CHARACTERIZING TRAFFIC VARIATIONS

V. A. BOLOTIN
Eatontown, New Jersey, U. S. A.

V. S. PROKOPOVICH
Leningrad, U. R. S. S.

ABSTRACT

Many important results have been obtained by authors studying the problem of traffic variation among time periods. However, there is much to be done, and therefore we suggest some ideas related to the problem.

Let us consider traffic (offered load) on a certain trunk group for all hours of the year. We assume that for every hour traffic is a r.v. with some d.f. A subset of hours with identical d.f.'s may be called a set of homogeneous hours, or a homogeneous set. Thus, traffic variations may be described in terms of time periods with 'homogeneous' traffic, which makes it possible to divide the analysis of traffic variations into two consecutive studies. The first one is analysis, in particular, nonparametric statistical analysis, of the set of all hours, which would result in separating homogeneous sets. Then it should be possible to analyze the different distribution functions for separate homogeneous sets.

Although a comprehensive model is the final goal in describing traffic variations, some results based on limited statistical data are useful both for understanding the whole phenomenon and for practical problem solutions. To illustrate this we discuss two examples.

One example suggests a mathematical model for the variance of offered traffic as a function of the mean value (equation (4)).

In the other example a classification of trunk groups in a local network is proposed. We define a class of trunk groups with the same position of the busy hour and with similar traffic variations during the day. Thus, some traffic characteristics can be measured and studied as they are related to classes of trunk groups rather than to individual groups.

INTRODUCTION

Consider a trunk group that serves calls through some switching network. The quality of service, or performance, is usually described by the loss probability, p, as a function of traffic intensity (offered load) A over some period of time, usually an hour.

The traffic intensity is a variable with its values defined on the set of all hours of the year. We assume that the traffic intensity A is a random variable with some distribution function \( F(x) = \Pr(A < x) \). Examples of the function p(A) (logarithmic scale) and distribution \( F(x) \) are shown in the third and second quadrants of Figure 1. They represent two fundamental parts of applied teletraffic theory. The loss probability curve, or p-curve, is the result of one of the traffic-within-hour analysis (methods for different connecting networks, different peakedness of traffic, etc.). The traffic distribution function \( F(x) \), or the A-distribution, is the result of the traffic-between-hours analysis. In other words, the p-curve is a characteristic of a traffic stochastic process, while the A-distribution is a characteristic of hour-to-hour traffic variations. We need both a p-curve and an A-distribution for a complete description of performance. The composition of \( p(A) \) and \( F(x) \) is presented in the first quadrant of Figure 1 as a distribution function of losses: \( F(y) = \Pr(p(A) < y) = \Pr(p(F^{-1}(x)) < y) \), or P-distribution (performance distribution).

The P-distribution is the complete performance characteristic, and with full knowledge of this we could supply solutions to many different practical problems. As soon as we recognize the importance of the P-distribution, we recognize the importance of the p-curve and A-distribution. There are many methods for obtaining p-curves analytically or experimentally, so the problem of deriving a P-distribution is mainly the problem of finding an A-distribution, i.e., the problem of hour-to-hour traffic variations.

Various mathematical models of A-distributions have been successfully studied and used in describing performance or service (e.g., [1-4]). This paper presents a viewpoint which may be helpful in understanding these problems.
SOME TYPICAL PROBLEMS

A number of problems of performance analysis and traffic measurements have long been considered and solved and more can be examined on the basis of the A-distribution. Here are several examples.

There exists an old but effective approach to traffic calculations for equipment dimensioning (see, e.g., [5]). The idea is that an extra amount of traffic has to be added to the mean value to account for the A-distribution effect, and the total result is considered as a value of traffic for equipment calculations. One practically realization of this idea suggests that we should consider some quantiles of the A-distribution. A simple example is a method in which two parameters of the A-distribution are considered. One parameter is (e.g.) the sum of the mean value and some factor times the standard deviation; the other is some quantile, e.g., 90th percentile. The first parameter is used for basic calculations under the prescribed level of loss. The quantile is used as a control value and provides some additional information about possible peak values in some small percentage of hours. Another quantile is used instead of the first parameter. It also may be reasonable to use more than two points of the A-distribution.

In this example of application we do not necessarily need full information about the A-distribution, and some quantiles may be sufficient. However, the quantiles provide some additional information about possible peak values in some small percentage of hours. Another quantile is used instead of the first parameter. It also may be reasonable to use more than two points of the A-distribution.

The next example of a typical problem is relevant where we have more advanced knowledge of the A-distribution. This is the problem of the choice between two approaches to performance characteristics of the A-distribution. The traditional approach is based on the busy hour period and on the average volume of traffic for the busy hour during a busy season. The other one is a new approach - extreme value engineering [6].

Suppose we possessed a full description of the A-distribution and, therefore, the P-distribution. Then we could analyze different aspects of the problem: which approach is more suitable and more effective in every particular situation? What kind of extreme value distribution should be chosen, bearing in mind both theoretical knowledge and the measurements obtained? It is possible that a solution may depend on some particular features of the A-distribution. Relations between both types of performance characteristics could also be derived.

One more example describes a way of possible extrapolation of a P-distribution, given for one trunk group, to other trunk groups. It is natural that many trunk groups carry traffic with the same type of variations. If we had a criterion for recognizing such identical groups, we could apply detailed knowledge of traffic variations on one group to another identical group, thus diminishing the volume of measurements needed or using only limited statistical data.

We describe an attempt at trunk group classification later in this paper.

FORMAL DESCRIPTION OF THE A-DISTRIBUTION

Consider a trunk group, and let $H$ be the set of all hours of the year. We assume that the traffic intensity for an hour is a random variable with some distribution function $F(x)$. A subset of hours, $H_1$, with the same distribution function $F_1(x)$ is defined as a set of homogeneous hours. Thus, $H$ has been separated into subsets $H = \bigcup H_1$. If $N$ is the number of hours within the whole year and $N_1$ is the number of hours in the set $H_1$, then $p_i = N_1/N$ is the proportion of occurrence of $H_i$. The distribution function for the A-distribution is then

$$F(x) = \sum_i p_i F_i(x), \quad p_i > 0,$$  

where $F_i(x)$ is the distribution function for the set $H_i$.

The mean value and $s$-quantile of the A-distribution are

$$\bar{X} = E(A) = \sum_i p_i \bar{A}_i = \sum_i p_i E_i(A), \quad X_s = \sum_i p_i X_{s,i},$$

where $X_{s,i}$ is the $s$-quantile for $F_i(x)$.

The last expression reminds one of a well-known and important fact - some extreme values of traffic, which may result in high losses, occur not only in the busy season but also during seasons and times of day with a relatively low mean value of traffic.

Although homogeneous sets of hours have been defined in terms of distribution functions, this does not mean that we need to know particular distribution functions in order to separate the whole set $H$ into homogeneous sets $H_i$, on the basis of statistical data. We are able to divide the statistical analysis of the A-distribution into two consecutive studies. The first one is an analysis of the set $H$, which is approximately non-parametric statistical analysis. For example, one could use nonparametric methods for testing whether two populations have the same distributions (see, e.g., [7]). The second is an analysis of the distribution functions on homogeneous sets of hours.

It should be noted that if homogeneous hours may consist of hours belonging to different seasons and to different hours of the day. For example, a set of busy hours that belong to a slow season may have the same distribution function as a set of non-busy hours that belong to a busy season. This may be useful knowledge both in analysis of the A-distribution and in measurement practices.

VARIANCE OF THE A-DISTRIBUTION

Let us consider a homogeneous distribution $F_i(x)$ with the mean value $A_i$. The traffic $A_i$, offered to the trunk group, is a sum of some original traffic flows $A_{ik}$, $k = 1, 2, \ldots$, could be considered as (approximately) mutually independent r.v. The variance of $A_i$ is

$$Var(A_i) = \sum_k Var(A_{ik}).$$

If we assume that the variance $Var(A_i)$ and $Var(A_{ik})$ are expressible, at least approximately, by a function of the form $g(A)$ and $g(A_{ik})$, then this formula becomes

$$g(\bar{X}_i) = \sum_k g(\bar{X}_{ik}).$$

The unique (continuous) solution to this functional equation is

$$Var(A_i) = g(\bar{X}_i) = h \bar{X}_i^2.$$  

(2)

The constant $h$ describes natural random hour-to-hour variations of offered traffic which may be
explained, e.g., as a fixed hour traffic variation from year to year.

Considering now the A-distribution for a non-homogeneous set of hours (not necessarily the set of all hours of the year) we can define the variance of the mixture distribution (1). The result is

\[ \text{Var}(A) = hA + \sum p_i A_i^2 - A^2, \]

(3)

where \( A \) is the mean value of the A-distribution. Let \( A_i = q_i A \). Then (3) changes to

\[ \text{Var}(A) = hA + r^2 - 1 > 0. \]

(4)

In this formula \( h \) represents the random component of hour-to-hour variations (homogeneous set), and \( r \) represents regular components of hour-to-hour variations (seasonal, during the day, etc.; non-homogeneous set).

Formula (4) suggests an additional explanation for using a polynomial variance approximation in processing measurement data. A very clear example of this type of approximation can be found in [8], formula (3):

\[ \phi = 1.5, 1.7, 1.84 \]

corresponding to low, medium and high amounts of day-to-day variation, respectively. Here \( A \) is the hourly traffic measured. This formula describes variations of the offered traffic as a function of the measured traffic, and it would be interesting to investigate relations between (4) and the first term of (5). (The second term is due to a finite measurement interval.) The first term of (5), which is empirically based, and (4) describe similar types of convex function.

The same type of approximation could be considered for quantiles of the A-distribution.

Estimation of the values of \( h \) and \( r \) (and corresponding coefficients for quantiles) leads to simple calculations of variance and quantiles on the basis of measured or forecasted mean values.

A CLASSIFICATION OF TRUNK GROUPS. BOUNDARY CASE

Let us consider a trunk group for traffic from one telephone area to another telephone area. We shall use the simplest type of subscriber classification: residential subscribers versus business subscribers. Let \( M_1 \) and \( M_2 \) be the number of residential and business subscribers in the first area. The ratio

\[ m = M_1/(M_1 + M_2) \]

is the fraction of the residential subscribers in the first area. In the same way we define \( N_1 \) and \( N_2 \) for the second area and the ratio

\[ n = N_1/(N_1 + N_2) \]

for the fraction of the residential subscribers in the second area.

We have \( \sum p_i = 1 \), and multiplying \( A_i = q_i A \) by \( p_i \) we obtain \( \sum p_i q_i A = \sum p_i A_i = A \). Then \( \sum p_i q_i A_i^2 = 1 - \sum p_i q_i^2 = 1 - \sum p_i (q_i - 1)^2 > 0. \)

Let \( A(t) \) be the average traffic intensity on the trunk group for the hour that starts at the moment \( t \). We can write as in [9]

\[ A(t) = M_1 N_1 a_{11}(t) + M_1 N_2 a_{12}(t) + M_1 N_2 a_{21}(t) + M_2 N_2 a_{22}(t), \]

where \( a_{ij}(t) \) is the average traffic for the same hour between one calling subscriber of the class \( i \) and one called subscriber of the class \( j \). We are not interested now in particular values of \( a_{ij}(t) \); the following approximate analysis does not depend on those values.

The average hourly traffic between one calling subscriber and one called subscriber is

\[ a(t) = A(t)/(M_1 + M_2)(N_1 + N_2) \]

\[ = ma_{11}(t) + m(1-n)a_{12}(t) + (1-m)na_{21}(t) + (1-m)(1-n)a_{22}(t). \]

(6)

With both \( m = 1 \) and \( n = 1 \) both areas are purely residential, and the maximum traffic value, i.e., busy hour traffic, usually occurs in the evening hours. This case of traffic variations during the day is represented by curve 1 in Figure 2.

![Figure 2. Different types of day traffic variation](image)

If \( m \) or \( n \) both decrease, the business traffic changes the shape of this curve 1 to something like the shape of the curve 2. There may exist such a pair of \( m, n \) values that the maximum value of the day traffic and the maximum value of the evening traffic become equal (curve 3). For this case we can write

\[ a(t_d) = a(t_e), \]

(7)

where \( t_d \) is the moment that starts the daytime busy hour, and \( t_e \) is the moment that starts the evening busy hour. From (6) and (7) we obtain (the exact values of \( t_d \) and \( t_e \) are not needed)

\[ a_{11}m + a_{12}(1-n) + a_{21}(1-m)n + a_{22}(1-m)(1-n) = 0, \]

(8)

where \( a_{ij} = a_{ij}(t_d) = a_{ij}(t_e) \) are constants.

We started from one pair of values \( m, n \) but equation (8) defines a set of points (a curve) inside the unit square that is shown on Figure 3. For all those points the evening value of the busy hour traffic is equal to the day value of
The average hourly traffic between two subscribers, as it has been defined by (6), is a function of time, percentage of residential (and business) subscribers in both areas, and average hourly traffic for a pair of subscribers in different classes:

\[ a(t) = a(t; m, n) = a(m, n; a_{ij}(t)), \quad i, j = 1, 2. \]  

(10)

Let \( g(t; m, n) \) be the derivative:

\[ g(t; m, n) = \frac{\partial a(t; m, n)}{\partial t} = \left[ g(m, n; \frac{da_{ij}(t)}{dt}) \right]. \]

(11)

For an arbitrarily fixed point \((m', n')\), other than any point on the curve (9), equation \( g(t; m', n') = 0 \) results in \( t' = t(m', n') \), which is the starting moment of the busy hour for this pair \((m', n')\) and for the corresponding set of points \((m, n)\) which satisfy the equation

\[ g(t; m, n) \bigg|_{t=t'} = g(m, n; a_{ij}') = 0, \]

(11)

This equation describes a family of trunk groups with some fixed position (time of day - from the moment \( t' \) to the moment \( t' + 60 \) minutes) of the busy hour. Equation (11) is the same as (8), but with \( a_{ij}' \) instead of \( a_{ij} \). Again, omitting the middle term we can describe a class of trunk groups by a hyperbola (9) with some other value of \( C: C = (a_{ij}', a_{ij}'')/a_{ij}' \). One more point \((m'', n'')\), other than any point on the curves (8) and (11), leads to a third family of trunk groups, \( g(m, n; a_{ij}'') \), and so on. The simplified solution gives \( \alpha \) the third hyperbola with the third value of \( C: C = C'' \) (see Figure 3). Thus, every value of \( C \) and its corresponding hyperbola on Figure 3 defines a class of trunk groups with the same position of the busy hour. From a practical point of view, a class may be defined by a family of neighboring hyperbolae. As before, for more strict analysis we should take into consideration the middle terms in (8) and for this derive some relations among the \( a_{ij}(t) \), which is possible to achieve from proper traffic measurements.

The described classification of trunk groups according to the position of the busy hour is an example of a classification related to a particular type of traffic variation. This specific classification is useful for analysis of the whole problem. But it is also important by itself because it is suitable for immediate practical application. This we will now explain.

AN EXAMPLE OF PRACTICAL APPLICATION

To demonstrate the usefulness of the simplified trunk group classification we used some statistical data from a local telephone network. Considering traffic measured every hour for 13 hours a day and for 24 days within a year, we located an approximate position of the busy hour for every trunk group and divided trunk groups into three main classes (with a morning, an afternoon and an evening busy hour) and four intermediate classes (with two or three nearly equal busy hours - morning and afternoon, morning and evening, and evening and morning). This classification was an experimental one. According to this simplified theoretical classification a class for a trunk group is defined by a pair of values \((m, n)\). Thus, every trunk group can be described by a point inside the unit square (Figure 3), and groups that have nearly the same busy hour are expected to belong to the same class.

This classification appeared to be in agreement with the experimental classification; nevertheless we prefer to refrain from final inferences and recommendations for two reasons. First, the statistical data were not reliable enough. Second, for better analysis some general cases of the equations (8) and (11) should be considered.

The analysis described here has been used in the
The main idea of the proposals is as follows. All trunk groups in the network are to be divided into classes according to the value of $C$ in (9). A range of values defines a class. At least one trunk group in every class has to be chosen as a "representative" group and assigned for detailed measurements. Measurements are limited for other trunk groups inside the class and a more detailed estimation of traffic is made on the basis of results obtained for the representative group(s). One particular example is the practice in which readings of traffic meters for a common trunk group are being carried out only twice a day - at the beginning and at the end of the measurement period (e.g., 9 a.m. - 4 p.m.). Then the traffic for the busy hour is estimated through the use of a busy hour load ratio, which is found from the results of traffic measurements on the representative trunk group. It has been found that the busy hour ratio for 5 to 8 hours (morning and afternoon) varies in the range of several percent within a class of trunk groups. Typical examples of the ratio ranges for two different classes (with two different values of $C$) are: 1.03 to 1.09, and 1.11 to 1.16. For a class of groups with an evening busy hour we had a range of 1.20 to 1.29, related to three hours.

CONCLUSION

Analysis of traffic variations has to be considered as one of the vital problems in the modern approach to performance definitions and evaluations. There are two types of analysis. A generalized approach provides definitions and studies of the whole distribution function which describes traffic variations during the year. Analysis and solution of some limited specific problems can serve two purposes. It often adds something to our understanding of the whole phenomenon of traffic variations and it demonstrates that there are immediate practical applications.

In this paper examples of both types of analysis was shown. The traffic variance has been examined as a function of the average traffic value on the basis of a generalized model for hour-to-hour traffic variations throughout the year. A model for trunk group classification according to the position of the busy hour has been suggested. Further comprehensive studies of the models described by formulae (I), (4) and (II) could justify them as useful tools for practical applications.

REFERENCES