policy headway, i.e., the volume of offered service is determined by the required service headway, rather than by passenger volumes. However, when passenger volume is heavy, capacity offered must cover passenger load profile, as shown in Figure 1.24. The level of offered capacity is determined by selection of the maximum value for the load factor α, i.e., by capacity utilization on MLS. This selection is based on the trade-off between the two basic requirements for transit operations:

- For passenger comfort and convenience, lower values of α result in less crowding, higher availability of seats, and more frequent service.
- Cost of operation is lower when a higher value of α is adopted, because smaller number of vehicles is required to transport a given passenger volume.

Several other factors may influence the selection of the maximum value for the load factor on the MLS, $P_{\max}$. For example:

- Rather uniform passenger volume along the line suggests lower α; uneven distribution allows higher α, because the maximum volume is found only on a short section of the line.
- High ratio of seated to standing passengers foreseen by vehicle design requires lower α (areas with seats cannot accept any overcrowding, only standing areas can).
- Long average trip length (e.g., express bus, regional rail) implies low passenger turnover rate; therefore, higher comfort should be provided—α should be lower.
- High percentage of senior citizens or shoppers with packages dictates lower α, larger volumes of school children, higher α.

- For transporting large crowds of passengers to special events, such as sport stadia, holiday events in the city, major concerts, etc., the maximum achievable value for α can be used.

As this review of different operating factors shows, the value of the load factor α is a result of the trade-off between two mutually conflicting requirements: passenger comfort, requiring lower α, and lower operating costs, dictating higher α. Other mentioned conditions may influence the selection of the load factor's value in this trade-off.

1.4.2.4 Selection of TU Size, Frequency, and Load Factor. Figure 1.26 presents a diagram that can be used to select the optimal combination of TU size n, service frequency f or headway h, and load factor α for any scheduling period of day. The diagram is based on the equation

$$C = n \cdot C_v \cdot \alpha \cdot f = \frac{60 \cdot n \cdot C_v \cdot \alpha}{h}$$
$$\left|\frac{C}{\text{sps/h}}\right|\frac{n}{\text{veh/TU}}\left|\frac{C_v}{\text{sps/veh}}\right|\frac{\alpha}{\text{prs/sp}}$$
$$\left|\frac{f}{\text{TU/h}}\right|\frac{h}{\text{min/TU}}\Bigg| \qquad (1.43)$$

The equation shows four different train consists: TU sizes of two, four, six, and eight cars, and the line capacity they can provide with operations at different

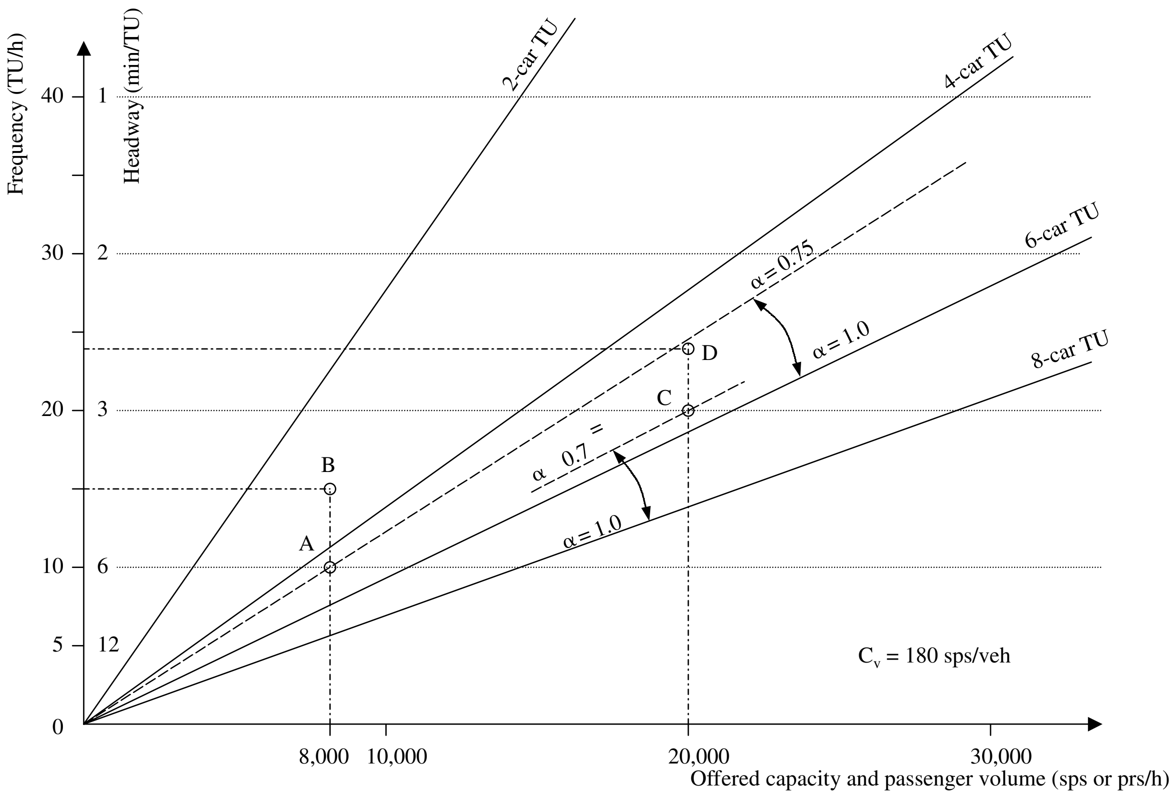


Figure 1.26 *Diagram for selection of TU size and headway for different scheduling periods.*

frequencies/headways and with different values of the load factor α. Each sloped line shows the values for a given TU size at full occupancy, $\alpha = 1.00$. A dashed line shows the capacities offered by six-car TUs with $\alpha = 0.75$.

This diagram can be used as follows. If during midday period $P_{max} = 8{,}000$ prs/h, reasonable choices would be to operate six-car TUs at $h = 6$ min with $\alpha = 0.75$—point A on the diagram; or four-car TUs at $h = 4$ min with $\alpha = 0.75$—point B. Suppose that the peak period has $P_{max} = 20{,}000$ prs/h; then the choices may be to operate eight-car TUs at $h = 3$ min and $\alpha = 0.70$—point C, or six-car TUs at $h = 2.5$ min and $\alpha = 0.77$—point D on the diagram.

1.4.3 Scheduling Procedure

There are many variations in the scheduling procedure, depending on the time period (peak, off-peak, owl, etc.), methods of bringing vehicles into and out of service, crew practices, etc. The basic physical relationships among operational elements are valid in all cases, however; scheduling for a regular case is presented here, with practical examples of the computational sequence.

The design hour volume P_d divided by the average number of passengers a TU will carry on MLS gives the required frequency:

$$f = \frac{P_d}{\alpha \cdot n \cdot C_v} \quad \left|\frac{f}{\text{TU/h}}\right|\frac{P_d}{\text{prs/h}}\left|\frac{\alpha}{\text{prs/sp}}\right|\frac{n}{\text{veh/TU}}\left|\frac{C_v}{\text{sps/veh}}\right| \tag{1.44}$$

and its inverse, the headway:

$$h = \frac{60}{f} = \frac{60\alpha \cdot n \cdot C_v}{P_d} \quad \left|\frac{h}{\text{min/TU}}\right|\frac{f}{\text{TU/h}}\left|\frac{\alpha}{\text{prs/sp}}\right|\frac{n}{\text{veh/TU}}\left|\frac{C_v}{\text{sps/veh}}\right|\frac{P_d}{\text{prs/h}}\right| \tag{1.45}$$

When $h \leq 6$ minutes, passengers are not particularly concerned about exact departure times, so that any value of h is acceptable, including, on precisely operated lines, such values as 1½ or 2½ minutes. However, if the computed value of $h > 6$ minutes, it should be rounded down to the nearest clock headway (see Section 1.1.4). This h should then be compared with the adopted policy headway h_p, and the smaller of the two should be used.

The actual values of f (which is an integer when $h > 6$ minutes) and α, after the rounding of h, are computed from Eq. (1.45):

$$f = \frac{60}{h} \quad \text{and} \quad \alpha = \frac{P_d \cdot h}{60\, n \cdot C_v}$$

To compute the number of vehicles needed to provide service with a given headway, cycle time T must be found by computing its components. *Operating time* T_o for each direction must be computed by field surveys. *Terminal time* t_t at each terminal depends on four factors, as explained in Section 1.1.6.1. For street transit systems, crew rest and delay recovery usually govern, while rapid transit has a constant terminal time (usually 5–10 minutes) irrespective of line length. Alternatively, instead of t_t, terminal time coefficient γ, expressing t_t as a percent of T_o, can be used for this preliminary computation of T:

$$T = 2(T_o + t_t) = 2\,T_o(1 + \gamma) \quad \left|\frac{T, t}{\text{min}}\right|\frac{\gamma}{-}\right| \tag{1.46}$$

here γ can be given a value of 0.1–0.3.

The number of TUs ($N_{TU} = N/n$) on the line is obtained as the ratio of the cycle time and headway:

$$N_{\text{TU}} = \left[\frac{T}{h}\right]^+ = \left[\frac{f \cdot T}{60}\right]^+ \quad \left|\frac{N_{\text{TU}}}{\text{TU}}\right|\frac{T}{\text{min}}\left|\frac{h}{\text{min/TU}}\right|\frac{f}{\text{TU}/h}\right| \tag{1.47}$$

The brackets indicate integer value equal to or greater than the computed value. In the following equations it will be assumed, for simplicity, that single vehicles are operated, $n = 1$, so that $N_{TU} = N$.

Since h is fixed, T must be adjusted to be

$$T = h \cdot N_{TU} \qquad \left|\frac{T}{\text{min}}\right|\frac{h}{\text{min/TU}}\left|\frac{N_{\text{TU}}}{\text{TU}}\right| \tag{1.48}$$

by increasing t_t (at one or both terminals). The adjusted value of T represents the final cycle time. The actual value of γ is recomputed with the new t_t by Eq. (1.20).

If it is desirable to minimize the number of TUs in service, terminal times can be kept at their minimum values by using two different headways, usually varying by one minute, satisfying the following equation:

$$k_1 \cdot h_1 + k_2 \cdot h_2 = T \tag{1.49}$$

where k's represent the numbers of the two respective headways in a cycle.

For example, the required headway is computed to be 5.4 minutes and the initial cycle time is 81 min. Ideally, a uniform headway of 5 minutes should be used, T should be rounded up to a multiple of h, i.e., to 85 minutes, and $N = 85/5 = 17$ TUs. However, it is possible to use nine 5-minute and six 6-minute headways, adding up to the cycle time of 81 minutes (Eq. 1.49): $9 \times 5 + 6 \times 6 = 81$.

The number of TUs is $9 + 6 = 15$, i.e., two TUs less than if uniform headways were used, and the average headway is exactly $81/15 = 5.4$ minutes.

Cycle speed V_c can also be related to operating speed via the terminal time coefficient γ:

$$V_c = \frac{V_o}{1 + \gamma} \qquad \left|\frac{V}{\text{km/h}}\right|\frac{\gamma}{-}\right| \tag{1.50}$$

but its most important relationship is with the number of TUs on the line, because that determines the cost of providing the service. From Eqs. (1.29) and (1.48), one obtains:

$$V_c = \frac{120L}{T} = \frac{120L}{h \cdot N_{\text{TU}}}$$

$$\left|\frac{V}{\text{km/h}}\right|\frac{L}{\text{km}}\left|\frac{T}{\text{min}}\right|\frac{h}{\text{min/TU}}\left|\frac{N_{\text{TU}}}{\text{TU}}\right| \tag{1.51}$$

From this equation and Eqs. (1.44) and (1.50) N_{TU} can be expressed as:

$$N_{\text{TU}} = \frac{120\ L}{h \cdot V_c} = \frac{2\ L \cdot P_d(1 + \gamma)}{\alpha \cdot n \cdot C_v \cdot V_c}$$

$$\left|\frac{N_{\text{TU}}}{\text{TU}}\right|\frac{L}{\text{km}}\left|\frac{h}{\text{min/TU}}\right|\frac{V}{\text{km/h}}\left|\frac{V}{\text{km/h}}\right|\frac{P}{\text{prs/h}}\left|\frac{\gamma}{-}\right.$$

$$\left|\frac{\alpha}{\text{prs/sp}}\right|\frac{n}{\text{veh/TU}}\left|\frac{C_v}{\text{sps/veh}}\right| \tag{1.52}$$

The first of these expressions, N_{TU} as a function of V_c, is plotted in Figure 1.27. The diagram shows that cycle speed strongly influences the required number of TUs, particularly at low speeds and for short headways. The second expression shows that, as is intuitively clear, the fleet size is directly proportional to the line length and passenger volume, while it is inversely proportional to TU capacity and operating speed. These elements, together with coefficients α and γ, therefore have a direct bearing on the cost of operation.

1.4.4 Procedure Summary, Examples, and Numerical Schedules

The preceding computational procedure is summarized here into five steps and illustrated by examples. It is assumed that the data on passenger volumes have been collected and analyzed and uniform scheduling periods determined. Four different cases are illustrated numerically:

- **a.** A street transit route operated by 45-seat buses for peak periods
- **b.** The same route for base periods
- **c.** A rapid transit line operated by vehicles with total (seats plus standing) capacity of 140, in trains of up to six cars, for peak hours

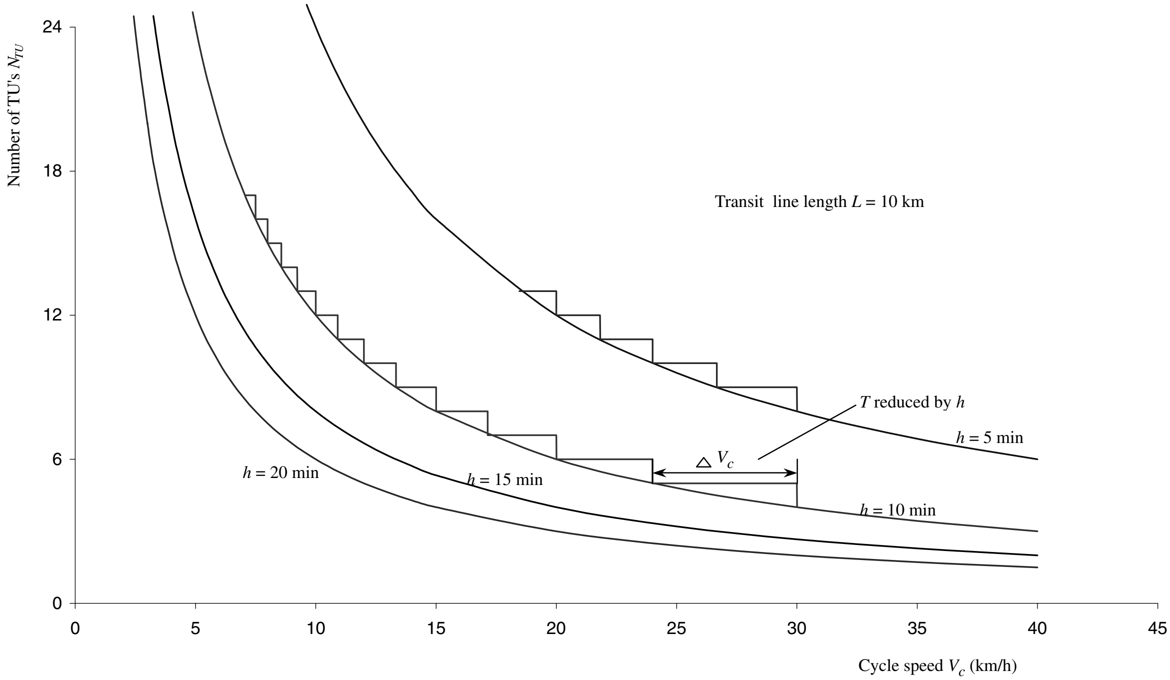


Figure 1.27 *Number of TUs on a line as a function of cycle speed and headway.*

d. The same line during base periods, minimum train consisting of a married pair

Scheduling steps are defined in a brief form. The data and computed values for the example cases a–d are presented in Table 1.5.

Step 1: Prepare data and determine factors.

- Line length: L (km)
- One-way operating time: T_0 (min)
- TU capacity: C_v (sps/veh) and n (veh/TU)
- Policy headway: h_p (min/TU)
- Load factor (on MLS): α (prs/sp)
- Design volume: P_d (prs/h) (includes PHC)
- Minimum terminal time: t_t (min) or minimum value of γ

Step 2: Compute headway and frequency. Headway is obtained directly from Eq. (1.45):

$$h = \frac{60 \cdot \alpha \cdot n \cdot C_v}{P_d}$$

if the computed $h > 6$ minites, round it down to the nearest smaller of the following numbers: 6, 7.5, 10, 12, 15, 20, 30, or 60 (sometimes 40 minutes is also used, repeating the times every two hours).

Compare the obtained headway with h_p and adopt the shorter one.

Compute frequency and the actual value of α using expressions based on Eq. (1.45):

$$f = \frac{60}{h} \quad \text{and} \quad \alpha = \frac{P_d \cdot h}{60 \cdot n \cdot C_v}$$

Step 3: Determine fleet size. Introduce the known T_o and t_t or γ in the first or second of the following expressions, Eq. (1.46):

$$T' = 2\,(T_o + t_t) = 2\,T_o(1 + \gamma)$$

Table 1.5 Examples of scheduling computations

Step	*Item*	*Symbol*	*Dimension*	*Source, Eq. ( )*	*Cases* *a* *Bus-peak*	*b* *Bus-base*	*c* *RT-peak*	*d* *RT-base*
1	Line length	L	km	given	8	8	12	12
	Vehicle capacity	C_v	sps/veh	given	45	45	840	280
	Max. passenger volume	P	prs/h	given	980	160	10,000	1500
	Operating time	T_o	min	given	40	30	24	24
	Load factor	α	prs/sp	given	1.1	0.9	0.8	0.6
	Policy headway	h_p	min/TU	given	5	12	5	10
	Terminal time coefficient	γ, (t_t)	–, (min)	given	0.18	0.15	(6)	(6)
2	Headway	h	min/TU	(1.44)	3	15	4	6
	Min (h, h_p)	h	min/TU		3	12	4	6
	Frequency	f	TU/h	(1.45)	20	5	15	10
	Actual load factor	α	prs/sp	(1.45)	1.09	0.71	0.79	0.54
3	Approx. cycle time	T'	min	(1.46)	95	69	62	62
	Fleet size	N	veh	(1.47)	35	6	16×6	11×2
4	Cycle time	T	min	(1.48)	96	72	64	66
	Terminal time	t_t	min	(1.52)	8	6	8	9
5	Cycle speed	V_c	km/h	(1.50)	10.0	13.3	22.5	21.8
	Summary data:							
	h		min/TU		3	12	4	6
	α		prs/sp		1.09	0.71	0.79	0.54
	N		veh		32	6	96	22
	T		min		96	72	64	66
	V_c		km/h		10	13.3	22.5	21.8

to obtain an approximate value of T, called T'. With that T', compute fleet size using Eq. (1.47):

$$N_{TU} = \left[\frac{T'}{h}\right]^{+}$$

rounding the obtained N to the next higher integer.

Step 4: Compute cycle and terminal times by Eq. (1.48): $T = h\ N_{TU}$, using the rounded N_{TU}, and

$$t_t = \frac{T - 2\ T_o}{2} \qquad (1.53)$$

The obtained t_t represents the average value of terminal times at the two terminals, which may not be equal.

Step 5: Compute cycle speed, using Eq. (1.29):

$$V_c = \frac{120\ L}{T}$$

Computations in Table 1.5 follow this sequence of steps.

The final products of the scheduling process are numerical and graphical presentations of transit line operation that are made for use by both the operator and the public. The numerical schedule for operation of each run, i.e., running of a TU from its leaving the depot to its return, is used by individual drivers and dispatchers (see Section 1.4.6). A tabulation of times for all TU runs is given to the line supervisors, while various public timetables are prepared in appropriate formats.

1.4.5 Graphical Presentations of Transit Operations

Graphical presentations, such as time-distance, time-speed, distance-speed, and time-energy consumption diagrams, can be used very effectively in planning, operations (scheduling), and analyses of transit systems.

They offer, for example, a much better overview (such as regularity of headways), as well as details (meeting locations) of a line operation, or impacts of schedule changes, than a numerical table can provide.

While there were speculations that use of computers in scheduling and various transit operations analyses would make graphical presentations obsolete, the opposite has happened: with computer plotter capabilities it has become easier to generate various diagrams and figures. Innovative types of plots and diagrams offer clearer understanding of operations than either mathematical equations or numerical tables. Thus, computers make complex types of graphical presentations easy to plot and make them very useful in operations.

1.4.5.1 Graphical Schedules for Single Lines. There are several types of graphical schedules, from the conventional time-distance diagrams (sometimes known in the transit industry as string charts) to presentations of schedules on complex multiline networks. They can also vary in the manner of plotting and degree of detail.

The basic graphical schedule is a *time-distance diagram,* which has the line (distance) plotted on the ordinate and time on the abscissa. As shown in Figure 1.28, the line is divided in sections with uniform speeds. The plot of every run of a TU, designated by a number, shows all schedule elements—travel times, speeds, etc.—for that TU on each section and at each terminal. The abscissa also shows headways as distances between subsequent TU runs, and cycle times as distances between two subsequent departures of the same run from the terminal.

The entire diagram shows TU arrivals/departures at each reference point along the line, terminal times, as well as locations and times where TUs meet. Short turns, those operating on a section of the line, can be plotted and their fitting in regular headways in both directions can be checked.

To avoid different slopes on the diagram, simplify the plotting, and improve readability, it is common to use a *real-time/operating-time diagram,* which has operating time between the two terminals instead of distance plotted on the ordinate. This diagram can be understood as one on which distance scale is stretched on slower sections until a straight line is obtained for the entire line; thus, the ordinate presents the line section lengths at different scales, depending on their speeds, or it presents operating time at a constant rate (therefore the designation time-time diagram). Figure 1.29 shows the schedule from Figure 1.28 as a time-time diagram.

With respect to scale and plotted details, graphical schedules may vary greatly. For example, the schedule for a light rail line in Figure 1.30 shows track layout with single and double track sections, pull-outs and pull-ins of TUs from depots for operations on some line sections, different terminal times, etc. Some diagrams, such as those for highly controlled rapid transit systems, particularly the automated ones, show precise times for stopping at every station. Large-scale schedules, which may show long sections of lines as the basic blocks only, are used for plotting networks with overlapping and branch lines.

1.4.5.2 Trunk Lines with Branches. For complex transit lines, such as a trunk line with branches, graphical schedules can be even more useful than for single lines. For example, for a line consisting of a trunk with two branches, each branch must be scheduled by itself, and then the joint schedule of the two lines on the trunk section must be checked with respect to service regularity as well as the total offered capacity. This relationship can be seen very clearly on a graphical schedule. For further details on trunk-and-branch scheduling, see Sections 2.5.3 and 4.3.3.

A graphical schedule for a line with two branches, A and B, is shown in Figure 1.31. It consists of a diagram of the trunk TX, consisting of three sections with different speeds, which continues directly into one branch, XA. Vertically separated, but plotted to the same real-time axis, is the diagram for the second branch, XB. Plots for TUs on line A continue from the trunk directly into that branch, while those to branch B are vertically transferred from point X to the XB diagram.

This plot shows TU movements and headways on all three sections, the trunk and the two branches; it also shows how the terminal times on the two branches must be coordinated if the trunk is to have uniform headways.

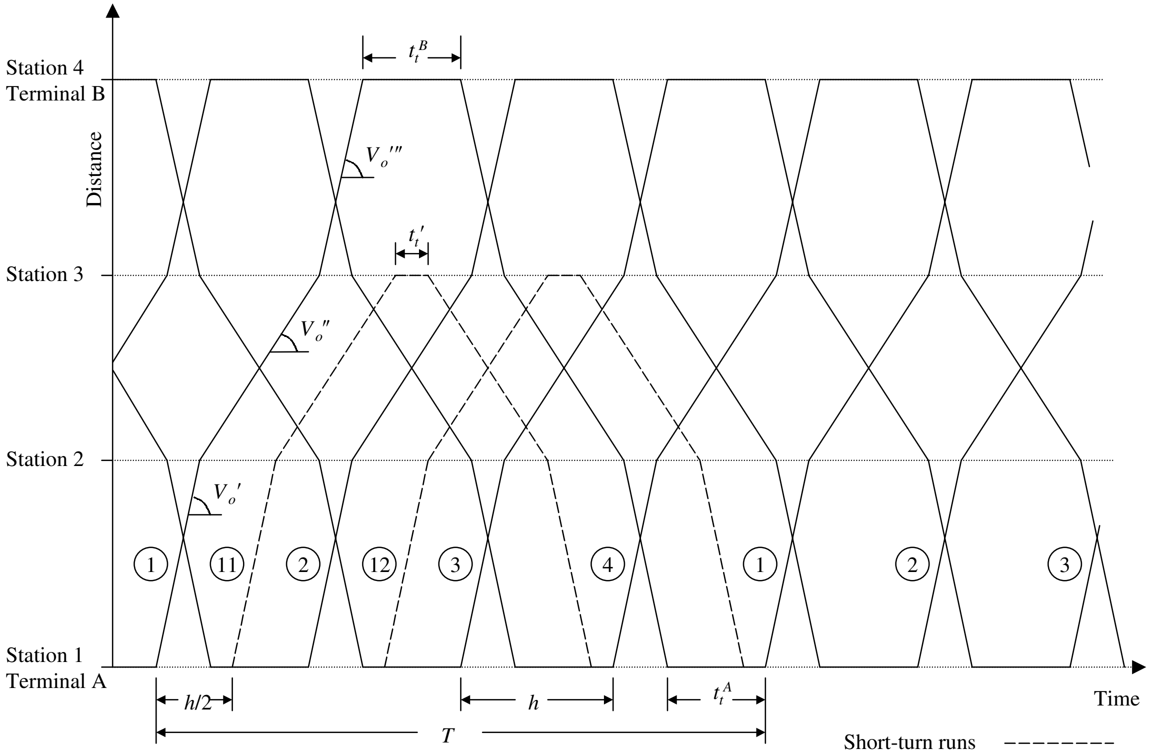


Figure 1.28 *Graphical schedule for a transit line with regular and short-turn runs.*

The same method of graphical schedule presentation can be used for a set of diametrical lines with a joint trunk section, typical for regional rail lines (Munich, Philadelphia). Such a diagram is shown in Figure 1.32. The basic purpose of such a plot is to present the relationship of TUs from different lines on the trunk section. Therefore, the branches can be shown only schematically, at a highly reduced scale, plotting only the entire travel time on each one.

This diagram should then be complemented by separate larger-scale diagrams of individual lines (such as the one in Figure 1.29), which have all the necessary details of sections, stations, terminal times, or short-turn runs.

1.4.5.3 Interconnected and Overlapping Lines. As the complexity of transit network increases, graphical schedules may become the only way to show the relationships of schedules clearly on joint sections as well as on individual branches.

An excellent example of the use of graphical schedule for a complex interconnected network is the San Francisco Bay Area Rapid Transit (BART) system. Figure 1.33 shows this network in 1996, when it consisted of a trunk line—M (Market Street)—and three branches—R (Richmond), C (Concord), and F (Fremont). The network was operated as three trunk/branch lines—MR, MC, and MF—and a fourth line between two of the branches, RF. The core point in the network is the Y junction in Oakland, where the three branches merge and the RF line crosses in the north-south direction, forming the triangular, Y connection. The MR and MC lines proceed jointly to the north, branching out at MacArthur Station (MA). Subsequently, in the late 1990s, several extensions as well as another branch, to Dublin/Pleasanton, opened, making the operations even more complex. The diagrams

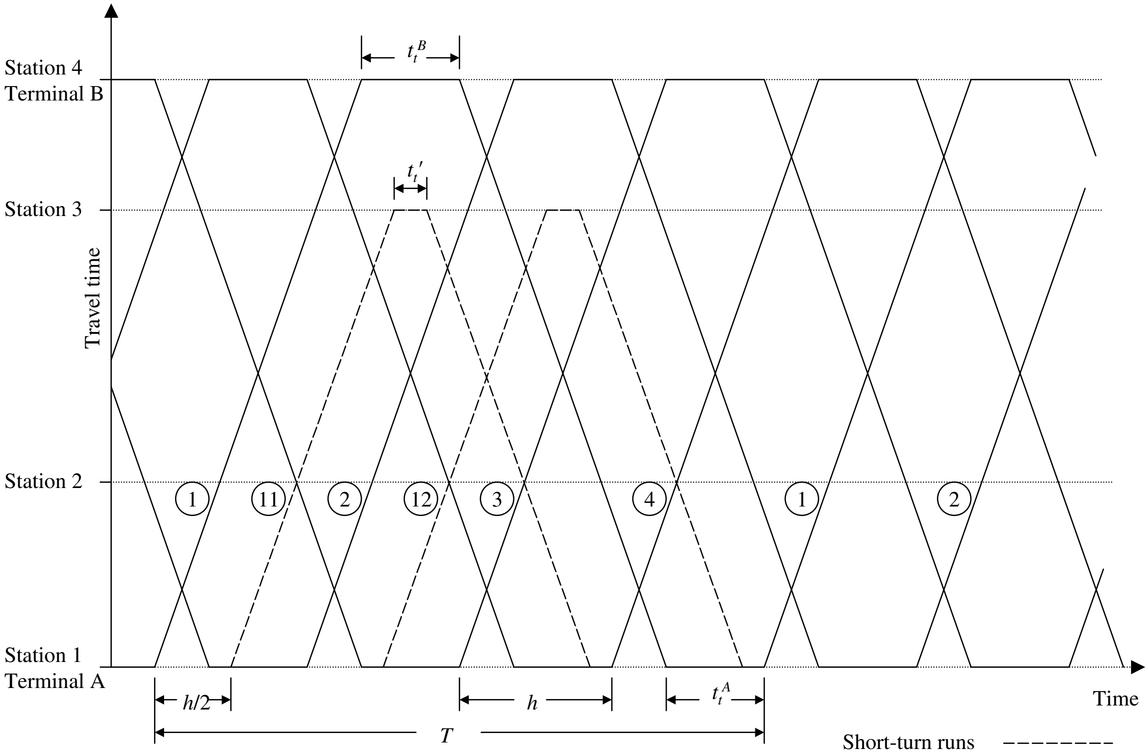


Figure 1.29 *Real-time/travel time schedule diagram for the line from 1.28.*

shown here are kept for the network prior to these extensions to avoid excessive schedule density that does not introduce any new concepts.

Scheduling of the BART network is quite complex because uniform headways, which are desirable on the trunk, impose certain relationships of headways on the branches, which then constrain the schedule for the cross-branch RF line. If the trunk headway is h_t and all branches should have equal headways, each one must have a headway of $3h_t$, and the sequence of trains on the trunk, or the basic module of the schedule, must be MR-MC-MF. If one line (in this case, MC) is more loaded than the others, another option is to operate the trunk with the basic module MC-MF-MC-MR. Then line MC has headways $2h_t$, while lines MR and MF have headways equal to $4h_t$. When combined with the cross-branch line RF, the average joint headway on each branch, R and F, also equals $2h_t$.

The headways on the branches, determined in this manner by the trunk operation, must also be synchronized with the RF line schedule. This must be done by an analysis of train arrivals/departures at the Y junction, because these determine the headways on all three directions radiating from it.

The relationships of headways on different lines and their regularity depend on the geometric setup of lines, headways that must be offered on each one, and other elements. It is therefore not always possible to provide uniform headways on all sections of the network.

Having determined the optimal relationship of schedules on all four lines, it is now necessary to schedule each line for its complete cycle, to determine terminal times and the required number of TUs. For this, a time-time diagram is used, again showing one branch as continuation of the trunk and the other branches on separate sections (Figure 1.34).

A special feature in this case is the plot of the RF line trains, i.e., those between two branches: inbound trains from R are shown as parallel to the inbound RM trains to the Y, but then their plot continues on the F

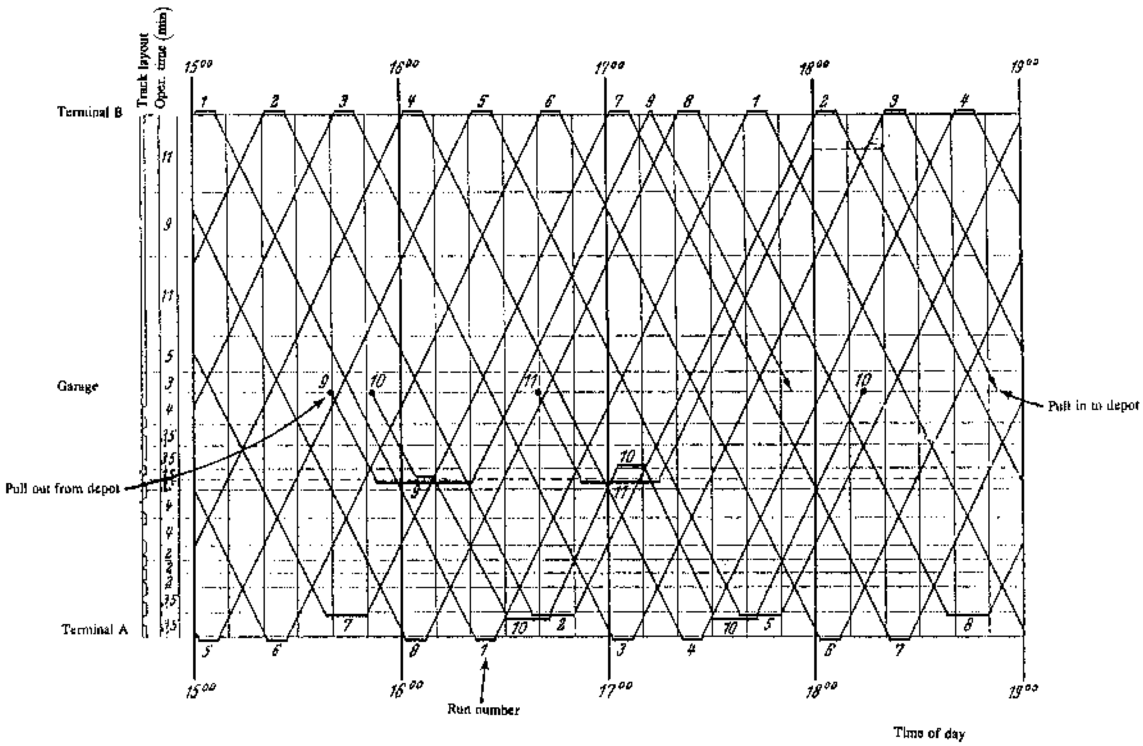


Figure 1.30 *An actual time-time graphical schedule with single track sections and operational details.*

branch as outgoing trains, i.e., with the opposite slope parallel to the MF trains. Thus, triangular interconnected lines can also be presented on a graphical schedule, and such a diagram gives an excellent overview of services on all network sections and their relationships, regularity of headways, terminal times, meeting points of trains, points critical for capacity, etc.

1.4.5.4 Other Types of Graphical Presentations. A number of other types of diagrams can be used effectively in analyses of transit operations. The most important ones are described here.

A *time-distance diagram* for different stopping regimes allows analysis of such operations as local (all-stop), skip-stop, express, and zonal, presented in Section 2.4. Plotting of different stopping schedules by subsequent TUs allows an easy overview of headways, travel times, meetings, transfer possibilities and overtaking points along a line, etc.

A *speed-time diagram* is the most useful for analysis of driving regimes and travel times along a line. This is presented in considerable great detail in Vuchic (1981), Chapter 3.

A *speed-distance diagram* is useful in planning and analysis of TU travel along a line because it shows speed at any one point. The speed profile is used for determining signal blocks on rail lines, positions of different signals, braking distances, etc.

An *energy consumption-time diagram* shows consumption of energy for traction during different regimes and for a station-to-station travel cycle. As explained in Vuchic (1981), Chapter 3, this diagram can be used to further develop a diagram of travel time versus energy consumption on a given station spacing. This relationship is important in planning operating re-

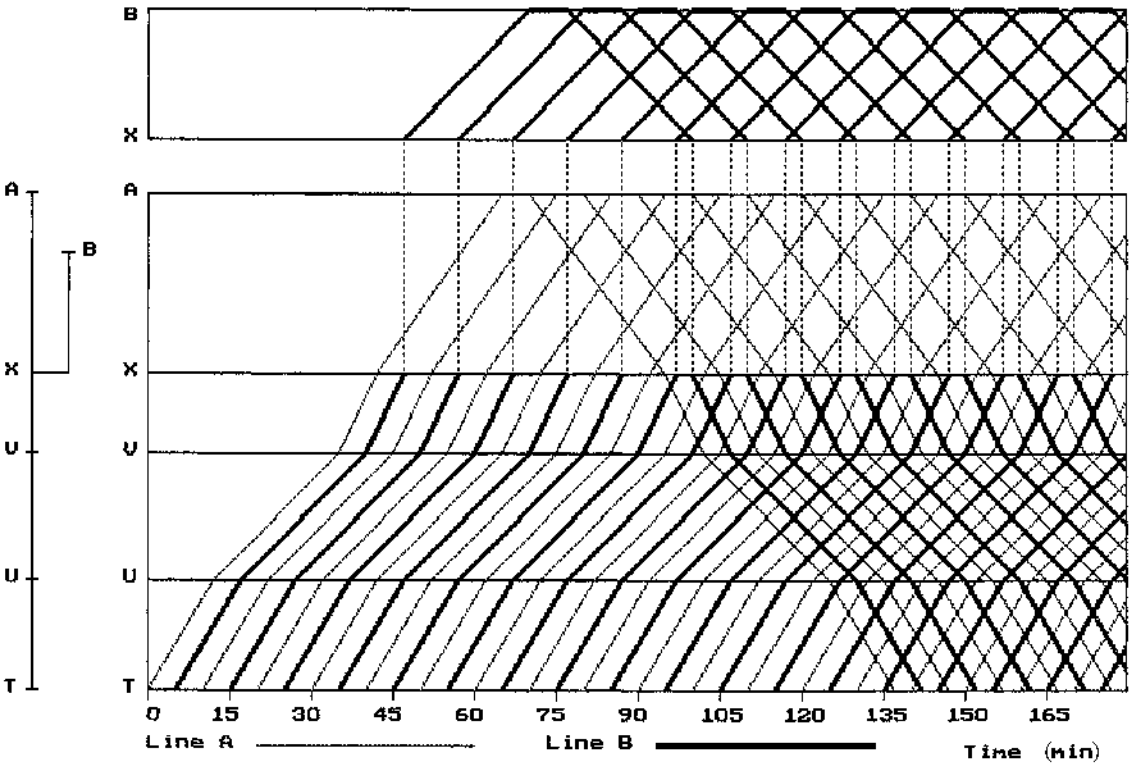


Figure 1.31 *A time-distance graphical schedule for a trunk line with two branches, A and B.*

gimes for electrically powered guided systems. Some rapid transit systems (e.g., BART and Washington Metro) have computer-controlled train operation which optimizes travel time-energy consumption relationship for different conditions.

1.4.6 Crew Scheduling or Run-Cutting

Assigning drivers or train crews to scheduled sequence of operations or TU runs is commonly known in the transit industry as *crew scheduling* or *run-cutting.* The task in run-cutting is to assign personnel to a given schedule of TU operations in such a manner that the total expenditure for wages is minimal, while satisfying the various operating and work rules set up by the agency or specified in its agreement with employee labor union.

Transit is a service that usually must be offered every day of the year for 16, 20, or even 24 hours. Since the quantity of service offered varies greatly among different days and hours of day, transit operating personnel must work during irregular hours, for periods of varying length, sometimes during nights, on weekends, and on holidays.

1.4.6.1 The Basic Concepts. The sharp peaking of passenger volumes that typically occurs on transit systems each working day in the morning and afternoon hours, about nine hours apart, requires different vehicle schedules for several periods during the day, as described in Section 1.4.2.2. The run-cutting—preparing schedules for drivers—is usually even more complex because sets of work assignments must consider workers' needs and meet established work rules and contractual standards.

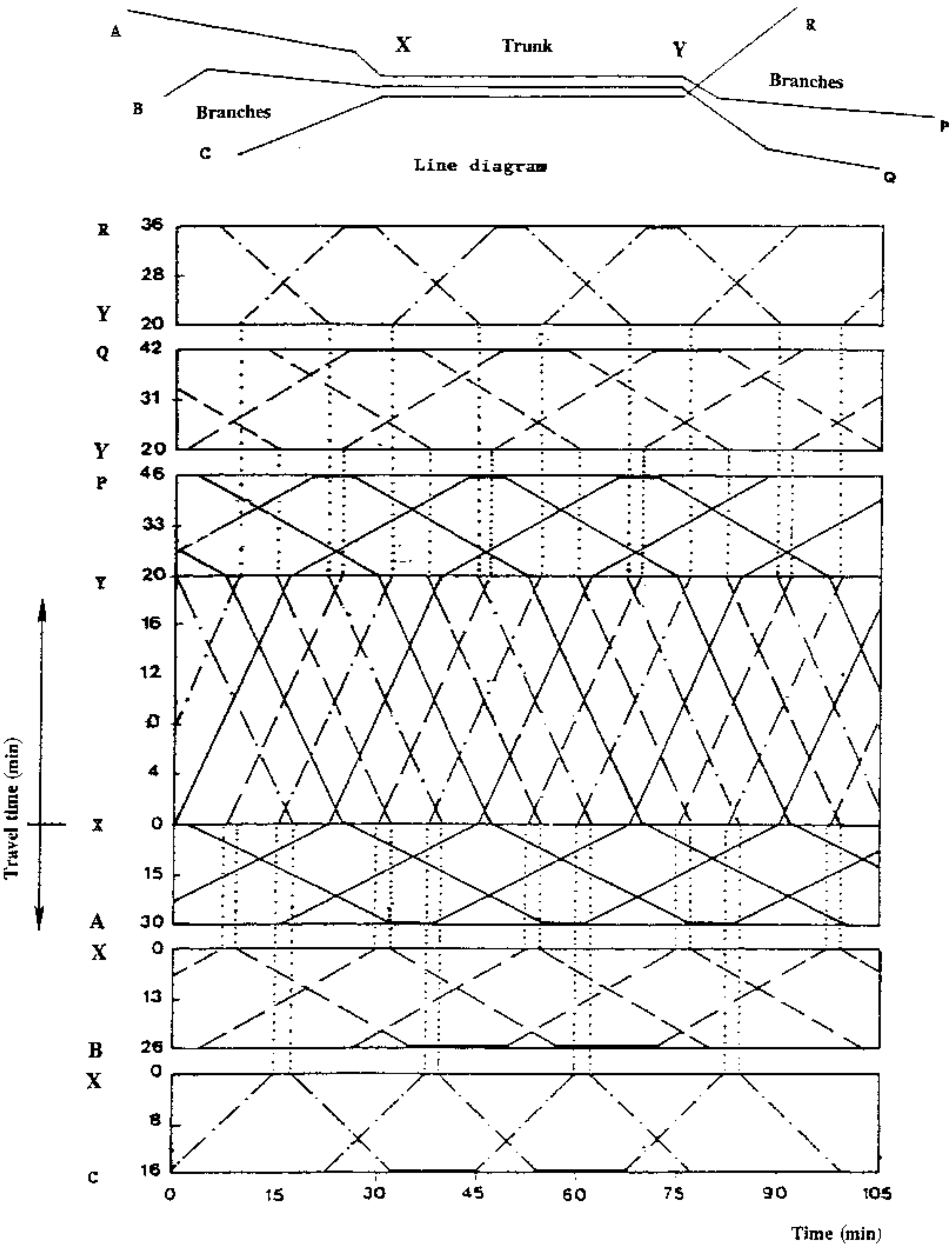


Figure 1.32 A graphical schedule for diametrical lines with a joint trunk.

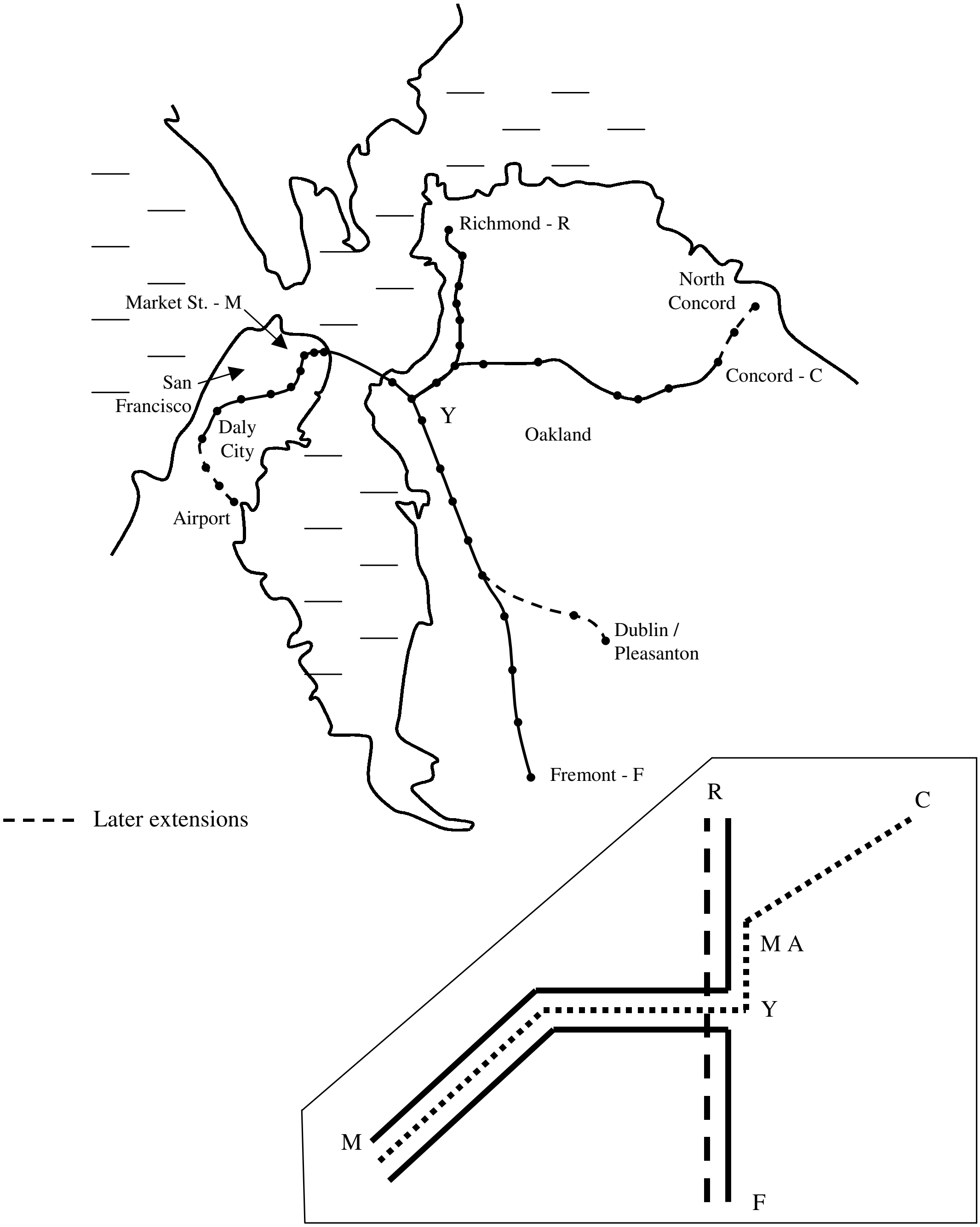


Figure 1.33 *San Francisco BART network and schematic presentation of its lines.* Note: *The scheduling is shown for the network from 1995, with later extensions as dashed lines.*

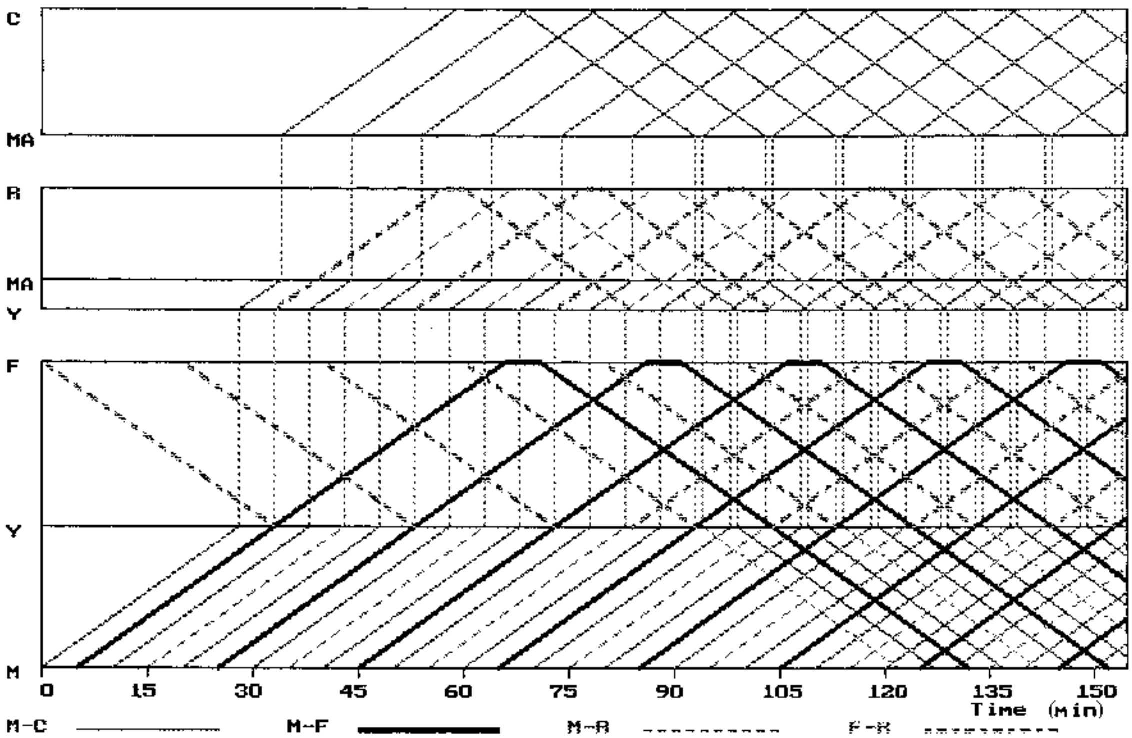


Figure 1.34 *Graphical schedule diagram for a trunk with three branches and the cross-branch line* (*San Francisco BART network*).

To cover the uneven demand for drivers during the daily hours on weekdays and weekends, working times are tailored in different forms. The basic and most desirable work shift or run is a continuous work shift, usually of eight hours. That is called a *straight run.* Then, since many more drivers are needed during peak than during off-peak periods, many drivers must be scheduled to work during the two peaks with several unproductive (paid or unpaid) hours between. Such a work shift is called a *split run*. The time interval between the beginning of work in the morning and its termination in the afternoon is referred to as *spread time*. Short time segments of work are called *extras* or *trippers.*

Work rules that the schedule must satisfy are determined in different ways. In some cases they are specified by the transit agency as a set of operating practices or standards, or spelled out in the contract between the agency and the labor union of its employees; further, there may be recommended or required practices developed by an association of agencies for the entire transit industry; finally, there are national government rules and regulations for maximum duration of driving between rests. Germany, for example, has extensive, carefully developed guidelines and regulations for work rules (as well as for safety, vehicle maintenance, and many other aspects of transit operations).

Consequently, there are wide variations in work rules and payment standards. For U.S. and European transit agencies, it is rather common for regular pay to be given for 8 hours of work, which are either continuous or within a 12-hour-spread time interval. Higher wage rates are usually paid for some or all of the following conditions:

- Overtime
- Longer spread time than prescribed
- Night shift
- Work on holidays

Alternatively, the contract may stipulate the frequency with which shifts involving night, holiday, or overtime work can be given to any one employee, after which a higher-than-regular wage rate must be paid. For example, the maximum percent of split runs, or ratio of split to straight runs, is often stipulated. When that is exceeded, an employee gets extra compensation.

Transit agencies usually have a scheduling department that produces schedule blocks for vehicles/trains and for crews. Specific assignments of individual workers to run-cutting blocks, which are then placed into work-assignment sheets, are usually done by individual rail yards and bus garages. This decentralization is more efficient because assignments can be more personalized at the local levels, to the extent that in some agencies drivers are always given the same vehicles, so that they tend to take better care of them. Moreover, work rules sometimes vary among yards and garages, particularly in such large agencies as those in London and New York, where there may be 10–20 rail yards and 50–100 bus garages.

1.4.6.2 Run-Cutting Procedure. The procedure of work assignments to employees for individual runs vary among agencies. In most U.S. transit systems, the runs are selected (picked) in the sequence of employees' seniority, so that workers with longest tenure can choose the most convenient runs. In most European transit systems, runs are programmed in a rotational manner, so that every worker gets generally the same overall schedule within a period of several weeks. An extremely systematic and detailed description of the methodology for making such rotating ("turnus") schedules, as well as analysis of work-assignment efficiencies, is given by Lehner (1978).

Various practices and requirements, together with many local conditions in each agency, make the task of achieving the minimum-cost driver assignments a very complicated one. There is no exact sequence of mathematical formulae or procedures for finding the optimal solution, but the basic procedure for run-cutting, developed through experience, consists of the following sequence of steps:

1. Develop as many straight runs as possible.
2. Form split runs within spread time.
3. Divide some straight runs into two or three segments (extras) and combine these segments with the extras left over from step 2 to form additional split runs.
4. Analyze the efficiency (basically, total expenditure) of the obtained solution. If it is not satisfactory, investigate possibilities for pairing with other routes so that TUs alternate between them, and repeat the procedure.

Figure 1.35(*a*) shows a typical diagram of operating hours that have to be served on a transit line. Actual numbered runs for such a service are shown in Figure 1.35(*b*), while the results of that run-cut are listed in Table 1.6.

If the peak-to-base ratio, i.e., the ratio of the number of TUs in operation during the peak hour to their number during midday period, is very large and the two peaks are spread far apart, it may be particularly advantageous to employ part-time workers. They can be used to cover one of the peaks, avoiding the runs with excessive spread times (>12 hours), which are inconvenient for workers and can be costly for the agency. Hiring part-time workers has other advantages and disadvantages (lower benefit payments but extra training costs) and is usually subject to agreement with the labor union.

Run-cutting procedures are also presented by Lehner (1978), Homburger (1982), Hoffstadt (1981), and Rainville (1982).

1.4.7 Use of Computers in Scheduling

The entire process of scheduling, including computations of headways, TUs, cycle times, etc., and then

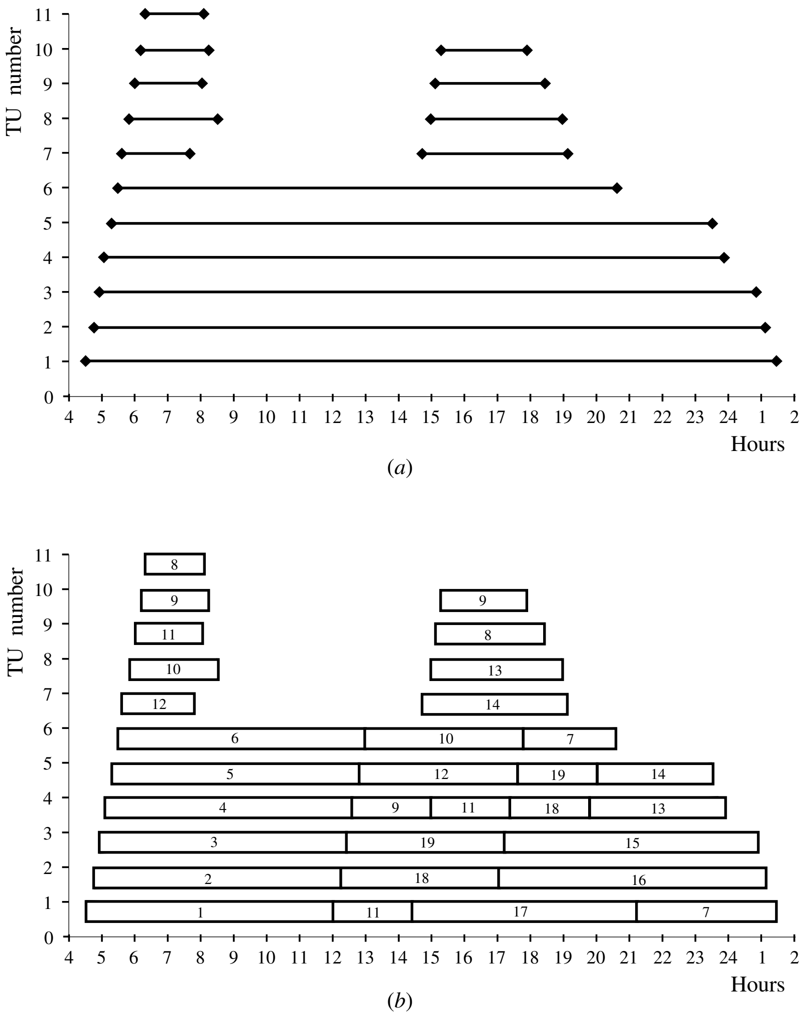


Figure 1.35 *Diagram of duty hours and run-cuts:* (*a*) *a typical diagram of duty hours;* (*b*) *run assignments for duty hours from* (*a*). (Source: *Lehner, 1978.*)

determination of specific TU blocks and driver runs, is very complex even for a single line. Scheduling for many lines is correspondingly more difficult. Moreover, the variations in schedules for different periods of day and among days of week, and considerations of various requirements and wage rates for labor, add to the computational complexity.

Efficient scheduling is, however, extremely important because it determines expenditure for what is usually the largest item of agency's operating expenditures: operating costs of vehicle-km operated and labor wages for hours worked. In large agencies, even small increases in schedule efficiency can translate into large annual funds. To facilitate this process, operations research algorithms with computer programs have been developed that can investigate hundreds of schedule and run-cutting variations. These may include line-pairing combinations, driver switches to other TUs after breaks in their work, variations in reserve TUs, and personnel shift changes at different locations and

Table 1.6 Summary of runs from the diagrams in Figure 1.35

Run	*Duration (min)*	*Gaps (min)*	*Work Length (min)*	*Breaks (min)*	*Total Work Spread (min)*
1	450	—	450	—	450
2	450	—	450	—	450
3	450	—	450	—	450
4	450	—	450	—	450
5	450	—	450	—	450
6	450	—	450	—	450
7	168 + 255 = 423	—	423	39	462
8	108 + 198 = 306	—	306	228	534
9	120 + 144 + 156 = 420	24	444	246	690
10	162 + 288 = 450	—	450	252	702
11	122 + 144 + 144 = 410	36	446	220	666
12	132 + 288 = 420	—	420	282	702
13	240 + 246 = 486	—	486	48	534
14	264 + 210 = 474	—	474	48	522
15	462	—	462	—	462
16	486	—	486	—	486
17	411	—	411	—	411
18	288 + 144 = 432	—	456	—	456
19	288 + 144 = 432	—	456	—	456
Totals	8312	108	8420	1363	9783

Characteristic elements:
1. Longest, shortest, and average work durations: 486, 411, and 443 minutes, respectively.
2. Longest, shortest, and average spread: 702, 411, and 515 minutes, respectively.

times of day, depending on the probability of demand variations, personnel absenteeism, etc.

Computer programs can produce schedules for several lines in a fraction of the time that a scheduler would take through manual computations. Even greater benefit is the ability to test variations in scheduling. When an initial schedule is produced, its total cost and various coefficients of utilization are computed; changes in the initial inputs, work assignments to individuals, wage rates, etc., can be assumed and immediately tested by the computer program. This testing of variations, shown by a dashed line as feedback in Figure 1.23, allows considerable improvements from an initial computation toward an optimal schedule.

Despite their advantages, however, computer applications for this purpose are not always straightforward because of numerous local details and procedural peculiarities that require the personal attention of a person familiar with the system. For example, travel times on some line sections may vary between peak and off-peak hours; acceptability of marginal values of load factor α must be subjectively decided; and labor contracts sometimes vary among divisions. All these details must be handled by experienced schedulers. For this reason, the introduction of computerized scheduling by transit agencies was historically a slow, often gradual process. For example, initial attempts to use computerized scheduling for buses of London Transport during the 1970s proved impractical because each one of its over 70 bus garages had a specific procedure of vehicle scheduling and different labor agreements. In many agencies, it took considerable efforts to develop computerized programs that could produce as efficient schedules as the experienced schedulers could prepare by traditional manual method.

With major efforts in the development of sophisticated and yet practical computerized scheduling programs, their applications have now become widely adopted and the size of agency in terms of TUs and operating personnel at which computer-based programs become more efficient than manual computations has decreased from hundreds to as few as 20–30 TUs, the exact number depending on local conditions and practices.

1.4.7.1 Structure and Contents of Packages. The best-known program for transit scheduling in the United States has been the Run Cutting and Scheduling (RUCUS) package. Developed under the sponsorship of UMTA/FTA, RUCUS was designed as a modular parametric package suitable for staged construction of schedules in any agency. Other programs were developed by several consulting firms, as well as by major transit agencies such as HHA in Hamburg, RATP in Paris, and SNCB in Brussels. Thus, there are several highly refined, comprehensive, but expensive packages sold internationally. Most packages consist of different modules: timetables, vehicle blocks, crew blocks, and various analytical tools used for revising inputs, as shown in Figure 1.23. In addition, there are numerous less expensive packages that perform fewer tasks but may be sufficient for small and medium-sized transit agencies.

The RUCUS package consists of three distinct modules: trips, blocks, and runs, corresponding to the three scheduling work elements in Figure 1.23. In addition, the program performs filing and preparing of passenger counts and travel time data for input into scheduling, and preparing and printing of results of scheduling—summary data for various reports. European transit operators utilize the convenient graphical schedules much more extensively than the U.S. transit agencies. Many European agencies also use different types of run-cuts than North America due to different labor practices. They use rotating duty rosters which give *equal time assignment cycle to each worker.*

Among these components, run-cutting is most frequently computerized, followed by determination of vehicle schedules (blocks) and timetables (trips). Summary data used for management reporting and regulation compliance are also frequently handled by computerized program packages.

Even with full computerization of the scheduling process, the work commonly remains modular because of the need for schedulers to intervene at different stages of the work. Actually, one of the reasons for the initial difficulties in introducing computers was that the first programs tended to be of the ‘‘black box’’ type, excluding the schedulers from the computations. The current trend is toward use of interactive programs that keep the scheduler informed and, more importantly, in charge of the computational process. The entire operation with computerized scheduling and information system requires considerable computer expertise, but that is not sufficient. A person thoroughly familiar with transit operations and experience in scheduling vehicles and personnel is always required to supervise the entire process.

1.4.7.2 Evaluation of Computer Applications. There are a number of benefits to as well as some costs of using computers for scheduling; there are also certain conditions that influence the benefits and costs. These should be evaluated to determine feasibility of computerizing that process for each transit agency.

Briefly stated, the benefits of computer-assisted scheduling are:

- Reduced manpower requirements for scheduling
- Increased probability of finding the optimal solution, i.e., minimal vehicle- and driver-hours for a given service level
- Ability to test quickly many alternatives and variations, including interlining, driver fallback after a break, and others, which is very useful for such actions as modifying schedules when passenger volume changes, for labor negotiations, cost considerations of different level-of-service policies, etc.
- Increased accuracy of computations
- Much greater ability to automate production of data and results of operations, such as timetables and payroll data
- Information produced that can be used for computer aided dispatching (CAD), automatic passenger counting (APC), real-time schedule adherence information, travel time data collection, scheduling of vehicle maintenance, and other computer-based operational functions

The costs or negative aspects of computer-assisted scheduling are:

- Implementation requires a major effort to upgrade and reorient the entire scheduling department information technology infrastructure and retrain its personnel.
- Agency must obtain the appropriate hardware and software.
- Agency must acquire competent programmers who understand transit operations.

Consequently, the benefits can be highly significant; the costs and effort are also major, but most of them are one-time investments. A few other factors should be borne in mind in evaluating the possible use of computers. First, for small agencies the use of computers is not appropriate because the benefits are so small (if any) that they cannot outweigh the efforts in preparing and adjusting the programs; benefits increase with the size of the agency. Hoffstadt (1981) estimated that a rationalizing effect is to be expected only when at least 30 vehicles are operated. Roy Lave in California also found 30 to be the threshold size for demand-responsive transit. And second, even the best computer programs cannot fully replace experienced schedulers. They replace the bulk of routine computations and person-hours of work, but good understanding of the process and experience in transit operations remain fundamental for efficient scheduling and optimal results.

1.4.8 Measures of Operating Efficiency

Strictly speaking, directly productive time of vehicles and personnel is only that when TUs actually travel and perform service on transit lines. Other times, such as terminal and deadhead, or sick leave for personnel, although necessary, should be minimized within the limits of various operating requirements and work standards. A very fundamental, systematic definition of concepts and analysis of schedules, run cuts, and their efficiencies was given by Lehner (1978). Based on his work, the following efficiency and utilization coefficients are defined.

1.4.8.1 Schedule Efficiency. The basic measure of efficiency of a transit line schedule is coefficient η_t, reflecting terminal time losses. It is defined as the ratio of the sum of operating times (in two directions) to cycle time:

$$\eta_t = \frac{T_o' + T_o''}{T} \qquad \left|\begin{array}{c|c} \eta & T \\ \hline - & \text{min} \end{array}\right| \tag{1.54}$$

Utilizing Eqs. (1.20) and (1.21), and assuming that $T_o' = T_o''$, η_t can be related to terminal time coefficients γ and γ':

$$\begin{aligned} \eta_t &= \frac{T_o' + T_o''}{T_o' + T_o'' + t_t' + t_t''} = \frac{2 \cdot T_o}{T} \\ &= \frac{1}{1 + \gamma} = 1 - \gamma' \qquad \left|\begin{array}{c|c} \eta, \gamma & T, t \\ \hline - & \text{min} \end{array}\right| \end{aligned} \tag{1.55}$$

If T and T_o are extracted from Eqs. (1.28) and (1.29) and introduced into Eq. (1.54), η_t can also be expressed as the ratio of cycle speed to operating speed:

$$\eta_t = \frac{V_c}{V_o} \qquad \left|\begin{array}{c|c} \eta_t & V \\ \hline - & \text{km/h} \end{array}\right| \tag{1.56}$$

As an example, if cycle time on a line is $T = 80$ min and operating times are $T'_o = 33$ min. and $T''_o = 35$ min, then terminal time coefficients are, from Eqs. (1.20) and (1.21), respectively:

$$\gamma = \frac{12}{68} = 0.18 \quad \text{and} \quad \gamma' = \frac{12}{80} = 0.15$$

while coefficient of schedule efficiency η_t is from Eq. (1.54):

$$\eta_t = \frac{68}{80} = 0.85$$

If it is assumed that $T_o' = T''_o$, which is true in most cases, and T is replaced by $N_{\text{TU}} \cdot h$ based on Eq. (1.48), then η_t can be expressed as:

$$\eta_t = \frac{2T_o}{N_{TU} \cdot h} = 1 - \frac{2t_t}{N_{TU} \cdot h}$$

$$\left| \frac{\eta}{-} \middle| \frac{T,\, t}{\text{min}} \middle| \frac{N}{\text{TU}} \middle| \frac{h}{\text{min/TU}} \right| \tag{1.57}$$

Since T_o is a function of operating speed V_o and line length L, this equation shows that η_t depends directly or indirectly on several operating elements. The relationships are not always simple, however, because the cycle time can only be a discrete number, an integer multiple of h. Thus, most relationships have a sawtooth shape, as shown in Figure 1.36, with sudden increases of η_t when T is shortened by h (min), so that N can be decreased by 1 for a given h.

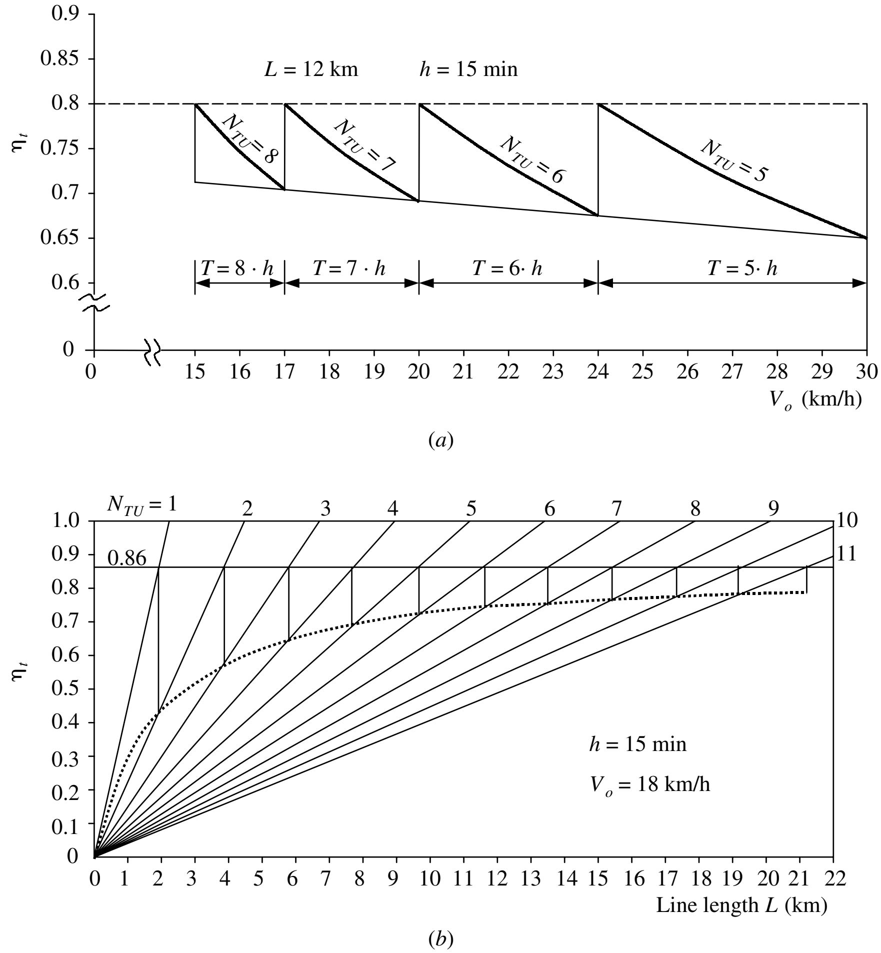


Figure 1.36 *Schedule efficiency coefficient* η_t *as a function of operating speed* V_o *and line length* L*: (a)* η_t *as a function of* V_o *for given* h *and* L*; (b)* η_t *as a function of* L *for given* h *and* V_o. (Source: *Lehner, 1978.*)

Table 1.7 Influence of operating speed on the schedule efficiency coefficient η_t

Operating Elements	Dimension	Case i	Case ii	Case iii	Case iv
V_o	km/h	18.0	19.5	20.0	21.5
T_o	min	40	37	36	34
T	min	80 + 25 = 105	74 + 31 = 105	72 + 18 = 90	68 + 22 = 90
η_t	—	0.76	0.70	0.80	0.76
V_c	km/h	13.7	13.7	16.0	16.0
N_{TU}	—	7	7	6	6

Line parameters: L = 12 km, h = 15 min, Min $(t_1' + t_t'')$ = 18 min.

The diagram in Figure 1.36(*a*) shows η_t as a function of V_o for fixed values of L and h. As V_o is increased, η_t decreases as long as T remains constant. When V_o is so high that T can be reduced by h [min], η_t jumps to the highest value possible for the minimum length of t_t, as determined by the crew rest and other operating requirements.

The diagram in Figure 1.36(*a*) shows that η_t generally tends to decrease as V_o increases. Lehner has shown that because of the discontinuous nature of the cycle time, impacts of other operating elements on η_t are also discontinuous; however, for a given line, η_t has a general tendency to increase with line length L (Figure 1.36(*b*)), but to decrease as headway increases.

The relationship from Figure 1.36(*a*) is also illustrated in Table 1.7 by schedule computations for four cases. As V_o increases from case i to ii, η_t decreases; then, from ii to iii T can be shortened, so that η_t increases and N decreases from 7 to 6. In case iv T remains the same as in iii, so that η_t decreases again.

1.4.8.2 Operating Personnel Efficiency. An important measure of labor force utilization is the ratio of net productive working hours on transit line to the total number of paid hours, designated as operating personnel efficiency coefficient η. That coefficient consists of the following elements.

- η_a, attendance coefficient, represents the ratio of reported hours (time) t^r, to paid hours t^p, accounting for losses due to vacation, illness, and other absences.
- η_s, coefficient of run-cutting, or the ratio of hours on transit line t^ℓ to reported hours t^r. It includes losses caused by split shifts, work preparation, deadheading, etc.
- η_t, coefficient of schedule efficiency, defined as Eqs. (1.54) and (1.55). Coefficient η is the product of the preceding three coefficients:

$$\eta = \eta_a \cdot \eta_s \cdot \eta_t = \frac{t^r}{t^p} \cdot \frac{t^l}{t^r} \cdot \frac{2 \cdot T_o}{T} = \frac{t^l}{t^p(1 + \gamma)} \qquad \left| \frac{\eta}{-} \middle| \frac{t, T}{\text{min}} \middle| \frac{\gamma}{-} \right| \tag{1.58}$$

It is difficult to give typical values for these coefficients because they depend on many local factors. For example, rapid transit systems usually have higher values of η_s than buses and regional rail systems; labor contracts or work laws affect η_a, while fully auto-
mated systems have a different structure of their personnel (no drivers, but more supervisors). Although their absolute values vary, these coefficients can be very useful for comparisons of alternative vehicle and personnel schedules (run cuts) for an agency, or for comparison of operating efficiencies of similar transit services.

EXERCISES

1.1. A bus line has stops at every intersection, 200 m apart. The speed of traffic and of the buses except for their stopping at bus stops is 30 km/h, but for every stopping, buses lose a total of $t_s = 20$ s. Passengers are uniformly distributed along the line and walk to/from its stops at a speed of $V_w = 4.8$ km/h. How would a passenger's total travel time, including access to/from stops and travel on buses, change if every other stop was abandoned, i.e., the spacing between stops was increased from 200 to 400 m, if the average trip length on the line is: a) 4 km; b) 6 km; c) 8 km? For what passenger trip length L_p will the total travel time remain the same?

1.2. A heavily traveled bus line has a length of 12 km. It has stops at every intersection and there are 5 blocks per kilometer. The speed of traffic and of the buses except for their stopping is 34 km/h. Terminal time is 9 min at each end. Assume that the buses stop at every stop, losing 20 s for deceleration and acceleration plus standing (dwell) time of 15 s. The average passenger trip length on the line is 4.8 km. Service headway is 5 min.

You want to propose a reduction of stop density to one per two blocks. In that case, time loss for each deceleration/acceleration would not change, but standing times would increase to 25 s due to increased number of passengers boarding and alighting per stop. The company management believes that the existing operation is desirable for passengers (short access) and that the changes would not make much difference in operating costs anyway. To convince the management that the change would be useful, compute and systematically present the following consequences of the change in stop density, assuming uniform distribution of origins-destinations along the whole line:

a. Additional average walking distance per passenger along the line, which includes access and egress (in m/prs).

b. Additional average access and egress time (in min/prs), if the speed of walking is 75 m/min.

c. Reduction of the average passenger travel time on buses (in min/prs).

d. Change in the passenger total travel time.

e. Break-even passenger trip length L_a between the two stopping regimes. HINT: Compute the travel distance on the line on which passenger saves as much time as he/she loses for longer walking to/from stops.

f. How many buses can be saved due to higher operating speed. HINT: Compute operating time $T_r = 60L/V_r$ rounded in min; then add the present additional time for stopping (converted in minutes) to obtain T and then N. Do the same for the new stopping times and find the difference ΔN.

1.3. Operating (travel) speed on an 8-km-long bus line with $h = 12$ min is $V_o = 16$ km/h. Terminal time at each end is 6 min ($\Sigma t_t = 12$ min). Bus preferential treatments are considered to improve the operation, but they would require a certain level of investment. The company will make the investment if it can reduce the number of buses by one due to these improvements. What is the minimum value of V_o that would allow saving one bus if the total terminal time at both ends should not be less than $T_t = 0.15\ T$, and the headway remains 12 min?

1.4. A bus line with a length $L = 2430$ m has 6 stations, including terminals. Interstation distances have the following lengths: 520, 280, 680, 450, 500 m. Running speed on the line is $V_r = 32$ km/h, headway is 4 min, and terminal times at each end are 5 min. Draw a general form of a graphical schedule for two buses operating on this line at headway h: plot a diagram with 1500 s on the abscissa and 2500 m on the ordinate. Show on the diagram straight lines of bus travel between stops and time lost per stopping of 30 s. Show also the following elements: h, T_o, t_t, T, V_o and V_c, assuming T_o and t_t are the same in each direction.

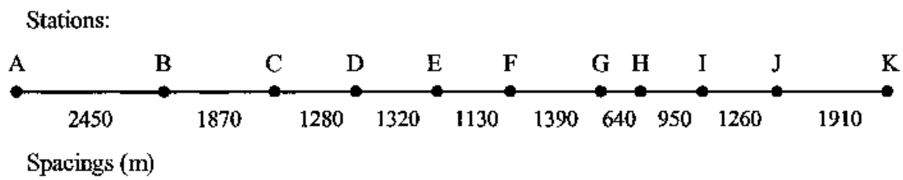
1.5. A 14-km (one-way) long trolleybus line has an operating speed $V_o = 12$ km/h; terminal time at each end t_t is at least 6 min.

a. How many trolleybuses are required for operation with a 10-min headway? What will be the cycle speed V_c for the line with that schedule?

b. What new operating speed V_o' should be achieved in order to reduce the number of trolleybuses by two, while maintaining the same headway and increasing the minimum required terminal time at one terminal to 8 min? Compute also the new cycle speed V_c'.

c. What will be the offered line capacity if $C_v = 80$ sps? Compute α_{max} if the line carries 300 prs/h on its maximum load section.

1.6. A rapid transit line is 14.2 km long (one way) and has the following passenger volumes boarding and alighting at individual stations in the peak direction during the peak hour:



Station	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*
Boardings (prs/h)	3300	1700	1900	3200	2900	1300	1600	600	400	700	0
Alightings (prs/h)	0	0	500	700	2100	700	4000	2700	2200	1700	3000

Capacity of each train is 590 spaces, α_{max} is selected to be 0.9, travel speed on the line is $V_o = 36$ km/h. The labor union contract requires that terminal times at both terminals together should be at least 15% of the two-way operating time, but no less than 10 min.

a. Compute the number of trains required for the service, excluding reserves. Using that number for N_{TU}, round the cycle time up to the next multiple of the headway, then compute the value of γ.

b. Assuming that there is no lower limit on terminal time, how much should the terminal time be reduced to permit withdrawal of one train, without changing the headway? What will be the new value of γ?

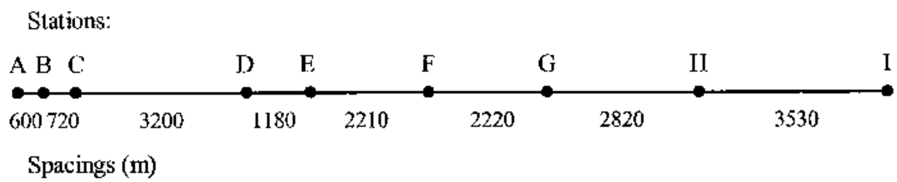
c. If terminal times remain as found under part a, how much should the travel speed be increased to reduce the number of trains on the line by one without changing the headway?

d. Draw a diagram of this line with distance and stations on the abscissa, passenger volume, and offered capacity on the ordinate (see Figure 1.5). Show the MLS.

e. Compute the work in person-km/h on this line and its work utilization coefficient α.

1.7. A 16.48-km-long regional rail line from center city to a suburb has the following characteristics during the afternoon peak hour:

The line has the following stations, station spacings, and boardings and alightings:



Station	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Boardings (prs/h)	2640	1968	3156	288	160	20	48	12	0
Alightings (prs/h)	0	96	144	1032	1240	3020	1032	528	1200

Train consists are 5 cars, and each car has 94 seats and a total capacity of 145 spaces. Operating speed on the line is 42 km/h. Terminal times cannot be shorter than 5 min at each end.

a. Compute and draw the load diagram (profile of passengers, offered seated, and total capacities). Compute and show the MLS.

b. Compute schedule I for the line adopting the maximum load factor $\alpha = 0.65$ on the MLS.

c. Compute two other schedules for a more economical operation when alternative trains would turn back at station F and at terminal I: schedule II would again have load factor $\alpha = 0.65$, while schedule III should be based on a higher load factor $\alpha = 0.87$.

d. Compare the three schedules by:

- The number of trains and cars required
- Work utilization coefficients

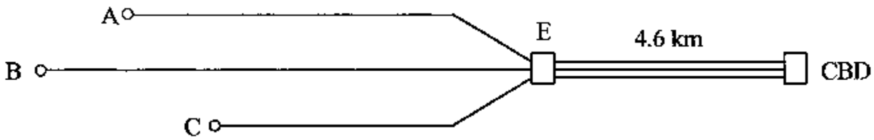
e. Draw a graphical schedule showing operation III. Note that the headway on the joint section must be uniform in both directions.

f. List other factors that should be considered in selecting between schedule I and schedule II or III with short-turn trains.

1.8. Five bus lines merge to a joint section on a trunk street. If their headways are 2, 5, 7.5, 15, and 20 min, what will be the average headway h_{av} on the common section, including buses from all five lines?

The MLS, which is on the common section, has a volume of 2480 prs/h. The capacity of each bus is 68 spaces. What is the average load factor α?

1.9. In a situation similar to Boston, Philadelphia, and San Francisco, three radial LRT lines-merge into one 4.6-km-long trunk line leading to the city center. Operating speed on the trunk is 19.6 km/h.



The branches have the following characteristics and passenger volumes during the peak hour:

Line	*Length** (*km*)	*Operating Speed on Branches* (*km/h*)	*Passenger Volume* (*prs/h/direction*)**
A	8.1	18.6	3520
B	10.1	24.0	2640
C	6.0	18.2	705

*From outer terminals to point E, one way.
**East of point E, which is the maximum load section for all lines.

Each LRT vehicle has a capacity of 156 spaces, including 58 seats; average maximum occupancy of each vehicle should be α_{max} = 66% (all seats occupied plus 45 persons standing).

a. Compute the headway for each line using single cars. If single-car operation would result in a headway shorter than 6 min, use 2-car TUs. The policy headway for all the lines should be 7.5 min.

b. Find average headway of service on the trunk section (including vehicles from all lines).

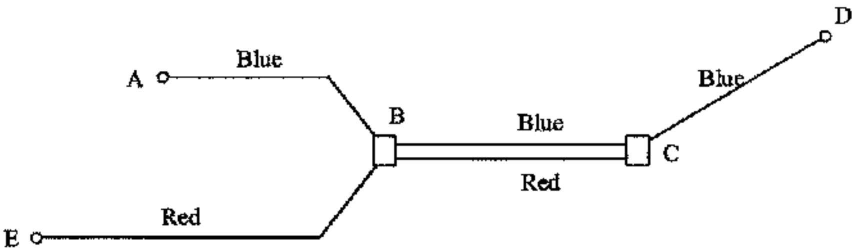
c. Compute the total fleet required for this service, assuming 10% reserves.

d. Suppose that operating speed on the trunk is reduced to 15.1 km/h. What is the required increase in fleet size in order to maintain the same headway on all three lines?

e. If the operating speed on the trunk remains 19.6 km/h, what operating speed should be achieved on the outer section of line B to reduce the number of TUs used by one?

Note: Round off headway on individual lines to the closest half minute and make them divisible in 60. Average headway should be exact, in seconds. Total terminal times for each line should be at least 0.15 T.

1.10. Two LRT lines, shown below as the Blue Line and the Red Line, should be scheduled. Data about their individual sections are given in the table. Service on the joint section, BC, should be with uniform headway. The lines are served by two-car trains, and each car has a capacity of 160 spaces; $\alpha_{max} = 0.80$.



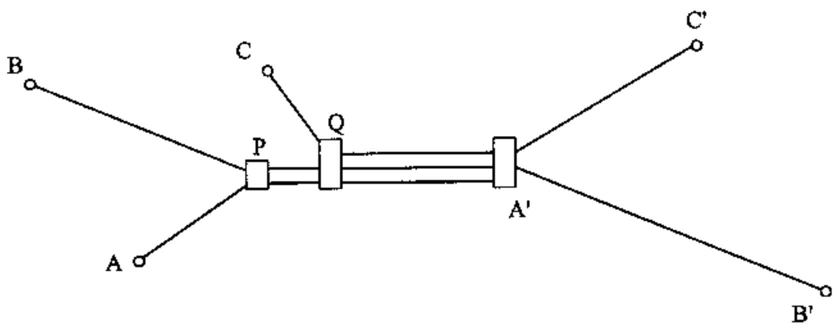
Line Section	*Length (km)*	V_o *(km/h)*	P_{max} *(prs/h)*
AB	3.6	18	1210
BC	4.2	15	2500
CD	2.8	16	1150
EB	6.4	18	1490

a. Begin by computing the headway for each line separately, assuming that the passengers on the joint BC section are distributed equally between the two lines. Then adjust these to meet the uniform headway requirement.

b. Determine the required fleet size, assuming that terminal times at both terminals together amount to at least 10% of the cycle time, and that 12% of all vehicles are in repair of reserve (i.e., the cars on the lines should represent 88% of the fleet).

c. Draw a graphical schedule with the derived values for a period including at least one cycle time of each line. The ordinate of the diagram should start from terminal D, then have CBA at given distances. From point A leave a 1-cm gap and then plot section BE, similar to Figure 1.31.

1.11 A network of three bus lines, AA′, BB′, and CC′, is shown here:



Each section of lines has the following characteristics:

Section	*Length* (*km*)	V_o (*km/h*)
AP	3	22
BP	6	16
PQ	1.5	13
CQ	1.5	14
QA′	5	13
A′B′	8	16
A′C′	6.5	22

Terminal time (at both ends together) is approximately $t_t = 0.15T$. Passenger volumes on the maximum load section, in persons/hour/direction, are: A-A′ = 980, B-B′ = 1515, and C-C′ = 500.

Two types of buses are available for service:

1. Standard buses with capacity of 53 seats for which the load factor is $\alpha = 1.0$
2. Articulated buses with 80 seats, with $\alpha = 0.9$

The maximum frequency of buses at any point without major delays is 60 veh/h.

a. Compute h for each line for each type of bus.

b. Determine which type of bus would be better for each line on the following basis. For headways longer than 6 min, select the type that provides the shorter headway; for headways shorter than 6 min, passenger waiting time is not a major factor, so that the bus type having the lower cost per seat-km (articulated) should be adopted.

c. Compute T, N, T', and V_c for each line for the selected vehicle type (1 or 2).

d. Compute the average headway on sections PQ and QA′.

e. Suppose that a number of improvements (such as reserved lanes, signal actuation by buses, presale of tickets, etc.) are introduced on section PA′, so that the operating speed is raised from 13 to 18 km/h. Compute how many buses each line would need.

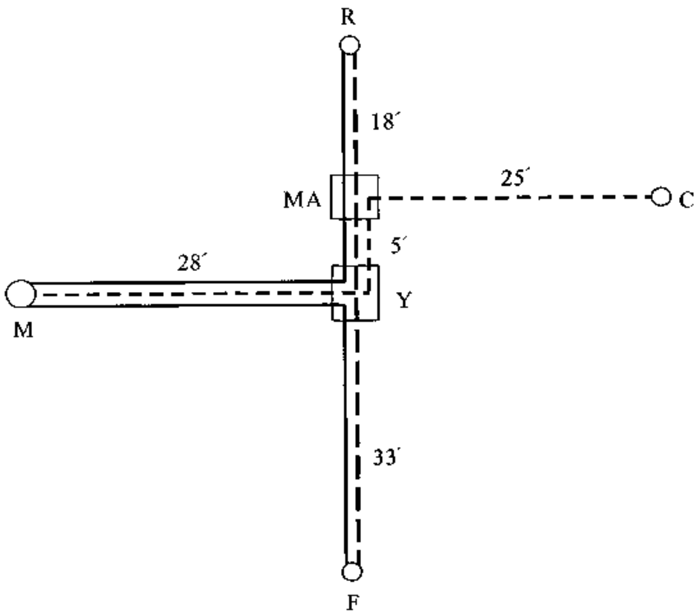
1.12 A network shown below consisting of a trunk line MY with three branches, R, C, and F, and a cross line between two branches RF (similar to the initial San Francisco BART network) has to be scheduled. One-way travel times for each line section are given on the diagram. Compute the schedules for all the lines based on the following headways:

Line MR: 15 min
Line MC: 7.5 min
Line MF: 15 min
Line RF: 15 min

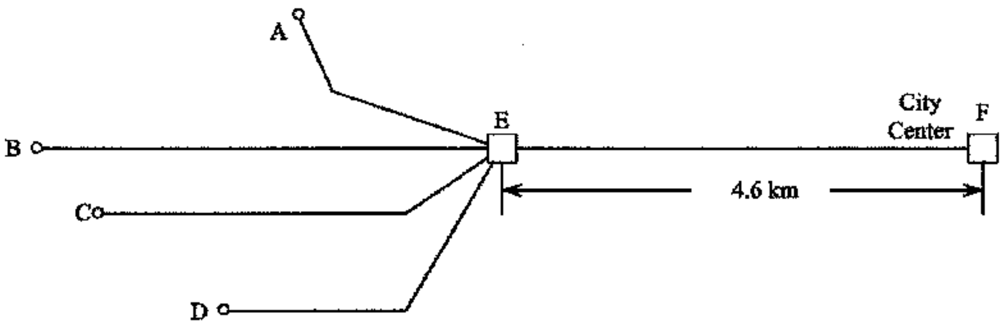
making sure that the headway on each line and on the joint section from Y to M are uniform.

Draw the computed schedules on a diagram using the patterns from Figures 1.33 and 1.34. Show all the trains leaving terminals from 12:00 noon, cutting off the diagram at 15:00. Note that the distance scales vary among the diagrams, so that the slopes of the scheduled lines will be different (but mutually parallel) on each section.

Compute the number of trains required for the service, assuming that terminal times at each terminal must be at least 5 minutes.



1.13. Suppose that, similar to the network of Green Lines in Philadelphia, four radial streetcar lines, AF, BF, CF, and DF, merge into a 4.6-km-long trunk line leading to the city center:



The lines have the following characteristics:

Line	*Branch Length from Outer Terminal to E (km)*	*Maximum Load on Each Branch (prs / h / direction)*
AF	4.8	1920
BF	7.6	1450
CF	8.2	2550
DF	5.4	1200

Each streetcar has a capacity of 125 spaces. Maximum occupancy of vehicles should be 80%.

a. Find headways for each line using only values divisible by 60.

b. Find the average headway on the trunk section (including vehicles from all four lines);

c. If the cycle speed on all lines is 25 km/h, compute cycle times T and find the minimum number of vehicles required for the derived headways on each line.

REFERENCES

Banković, R. 1982. *Javni gradski putnički prevoz* (Urban Public Transportation). Belgrade: Gradjevinska Knjiga.

Bowman, L. A. and A. M. Turnquist. 1981. Service frequency, schedule reliability and passenger wait times at transit stops. *Transportation Research* 15A(6):465–471.

Canadian Urban Transit Association (CUTA). 1993. *The Canadian Transit Handbook.* Toronto: CUTA.

Ceder, A. 1987. Methods for creating bus timetables. *Transportation Research* 21A(1): 59–83.

Ceder, A., and H. Stern. 1981. Deficit function bus scheduling with deadheading trip insertions for fleet size reduction. *Transportation Science* 15(4):339–363.

Fruin, J. J. 1971. *Pedestrian Planning and Design*. New York: Metropolitan Association for Urban Designers and Planners, Inc.

Furth, P. G. 1980. *Setting Frequencies on Bus Routes.* Research Report CTS-RAMP-80-2. Cambridge, MA: Massachusetts Institute of Technology.

Groche, G., and E. Thiemer. 1980. *Handbuch für den Öffentlichen Personen-Nahverkehr-ÖPNV* (Urban Transit Handbook). Darmstadt: Otto Elsner Verlag.

Hoffstadt, J. 1981. Computerized vehicle and driver scheduling for the Hamburger Hochbahn AG. In *Computer Scheduling of Public Transport,* ed. A. Wren, 35–52. New York: Elsevier.

Homburger, W. S. 1982. Urban transit. In *Transportation and Traffic Engineering Handbook,* 2nd ed., ed. W. S. Homburger, Chap. 7. Englewood Cliffs, NJ: ITE/Prentice Hall.

Hulbert, S. 1982. Human factors in transportation. In *Transportation and Traffic Engineering Handbook,* 2nd ed., ed. W. S. Homburger, Chap. 8. Englewood Cliffs, NJ: ITE/Prentice Hall.

Kittelson & Associates. 2003. *Transit Capacity and Quality of Service Manual*. TCRP 100. Washington, DC: Transportation Research Board.

Krstanoski, N. 2000. *Zbirka Rešeni Zadaci po Javen Gradski Patnički Prevoz* (Collection of Solved Problems in Urban Transit). Technical University Bitola, Macedonia.

Lehner, F. 1950. Menge, Arbeit, Leistung and Wirkungsgrad im Verkehr. *Verkehr und Technik* 1 and 2.

———. 1957. *Die Verkehrsbedieunung im Personennahverkehr.* In *Der öffentliche Personennahverkehr,* Risch und Lademann, Chap. 3, 44–143. Berlin: Springer.

———. 1978. (originally published in 1953), *Der maximale Wirkungsgrad des Personaleinsatzes*. Düsseldorf: Alba Buchverlag.

Levinson, H. S. 1992. Urban mass transit systems. In *Transportation Planning Handbook,* ed. J. O. Edwards, Chap. 5. Englewood Cliffs, NJ: Prentice-Hall.

Molinero, A. R. M., and L. I. S. Arellano 1997. *Transporte Publico: Planeacion, Diseno, Operation y Administracion*. Universidad Autonoma del Estado de Mexico.

Rainville, W. S. 1982. (originally published in 1947), *Bus Scheduling Manual: Traffic Checking and Schedule Preparation*. Washington, DC: U.S. DOT and American Public Transport Association.

Robertson, D. ed. 1994. *Manual of Transportation Engineering Studies,* 1st ed. Washington, DC: ITE.

Rüger, S. 1974. *Betriebstechnologie städtischer öffentlicher Personenverkehr.* Berlin: Transpress VEB Verlag für Verkehrswesen.

Vuchic, V. R. 1981. *Urban Public Transportation Systems and Technology*. Englewood Cliffs, NJ: Prentice-Hall.

Vuchic, V. R., et al. 1978. *Transit Operating Manual,* 2nd ed. Harrisburg: Pennsylvania DOT.