HIGHWAY CAPACITY AND TRAFFIC CONGESTION:
A PRELIMINARY STUDY 1/2/

C. B. McGuire

July 23, 1952

1. Introduction

In the earliest activity-analysis models of transportation, \[9\] no account was taken of the phenomenon of congestion. The reality of these models was probably not seriously affected in the case of ocean shipping, but it was clear that interference between vehicles had to be considered if highway and rail transportation were to be accurately analysed. Recently a suggestion of Koopmans had led to the development of some models of transportation on highway networks which assign to each segment of the network a limited capacity. The purpose of this paper is to look at this notion of capacity a little closer, and in part to examine the uses to which it has been put in these recent transportation models.

We shall first discuss in Section II a pure capacity concept which for many years has been a major concern of traffic engineers, and which has been

---


2/ I am indebted to Martin Beckmann and Tjalling C. Koopmans for many valuable suggestions. The extent to which the present paper leans on the work of C. K. Normann and his associates is obvious.
used by Beckmann [1] in a model of centrally directed traffic movements. So that this not always realistic element of central direction may be removed, we shall digress in Section III to discuss the preferences of individual drivers. In Sections IV and V we shall describe and enlarge upon the more sophisticated traffic analysis resulting from the empirical studies of Norman and the Bureau of Public Roads. This latter approach is embodied by Beckmann and McGuire [2] in a model which is based on individual choice (by drivers) of routes and speeds. In Section IV the concept of capacity will be somewhat extended to include a description of the relationships between flows and speeds, and in Section V an attempt will be made to define a measure which indicates the average or expected nuisance value of congestion. The last section will devote a few remarks to applications.

It will be seen that much of our discussion consists of a review of the procedures used by traffic engineers in dealing with highway capacities and problems of congestion. Our primary interest of course is not whether these students of the subject are right or wrong; their interests are somewhat different from ours, and in any case we are in no position to judge their work. However, the alternative to the course we are taking would be a statistical analysis, probably of large scale; the present course seems to be a logical first step. Expert opinion and cited facts for typical cases form the material we shall use.

Throughout the discussion it will be assumed that particular roads may be considered for present purposes without reference to their position in a network of roads. Strictly speaking, our "road" (or "highway") is of infinite length and has no points of access from side roads; in fact, the conclusions obtained are applicable to any fairly long unintercepted road section. A configuration of such "roads" we shall call a "network". City streets are thus
excluded from the study, since their length is probably not great enough to justify disregarding the influence of intersections. Urban expressways and rural highways are more the type of road we have in mind. Much more serious is our neglect of the problem of traffic movements at intersections. A satisfactory solution to this more difficult problem probably calls for a study of the meshing of random streams of traffic. Some recent theoretical advances in rather removed fields appear to be applicable to this and some related problems of traffic movement as well, but we shall postpone discussion of these for the present.

II. Uniform Speeds and Capacities

Consider a single-lane road over which traffic moves in one direction. A uniform speed will be said to exist when every vehicle moves at the same speed $u$. The uniform capacity $c(u)$ at a given point on the road is defined as the maximum traffic flow (measured in vehicles/hour) which can pass that point at uniform speed $u$ under certain assumed conditions. Obviously flow will depend on the distance-spacing between vehicles, and will be at a maximum for any given speed when spacing is at a minimum. The uniform capacity function can be written

$$c(u) = \frac{u}{d(u)} \quad (1)$$

where $d(u)$ is the spacing function, which tells us the average center-to-center trailing distance which drivers maintain between their own vehicle and the one just ahead of them.

---

$^{3/}$ [8], [16].
Four ways of determining $c(u)$ suggest themselves:

(a) Hypothetical spacing functions can be constructed and substituted in (1),

(b) Empirical spacing functions can be substituted in (1),

(c) An empirical uniform capacity function can be determined by direct observation of existing traffic flows,

(d) An experiment can be performed.

To our knowledge (d) has never been carried out. Each of the others will be described below.

(a) A minimum "safe" spacing function $d(u)$ may be postulated, and an attempt made to estimate its parameters. This has been the most common approach.\footnote{Cf. [5], [6], and bibliography in [13, p. 122].} Without much justification engineers (and also physicists trying their hand at the problem) have usually assumed that safety requires a distance between vehicles at least equal to that necessary to come to a full stop. Thus typically,

$$d(u) = a + bu + cu^2$$

where $a =$ average length of vehicle

$b =$ perception time plus reaction time

$c =$ $1/2r$ where $r$ is the maximum rate of deceleration.

All three values are affected by the composition of traffic (number of trucks, age of drivers, etc.). Time of day and location affect $b$; and road surface, weather, and grade and curvature of the road affect $c$. The value of $a$ is easily determined. Some of the most careful studies of $b$ have been made by De Sylva [3], and of $c$ by Moyer [10]. Various values used in constructing $d(u)$ functions are interestingly compared by Normann [13, p. 120]. While much disagreement is evident the values
\[ a = 2.84 \times 10^{-3} \text{ miles} \quad (15 \text{ feet}) \]
\[ b = 2.78 \times 10^{-4} \text{ hours} \quad (1 \text{ second}) \]
\[ c = 9.47 \times 10^{-6} \text{ hours}^2/\text{mile} \quad (\text{i.e., deceleration} = 21.6 \text{ ft/sec}^2) \]

are fairly typical and give an idea of the magnitudes involved. It will be seen that the "safe" trailing distance grows more and more rapidly as speed increases. The resulting uniform capacity function is at a maximum when \[ u = \sqrt{\frac{a}{c}} \equiv \hat{u}, \] which means that as speed increases beyond \( u \), the spacing necessary for safety increases so fast that actual flow declines. This simple proposition, long a matter of dispute, is now almost universally accepted. The general form of \( c(u) \) is shown in Figure 1 (Curve a).

![Figure 1. Uniform Capacity Curves](image)

\[ 5/ \quad [15, \text{ p. 202}] \text{ and } [13, \text{ p. 120, 122}]. \]
In most of these studies has been found to be surprisingly low, in the 11–25 mile/hour range, and the corresponding single-line capacity around 2000 vehicles/hour.\[^6\]

A criticism which might be levied against this kind of uniform-capacity function is the concept of safety it uses. Minimum safe spacing is a difficult quantity to define to start with, and furthermore, there is ample reason to believe that drivers do not space themselves at distances which are safe anyway.\[^7\] While some attempts have been made to refine this concept of safety in order to secure a better correspondence between theory and observation, they are more properly classified under (c).\[^8\]

(c) The c(u) function can be estimated directly by fitting a curve to a set of maximum observed flows for several different uniform speeds.\[^9\]

The main difficulty encountered here is that very seldom are the higher uniform speeds observed. When speeds are high some passing nearly always takes place. This difficulty points up one of the rather artificial features of a uniform-capacity function. No roads of importance are one-lane roads, yet the function is defined only for the latter. The complications introduced by passing will be dealt with in the following sections.

\[^6\] [13, p. 120].

\[^7\] [11, p. 226].

\[^8\] Cf. the interesting dispute over "safety" in [7].

\[^9\] This procedure is hinted at in the literature; we are not at all certain that it has been carried out.
(b) An empirical $d(u)$ function can be found, and entered in (1) to yield $c(u)$.\(^{10}\) The procedure in a typical instance was as follows. Of 8,500 vehicles recorded at one location, about 2,000 were traveling at the same speed as the vehicle just ahead of them. These were classified into speed groups, and the time-spacing\(^{11}\) and distance-spacing distributions found for each. The distribution of time-spacings for the 30 mile/hour group is shown in Figure 2, where it can be seen that very few spacings are found to be greater than four seconds. Since this held for all speed groups it was decided to throw out observations for

![Figure 2. A Distribution of Time-Spacings [11, p. 226].](image)

\(^{10}\) Four references cited in [13, p. 120, 122]. The procedure described is Normann's [11, p. 226-227].

\(^{11}\) If $d(u)$ is the distance-spacing between a vehicle moving at speed $u$ and the preceding vehicle, then time-spacing is defined to be $\frac{d(u)}{u}$. 
which spacing was greater than four seconds on the basis that these vehicles just happened to be going at the same speed as the vehicle ahead but were not traveling at the minimum spacing. Figure 3 shows the average minimum spacing in feet of all vehicles spaced at less than four seconds, for each speed group. Using the data in Figure 3, a uniform capacity curve was constructed which indicates for each speed the

![Spacing Graph](image)

**Figure 3. Observed Minimum Spacings** [11, p. 227].

maximum flow possible if all traffic moved at just that speed. The resulting curve (b in Figure 1) differs from a typical theoretical curve (a in Figure 1) mainly in indicating greater flows at the higher speeds. Notice also that the speed, denoted $\hat{u}$, at which this empirical $c(u)$ is maximised is greater than before.

One of the reasons for the variety of methods used to determine uniform capacity functions is that congestion often appeared to be present when flows were far short of the maximums indicated by the particular $c(u)$ function being used. It was then thought that a more careful procedure
would lead to a function without this defect. But the trouble probably was that an average rather than a uniform speed was being inserted in $c(u)$. The difference between the two cannot be neglected, as we shall see.

Difficulties such as these make it impossible to use a uniform capacity function alone to explain flow limitations. There are, however, still some reasons for finding respectable uniform capacity functions for typical roads or points. Such information would be extremely useful to an authority (such as the army) which had complete control of a traffic movement, so that all speeds could be specified. These functions also represent ideals in the sense that they show the loss in terms of flow when drivers select a variety of speeds. In addition, they will be found useful below in the discussion of more sophisticated capacity concepts which take into account the fact that universal speeds are not ordinarily found in practice.

III. Desired Speeds

In an early formulation of the uniform capacity concept Simpson [18] claimed that one of the reasons these capacities were never attained was that the inconvenience caused drivers as flow approached $c(u)$ resulted in their choosing other routes. Until recently few such explicit mentions were made of the importance of drivers' inconvenience to an adequate definition of highway capacity. It is not enough to know that a road can carry a flow of 2,000 vehicles per hour; another road may be able to pass the same flow with a smaller degree of individual inconvenience to its users. Shouldn't the second road be said to have a higher capacity? Throughout this discussion the
distinction must be kept in mind between these two facets of capacity — drivers' inconvenience and traffic flow. It will usually be found that, for any given road, a greater flow can come about only through greater inconvenience to drivers.

Before discussing this matter any further, a word should be said about the assumptions we make concerning drivers' preferences, which enter an activity-analysis model of a highway network in two ways. Drivers going from one given point to another must select their routes, and they must select the speeds at which they want to travel on each particular road. Our concern here is not with networks but only with individual roads, so we shall not discuss route selection beyond saying that in the individual-choice model of [2] drivers are assumed to select that route which minimizes time of travel.³²/ As for speed selection, we shall assume that for any given road each driver chooses a speed at which he wishes to travel. This desired speed probably depends on the driver's estimate of the risks involved, which in turn depends on the characteristics of the road and his own vehicle. Perhaps operating costs as a function of speed also enter, but this is much less certain. For a given road then and a particular group of users we have a distribution of desired speeds or "free-speed distribution". The term "inconvenience" will refer to the inability of drivers to travel at desired speeds on congested roads.

We intend to approach the study of this inconvenience aspect of congestion from two quite different points of view. At first, it might be thought that a transportation model which by its very nature views a highway network as a whole should not have to take inconvenience into account explicitly: that route will be least inconvenient, for any given driver, over which travel time is least.

³²/ For empirical justification of this assumption, see [19].
However, it might very well be that congestion is a nuisance in still another sense. Drivers may prefer uncongested roads even when they involve no relative time advantage over congested roads. Therefore we shall consider separately the relationships between flows and the average of speeds of all vehicles (Section IV), and a measure of the average or expected nuisance of congestion (Section V).

IV. Flows and Average Speeds

The only precise attempt to examine highway capacity in terms of inconvenience and speed reductions has been made by Normann and his group. This approach will be explained and enlarged upon in what follows. The description will refer to a two-lane road which carries traffic in both directions. The composition, in terms of physical characteristics and drivers' preferences, of the two streams of traffic will be assumed known. That is, we know what portions of the flows are trucks, etc., and we have complete information about the two free-speed distributions. We shall assume that these influences are fixed.

The distribution of desired speeds will ordinarily not be the same as the distribution of actual speeds. For a fast car to pass a slow car some empty space in the opposing lane is necessary. If the density\(^{13}\) of vehicles in this opposing lane is high, then such spaces will be seldom available, and the passing maneuver will be delayed or prevented. As a result, the actual speed distribution will differ from the free-speed distribution.

Suppose that for traffic on one lane the free-speed distribution is represented by a distribution function \(f(u)\). The function \(f\) is to be interpreted as referring to the flow passing a particular point. That is, if \(x\)

\(^{13}\) The number of vehicles per mile of road at any instant.
is the flow, then $xf(u)$ is the part of the flow which travels at speed $u$. The density of vehicles traveling at speed $u$ is then $\frac{xf(u)}{u}$. Let $A$

denote a car traveling at speed $u'$. If $u < u'$, then $A$ must in one hour pass
\[
\frac{xf(u)}{u} (u' - u) \] vehicles traveling at speed $u$. Altogether, $A$ must pass
\[
x \int_0^{u'} \frac{f(u)}{u} (u' - u) \, du \] vehicles per hour, and $\frac{x}{u'} \int_0^{u'} \frac{f(u)}{u} (u' - u) \, du$
vehicles per mile. Thus the total number of passings per mile per hour for all
vehicles is
\[
x^2 \int_0^{\infty} \frac{f(u')}{u'} \int_0^{u'} \frac{f(u)}{u} (u' - u) \, du \, du'.
\]
This means that if the free-speed distribution were maintained, the number of
passing maneuvers per mile per hour would have to increase as the square of
flow. If opposing traffic flow does not permit this increase then the
faster vehicles will at times be prevented from passing the slower vehicles.
A new speed distribution will result, the mean of which will be lower than the
mean of the free-speed distribution. Investigations have revealed the way the
speed distribution changes with increases in flow for typical cases, and have
shown that the actual number of passings per mile per hour increases with flow
up to a certain point, and then decreases to zero as flow becomes so great that
all vehicles are forced to move at the same speed.$^{15}$

\[14\] Cf. [14, p. 69].

\[15\] [14, p. 70].
A linear regression of average north-bound speed, $\bar{u}_n$, on north- and south-bound flows, $x_n$ and $x_s$, was found by Normann to be

$$\bar{u}_n = 44.12 - 0.0104 x_n - 0.00729 x_s$$

(3) 16/

for one road examined. Just how representative of all such two-lane roads these coefficients are is not clear, since the results of other analyses have not been published in as great detail. However the use made of equation (3) suggests that we are safe in assuming the existence of functions $\bar{u}_n = \phi_n(x_n, x_s)$ and $\bar{u}_s = \phi_s(x_n, x_s)$, with negative slopes as in (3), for each road.

We shall call such functions capacity functions. It should be emphasized that $\phi_n$ does not state only that $\bar{u}_n \leq \phi_n(x_n, x_s)$ when $x_n$ and $x_s$ are the flows, but rather that equality holds. The fact that some drivers will wish to go faster (unless $\bar{u}_n$ is as high as the mean of the free-speed distribution) prevents strict less-than inequality, and the size of the road and the presence of traffic prevents the opposite. The use of these capacity functions can better be discussed after a discussion of a measure of congestion in the next section.

V. Measures of Congestion

In an attempt to bring into the capacity concept some measure of the inconvenience caused drivers as flow increases, Normann's group examined such variables as standard deviation from average speed, standard deviation of differences in speed from the preceding vehicle, and the actual number of

16/ [11, p. 229].
passings made within a given length of highway. The most sensitive index of 
relative interference between vehicles was found to be the mean difference in 
speeds between successive vehicles. It is assumed that when the difference 
in speeds between two successive vehicles is nearly zero, then the second 
vehicle is being impeded by the first. Since the chance that the desired 
speed of the second vehicle is exactly equal to that of the first is very 
small, the assumption is not unreasonable. The mean of these successive speed 
differences thus gives some indication of the number of impeded vehicles. 
As flow increases there is a growing tendency for the speeds of individual 
vehicles to be governed by the speeds of the preceding vehicles. A typical 
result of a linear regression (for the same road as in equation (3)) was 

\[ d_n = 6.14 - 0.0032h x_n - 0.0025h x_s \]  

(4)$^{17}$

where $d_n$ is the mean difference in speeds of successive vehicles on the lane 
carrying north-bound traffic.

According to this criterion then, two flow combinations $(x_n, x_s)$ and 
$(x'_n, x'_s)$ cause equal states of congestion on the north lane if they give 
rise to the same values of $d_n$ in equation (4). The effects on the south lane 
of course might be quite different.

It should be mentioned here that some more sophisticated measures of con-
gestion appear throughout the Normann articles. These involve first, the 
establishment of a general relation between time-spacing between successive 
vehicles and differences in speeds, and second, a complete determination (for 
a particular road) of the distribution of time-spacings for various values of 
$x_n$ and $x_s$. Then for given $x_n$ and $x_s$ this information tells us just what

$^{17}$ [12, p. 228].
fraction of the total flow finds itself at less than a specified time-spacing from the vehicles ahead. In particular it tells us what fraction of the flow is at zero time-spacing, and therefore the fraction which is prevented from traveling at desired speeds. This approach would seem to be almost perfect as an indicator of the incidence of congestion. Notice however that while such a measure would tell us how many vehicles were affected, it would not tell us how seriously they were impeded unless we were willing to make an assumption, say, that fastest vehicles are the first to meet with speed reductions. Whether or not this procedure is simpler, and is intended eventually to supplant the mean-speed-difference procedure described above is not clear. We shall not discuss this work further except to say that a careful statistical study of the way these spacing distributions are altered when \( x_n \) and \( x_s \) and some of the parameters change might lead to some very fruitful generalizations which would obviate the repeated measuring procedures which now appear to be necessary for precision.

Besides mean difference in speeds, what other simple measures of congestion might be proposed? The economic point of view suggests that we let the total amount of time which lost through speed reductions be our measure of nuisance. The average time lost per mile is

\[
k = \int_0^\infty \frac{1}{u} f(u) \, du - \int_0^\infty \frac{1}{u} f^*(u) \, du
\]

(5)

where \( f(u) \) is the distribution of existing speeds and \( f^*(u) \) the free-speed distribution. And since average time per mile is the inverse of harmonic mean speed it follows that

\[
k = \frac{\hat{u}^* - \bar{u}}{\bar{u}^* \hat{u}}
\]

(6)
where \( u \) and \( u^* \) are the harmonic means of the actual and the free-speed distributions respectively. In form (6) \( k \) is seen to be a normalized index. And since the flow would be less if \( f^*(u) \) really prevailed, it is true that the actual time loss is not as great as (6) indicates. This has a subjective aspect; it is an index of the average time loss which drivers think they are incurring. This is particularly appropriate for the purpose of indicating the nuisance of congestion for it takes into account the effect of the difference between actual and desired speeds, and leaves out the comparison of actual speeds on alternate routes which is handled separately and explicitly in the network model of [2].

We shall next show that, at least for the purposes of transportation models in their present stage of development, we may safely ignore the index suggested by Normann and described above -- mean difference in speeds. By equation (4) we know what flow combinations \((r_n, r_s)\) represent equal degrees of congestion, accepting for the moment that definition of the term. But using equation (3) it is seen that

\[
d_n - \frac{1}{3} \bar{u}_n = -8.53 + 0.00024 r_n - 0.00014 r_s
\]

so that, approximately,

\[
d_n \approx \frac{1}{3} \bar{u}_n - 8.53
\]  

(7)

since the coefficients of \( r_n \) and \( r_s \) are small, in each case less than a tenth the size of the coefficients in equation (4). If this relation holds in general, then it is possible to ignore \( d_n \) in determining the degree of congestion present; the same information can be derived from the values of mean speeds. Thus if

\[18/\] If the measure is acceptable which views congestion mainly in terms of lost time, it is not unreasonable to believe that comparisons of average speeds will also suffice, since they do not differ greatly from harmonic mean speeds.
\( \bar{u}_n \) and \( \bar{u}'_n \) are mean speeds for speed distributions \( g(u) \) and \( g'(u) \) respectively. Then \( g(u) \) will lead to a higher state of congestion than \( g'(u) \) if and only if \( \bar{u}_n < \bar{u}'_n \). One possible way to normalize this measure for purposes of comparison with other roads is given in equation (6). We shall proceed on the assumption that for all roads \( d_n \) may be expressed in terms of \( \bar{u}_n \) as in (7). It seems very likely that this sort of relationship should hold, and it permits considerable simplification. Only average speeds need be considered. 19/

VI. Concluding Remarks

In the model of centrally directed traffic [1] uniform capacity functions are used. Flows and speeds are stipulated for each road, the only requirement being that \( x \leq c(u) \) in every case. If the network happens to contain a two-lane road carrying both directions of traffic, the two lanes are treated as separate roads.

In the individual choice model [2] capacity functions \( \phi_n \) and \( \phi_s \) are used for two-lane roads carrying both directions of traffic. An extension to roads carrying only one direction of traffic, and to roads of more than two lanes meets with no new difficulties in principle. Capacity functions similar to those described above have been studied, and the data necessary to formulate them in the same form as (3) probably already exist. 20/ A full understanding of the way these \( \phi \) functions are used can only come about through a study of the network models themselves. However, we shall attempt to give some idea of their use.

19/ For a measure of congestion involving flows, see [17].

20/ See [15, p. 224 ff.].
We may first think of flows \( x_n \) and \( x_s \) as being given for each road of the network. A corresponding set of average speeds will result which satisfy the equations \( \phi_n(x_n, x_s) = \bar{u}_n \) and \( \phi_s(x_n, x_s) = \bar{u}_s \) for each road.\(^{21}\) For flow-pairs \((x_n, x_s)\) to be possible not only must \( \phi_n(x_n, x_s) \geq 0 \) and \( \phi_s(x_n, x_s) \geq 0 \), but also the mean speed differences (given by equations such as (4)) must be non-negative. Since \( d_n \) and \( d_s \) are reduced to zero at positive speeds, only the latter limitations are effective.\(^{22}\)

Figure 4 depicts a \( \phi_n \) surface in \( x_n - x_s - \bar{u}_n \) space. The heavy line on the \( \phi_n \) surface represents the boundary defined by \( d_n = 0 \) and \( d_s = 0 \). All possible points \((x_n, x_s, \bar{u}_n)\) are on the \( \phi_n \) surface to the right of this boundary. At the present stage of our information nothing very specific can be said about the relation of \( d_n \) and \( d_s \). While we should expect them to reach zero simultaneously, the published data are not detailed enough to check this. Figure 4 is based only on the considerations given previously.

\(^{21}\) The capacity functions for different roads are in general different of course. We omit the subscripts for brevity, since no confusion should result.

\(^{22}\) Compare equations (3) and (4).
In the actual analysis of the network the flow on a particular road will not be given, of course, but will depend on the prevailing speeds and upon the nuisance of congestion on that road. Since we wish here to avoid these questions of the "demand" for road use, we shall neglect the fact that speed (and desired speeds) affect flow. We shall speak of changes in flow and beg the reader not to ask where the additional flow comes from or why.

Suppose \((x_n, x_s)\) to be a possible flow-pair not on the boundary of such points. How is \(u_n\) affected when \(x_n\) increases? If \(x_s\) is unaffected by the change in \(x_n\) then \(\frac{\partial u_n}{\partial x_n} = (\phi_n) \frac{x_n}{x_n}\). The increase can continue until \(d_n\) or \(d_s\) becomes zero; a greater value of \(x_n\) would be incompatible with the given value of \(x_s\). This simple speed-flow mechanism is the one used in [2]. It asserts that neither flow has any superiority over the other and seems rather artificial if, in the above example, \(x_s\) is high: one would expect the north-bound flow to have a "right" to half of the road.

In his studies of intersections Greenshields has found that very often a heavy traffic flow may dominate a lighter cross flow; that is, the heavy flow can increase at the expense of the lighter flow, but not the reverse. This effect of domination by a heavy flow probably also holds for opposing flows on a two-lane road under certain conditions. Unless however the heavy flow is many times greater than the light flow, one would expect to see instead the reverse phenomenon: a domination of the heavier flow by the lighter. To the best of our knowledge,

---

23/ See [2].

24/ [4, pp. 66-67].
none of the published investigations throws any light on this matter whatsoever.

Some models of this speed-flow mechanism which embody the domination concept can be easily constructed, but it is felt, that, as yet, they are a little too speculative for presentation.


