Subscriber Traffic: Year-to-Year Variation, Day-to-Day and Group-to-Group Variation

Vladimir A. Bolotin

Bellcore (Bell Communications Research, Inc.)
Red Bank, New Jersey, U.S.A.

Several types of subscriber traffic variability should be accounted for in switching equipment engineering. Theoretical analysis and comprehensive quantitative studies of traffic in Bell companies have been performed over the last several years. Variation by geographical area, class of service mix and time of day was described in the ITC 10 paper \[1\]. In the ITC 11 paper \[2\] day-to-day variability of large subscriber traffic parcels (total exchange traffic) was described, in particular by a detailed analysis of the highest day traffic.

This paper extends studies of subscriber traffic variability to describe

- year-to-year variation of peak period traffic in large subscriber groups,
- the combined effect of day-to-day and group-to-group variability of traffic generated by subscriber line groups of limited size.

As in all previous analyses, these results have been applied to traffic engineering and traffic capacity analysis of modern digital switching systems.

The complete picture of subscriber traffic variability is intended primarily for the refinement of engineering methods used by the Bell companies. It may also provide guidance for engineering in the developing telephone networks of the world, and aid in identification and analysis of traffic characteristics specific to ISDN new services.

1. Year-to-Year Variability of Extreme Traffic

1.1 Time Variability of Subscriber Traffic

Several types of traffic time variability have been analyzed and included into traffic engineering methods. Historically, the time of day variation was taken into account first - the average (ABS) traffic used for engineering was related to the busiest hour of the day. Later, the day-to-day variation was analyzed, and time consistent busy hour as well as extreme value engineering methods were developed.

A major problem of digital switch traffic engineering is the boundary traffic value that separates expected peak values, used in engineering, from overload values. The boundary value is important because normal functioning of a device, e.g. a central processor, is fundamentally different from its functioning at overloads. The quality of service characteristics are also fundamentally different. According to the existing engineering methods, the capacity is normally quantified as the boundary between normal functioning and overload, and this boundary is to coincide with the highest day (HD) traffic volume. The HD traffic may be defined as the highest one-hour (half-hour, fifteen-minute) traffic during one year, at a fixed or undetermined hour depending on the engineering method. The day-to-day variation of traffic belongs to the traffic pattern within the largest well defined
time period for projecting peak traffic: one busy season (one year). The year-to-year variability cannot be captured by the day-to-day variation characteristics and presents an additional uncertainty factor needed for the engineering decisions about the aforementioned boundary value. In some years, the offered traffic could be below the switching system capacity, in other years it could cause system overload. We will describe this variation and show that it is large enough to be taken into account in the engineering process.

1.2 Year-to-Year Variability of Subscriber Traffic

The analysis of the year-to-year variation is essentially based on the linear relation between peak subscriber traffic and the average busy season (ABS) traffic [2]. Fig. 1 shows a typical example of the linear relation phenomenon. Both HD and ABS call rate shown in Fig. 1 are linear functions of the class of service mix. Consequently their difference, called the ABS-to-HD increment, is a linear function of the ABS call rate.

The ABS and HD values shown in Fig. 1 are the expected traffic values that could be used in engineering. In each particular year the real (observed) ABS or HD value deviates from the expected value. For ABS engineering procedures, this deviation is not very important but for HD engineering it is critical since the HD capacity is the boundary separating normal functioning and overload condition.

A direct estimation of the year-to-year HD traffic variability is impossible since the ABS and, correspondingly, the HD values may have a year-to-year deterministic change component, e.g. because of gradually increasing traffic. As we will show it is their difference that contains only random year-to-year variation.

Fig. 2 shows the expected value and some sample points of the ABS-HD increment. Typical year-to-year variation (1978 - 1983) of the observed ABS-HD increment is shown for six offices. Also shown are two regression percentile lines for the year-to-year variation. Analysis of residuals showed that the range of the increment variability is practically the same within the full range of ABS call rate, from one call hour - residential traffic - to 4.5 calls hour - predominantly business traffic. Direct analysis of the increment standard deviation based on later observations, 1981-1986, also showed that the increment variability is practically independent of the ABS traffic, i.e. of the class of service mix. This is
illustrated in Fig. 3 by sample points and a regression line for 14 offices, six years of data each office.

This observation is crucial for studies in which available data do not represent enough years to estimate year-to-year variability for individual offices. As an extreme case, a regression line similar to the one depicted in Fig. 2 can be obtained even with only one year of data for many offices with different class of service mix populations. The standard deviation of the individual observation in this regression would be taken as an estimate of the year-to-year increment variability. That would mean that observations at different offices with nearly the same ABS call rate are viewed as if they characterized different years for a single office. As a result of this simplification, the increment standard deviation will be overestimated since there exist differences among offices caused by factors other than the class of service mix. A direct estimation of the increment variability office by office based on data for several years would yield unbiased results.

A comparison of the two estimation methods is depicted in Fig. 4. (the same data as for Fig. 3 were used). Each curve shows the increment standard deviation divided by the ABS call rate. The lower curve represents the average of the standard deviations for single offices shown in Fig. 3. In the upper curve, the standard deviation of the regression shown in Fig. 2, i.e., 0.2 calls/hour, is used. (The 90% upper band for individual observations shown in Fig. 2 coincides with approximately one standard deviation from the regression line - see also section 1.4.) If data for the more accurate method corresponding to the lower curve are not available, variability estimation should be made by the first method and the results scaled down according to Fig. 4.

1.3 Analysis

Standard deviations of the ABS-HD increment based on the variance in regression were obtained for each Bell company. For two Bell companies the results are shown in Table I. Table I contains typical examples of the year-to-year high day traffic variation for the three busy periods of the day, and for each period the range of HD values generated by the range of class of service mix. These examples represent typical low and high variability in Bell companies. The standard deviation of the ABS-HD increment is shown in column 4 (.0041 Erl. in the first row) In column 5, the expected HD value range is shown. For the morning and the afternoon traffic, the lowest value (0.07 Erl. in the first
row) is associated with residential lines, the highest (0.18 Erl.) with business lines. For the evening traffic, the lowest and the highest values describe residential traffic within the range of special service (Call Waiting etc.) penetration. The year-to-year coefficient of variation is the ratio of the standard deviation in column 4 to the expected HD value in column 5 (6% for the lowest, residential, usage and 2% for the highest usage). This variability range is shown in column 6. (The ABS call rate and load are assumed to be fixed.) These numbers would be reduced by a factor of about 0.6, according to Fig. 4, but indicate clearly that the year-to-year variation of high day traffic is very significant from an engineering standpoint.

### Table I: Year-to-Year Variability of High Day Subscriber Traffic

<table>
<thead>
<tr>
<th>Bell Company</th>
<th>Line Traffic Variable</th>
<th>Time of Day</th>
<th>Standard Deviation of Increment</th>
<th>Expected HD</th>
<th>Coeff. of Var.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TelCo I</td>
<td>Usage (Erlangs)</td>
<td>Morning</td>
<td>.0041</td>
<td>.07 – .18</td>
<td>6% – 2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Afternoon</td>
<td>.0044</td>
<td>.07 – .11</td>
<td>6% – 4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evening</td>
<td>.0036</td>
<td>.07 – .12</td>
<td>5% – 3%</td>
</tr>
<tr>
<td></td>
<td>Call Rate (Calls Hour)</td>
<td>Morning</td>
<td>0.20</td>
<td>1.3 – 5.5</td>
<td>15% – 4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Afternoon</td>
<td>0.15</td>
<td>1.2 – 3.0</td>
<td>13% – 5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evening</td>
<td>0.12</td>
<td>1.2 – 2.4</td>
<td>10% – 5%</td>
</tr>
<tr>
<td>TelCo II</td>
<td>Usage (Erlangs)</td>
<td>Morning</td>
<td>.0072</td>
<td>.07 – .18</td>
<td>10% – 4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Afternoon</td>
<td>.0056</td>
<td>.07 – .12</td>
<td>8% – 4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evening</td>
<td>.0069</td>
<td>.07 – .19</td>
<td>10% – 4%</td>
</tr>
<tr>
<td></td>
<td>Call Rate (Calls Hour)</td>
<td>Morning</td>
<td>0.24</td>
<td>1.4 – 5.8</td>
<td>17% – 7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Afternoon</td>
<td>0.24</td>
<td>1.4 – 3.6</td>
<td>17% – 5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evening</td>
<td>0.23</td>
<td>1.3 – 3.2</td>
<td>18% – 7%</td>
</tr>
</tbody>
</table>

1.4 Application

The coefficient of variation translates into probabilities of traffic volume being under or over system capacity.

Comparison of observed increment distributions with the Normal and Gumbel [3] distributions showed that practical estimation of the frequency of increment values causing overloads could not be based on these distributions quantiles. The median of the increment distribution is significantly lower than its expected value. the expected value is between the 65th and 70th percentiles. The 90th percentile is at the level of one standard deviation over the expected value (vs. 1.3 standard deviations in the Normal and Gumbel distributions).

Therefore if the average increment value is used for traffic engineering (the middle line in Figure 2), then the offered traffic will exceed the system capacity roughly every third year. Adding one standard deviation to the HD call rate would reduce this risk to every tenth year. Comprehensive data similar to those in Table I are being used for developing recommendations with respect to the year-to-year variability of HD traffic and the engineering margin needed for overload protection.
2. Group and Day Variability of Subscriber Traffic

2.1 General Model

This section outlines the model of day and group variation described in [14]. The model was developed in order to reconcile different approaches to traffic variation among and within subscriber groups. The idea of the model is to extract the "pure" group-to-group variation component from the combined group and day variation of subscriber line traffic.

Generally, usage and call rate measured on subscriber line groups have both group-to-group and day-to-day variability components in them. Increasing the group size one would asymptotically decrease the group-to-group variation. Performing a series of group-to-group variation studies with one day traffic, one would eliminate the day-to-day variation. Consequently, we can define the day variability as day-to-day variation of the general population traffic and the group-to-group variability as group-to-group variation of one day (usually fixed hour) traffic.

Let the r.v. \( Y \), with the c.d.f. \( L(y) \), be one-day one-hour line traffic (line usage or call rate) in a given population of lines. For a group of \( n \) lines \( (n \geq 1) \), the traffic volume and the c.d.f. are

\[
Y_n = \sum_{s=1}^{n} Y^{(s)}, \quad Y^{(s)} = Y; \quad L_n(y) = L^n(y).
\]

These c.d.f.

\[
\text{describe "pure" group-to-group variation.}
\]

A family of distribution functions \( L_n(x,y) \) with a random parameter \( X \) is introduced. The c.d.f. of \( X, U(x) \), characterizes day-to-day variation.

The c.d.f. of the day-to-day variation of an \( n \)-line subscriber group traffic (combined group and day variability) is

\[
F_n(y) = \int L_n(x,y) \cdot dU(x), \quad E(Y_n) = E(E(Y_n | X)) = n \cdot E(X),
\]

(2-1)

In the following, analysis of coefficients of variation is based on the variance

\[
Var(Y_n) = E(Var(Y_n | X)) + Var(E(Y_n | X)) = E(Var(Y_n | X)) + n^2 \cdot Var(X).
\]

(2-2)

If the group coefficient of variation is the same for all days, i.e., depends on \( n \) but does not depend on \( x \), then

\[
CV_n \equiv CV(Y_n | X) = \frac{\alpha}{\sqrt{n}}, \quad \alpha = \text{const}.
\]

(2-3)

From that follows a clear-cut separation of the combined day-to-day and group-to-group variation into components. The day-to-day variability of the \( n \)-line group traffic is described by

\[
CV_{d,n}^2 = CV_{d,x}^2 + CV_n^2 + CV_{d,x}^2 \cdot CV_n^2.
\]

(2-4)

\( CV_{d,x} \) describes the day-to-day variation for the total population of lines (pure day-to-day variability).
Figure 5 illustrates formula (2-4). The four curves represent combinations of two fixed day variability levels, $CIV_{d, \infty} = 0.075$ and $0.15$, and two fixed group variability levels, $CIV_1 = 1.1$ and $2$. The point of intersection of two curves shows that the same combined variation may occur as a different combination of the day and group variation components. At the point of intersection, the two curves define the same total variability, but they have different day variability and different group variability. These two cases represent two different classes of subscribers.

Limitations of this model and generalizations removing these limitations are discussed in the following section.

2.2 New Results

The model developed in $^4$ is limited by assumption (2-3) and refers to a homogeneous subscriber population (one class of service) and to line groups that are subject to load-directed assignments for balancing.

If assumption (2-3) does not hold, $CIV_n$ depends on $X$. This model is discussed in more detail in section 2.2.1. A model for unbalanced traffic is given in section 2.2.2.

For several classes of service (section 2.2.3), the expected values and variances of the subgroup of each class can be obtained as before, but for the total population the coefficients of variation will not add as they do in formula (2-4).

2.2.1 Coefficient of Variation Depends on Total Day Traffic

There are indications that the variability of line unit traffic may deviate from the assumption (2-3). For example, on a day with a large total load, group loads may be relatively closer to each other than on a day with a small total load. $CV(Y_n | X)$ in (2-2) needs to be described by a more complicated model. Generalizing (2-3), we assume that

$$ CV_n^2(X) = CV^2(Y_n | X) = \frac{1}{n} \left( \alpha^2 \pm \frac{\beta}{X} \right). \quad (2-5) $$

$\beta = 0$ corresponds to (2-3). It can be shown that this model leads to the following generalization of (2-4)

$$ CV_{d,n}^2(\beta) = CV_{d,n}^2 \pm \frac{\beta}{n \hat{E}(X)}, \quad (2-6) $$

where $\hat{E}(X)$ is an estimate of the average (i.e. ABS) traffic. The coefficient $\beta$ should be estimated by values of $CV_n(X)$ measured on days with different values of $X$.

A simpler approach would be based on direct estimates of $E(X)$ and $E(Var(Y_n | X))$:

$$ \hat{CV}_n^2 = \frac{\hat{E}(Var(Y_n | X))}{n^2 \hat{E}^2(X)}. \quad (2-7) $$

2.2.2 Unbalanced Traffic

Usually, $n$-line groups are not balanced ideally: the long-term average (ABS) value is different in different groups. To account for this phenomenon we randomize the day-to-day variation distribution $U(x)$:

$$ F_n(y,z) = \int L_n(x,y) u(x,z) dx. \quad (2-8) $$
If \( s(z) \) is the distribution of the random parameter \( z \), the c.d.f. for the total population becomes a mixture

\[
F_n(y) = \int s(z) dz \int L_n(x,y) dU(x,z).
\]

This is the same as (2-1), with

\[
U(x) = \int u(x,z) s(z) dz. 
\]

Similarly to (2-4)

\[
CV_{d,n}^2 = CV_n^2 + CV_d^2 + CV_b^2 + CV_{b,n}^2 (CV_d^2 + CV_b^2) \approx CV_n^2 + CV_d^2 + CV_b^2.
\]

(2-11)

\((CV_d \equiv CV_{d,\infty})\) where \( CV_b \) is the coefficient of variation among long-term averages in different \( n \)-line groups. The smaller \( CV_b \) the better these groups are balanced.

### 2.2.3 Nonhomogeneous Subscriber Population

Generally, an \( n \)-line group contains lines belonging to different classes of service, so that

\[
n = \sum_{i=1}^{k} n_i,
\]

(2-12)

where \( n_i \) is the number of lines of class \( i \) in the group. Each subscriber group of \( n_i \) lines is homogeneous, and the models described in the previous sections are applicable to it. In particular, analogously to the definition in section 2.1, let \( Y^{(i)} \) be one-hour one-line traffic (usage or call rate) in the \( i \)-th sub-population of lines. Similarly to (2-0), the total traffic in the \( n \)-line group is a r.v.

\[
Y_n = \sum_{i=1}^{k} Y_{n_i}^{(i)} = \sum_{i=1}^{k} \sum_{i=1}^{n_i} Y^{(i,s)} = Y^{(i)} - i.i.d. \text{ r.v.}.
\]

(2-13)

where \( Y^{(i)} \) is the \( i \)-class of service component of the total traffic.

Analogously to formulae in sections 2.2.1 and 2.2.2,

\[
L_{n_i}^i(y) = L_{n_i}^{n_i}(y), \quad F_{n_i}^i(y) = \int L_{n_i}^i(x,y) dU_i(x).
\]

(2-14)

All subsequent formulae are applicable to each sub-population with \( n_i \) lines. The day-to-day variability characteristics may be different for different classes. Finally, it can be shown that

\[
Var(Y_n) = \sum_{i=1}^{k} n_i \cdot CV_2(Y^{(i)}) \cdot E(X_i^2) + \sum_{i=1}^{k} n_i^2 \cdot CV_2(X_i) \cdot E^2(X_i).
\]

(2-15)

These formulae may be used for obtaining variability characteristics similar to those shown in Fig. 5.

### REFERENCES


