The Integration of Domestic and International Networks for Switch Planning

S. G. Low and J. A. Schmitt
AT & T Bell Laboratories
Holmdel, New Jersey 07733

Capacity planning for International Switching Centers requires that the effects of the international alternatives on the AT&T nationwide domestic network be integrated into the plan evaluation. The methods to solve the unique technical issues arising from this integration are presented.

1. INTRODUCTION

The current rapid growth in the demand for global telecommunications makes the planning of International Switching Center (ISC) capacity critical to the quality of international service. Changes in the locations of the U.S. ISCs serving the foreign countries and reconfigurations of the international networks between them and the ISCs in the foreign countries cannot be effectively planned in isolation. The analysis of these alternatives must take into account the impact on the nationwide domestic AT&T Switched Network (ASN) over which the traffic is routed and the subsequent effect on the utilization of the capacity of switches in the network. Thus, the international and domestic ASNs must be planned as an integrated entity.

This paper focuses on three issues that must be resolved to enable the analysis of the impact on the domestic network to be integrated into the planning of the international networks and ISC switch capacity.

1. A method of reallocating the international component of the load that traverses the domestic network based on changes to the international network topology and loads is described. For example, a call from southern California to a European country may currently be routed to an ISC switch in New York City. This traffic is routed domestically through the access network, perhaps through a portion of the hierarchy to a Dynamic Non-Hierarchical Routing (DNHR) switch, and then across the US to the east coast over the DNHR network to the New York ISC. See Figure 1. Changing the serving ISC to the ISC in Atlanta shifts the international traffic to and from the foreign country currently terminating on the New York ISC to the Atlanta ISC and changes the routing of this load over the domestic network. To evaluate the impact of the international network changes on the national domestic network, the traffic between the effected U.S. region and the country must be reallocated on different domestic links and the effects on the trunking requirements across the domestic network must be determined.

2. International calls may originate or terminate anywhere in the United States. The resizing of the network based on the above reallocation of the international loads must be done on a nationwide basis. The AT & T hierarchical network currently involves approximately 500 toll (transit) switches. To enable an integrated planning tool to operate in near real time with reasonable computer resources, an aggregation scheme for the hierarchical network is used. Although the idea of significantly reducing the number of switches by aggregation is not new, to do so in a manner that preserves terminations and trunks group sizes required for switch utilization and economic analysis calculations is. Because all the ISCs are DNHR switches, international traffic is routed through the hierarchical portion of the network to the DNHR network and then to the serving ISC. We implement aggregation at the bottom of the hierarchy (Class 4) where the amount of international traffic is smaller.
ITC 12 Torino, June 1988

and neglect these switches in any evaluation of the alternatives. This is appropriate because the small amount of international traffic between class 4 hierarchical switches remains invariant between most switch planning alternatives.

3. The physical termination limit of digital switches may not be attained due to processor throughput and memory limitations. This is particularly true of ISC gateways that require even more processor and memory intensive operations than their domestic counterparts. Models of these capacity constraints must be formulated to give a reliable estimate of switch exhaust dates.

![Diagram of network routing](image)

**FIGURE 1** SCHEMATIC OF THE DIFFERENT ROUTINGS OF A CALL BETWEEN THE U.S. AND THE FOREIGN COUNTRY THROUGH DIFFERENT PARTS OF THE NATIONAL NETWORK TO THE INTERNATIONAL NETWORK FOR TWO INTERNATIONAL CONFIGURATIONS

2. INTERNATIONAL CALL ROUTING

The international traffic between a foreign country and the domestic switch within a switching region is routed to an unique serving ISC. Each switching region is defined in terms of a set of Numbering Plan Areas (NPAs). The union of these switching regions is the U.S.. The relationship between the switching region and its serving ISC is referred to as the *switching region homing*, and it can vary with each country. The fraction of busy season-busy hour load to a given foreign country that originates or terminates in a switching region is referred to as the *switching region fraction*. International traffic traverses the domestic network between the domestic switch that originates or terminates the call and the serving ISC. Between the serving ISC and the ISC in the foreign country, the traffic is routed on the international network. See Figure 1.
The analysis of alternative planning scenarios for one or many countries involves changes to the international loads and network topologies, and to the switching region homings and fractions. The assumption is made that the domestic network topology is not affected by changes to the international network, that is, the end office and hierarchical switch homings do not change and switches designated as DNHR are fixed. Thus, the size of the access trunk network is invariant. However, based on the reallocation of the international load, the hierarchical and DNHR trunk group sizes can change, which in turn impacts the ISC and utilizations of the domestic switches.

The methods to accomplish the load reallocation for arbitrary changes of the type described above, the aggregation of the hierarchical network, and the modeling of real time and memory capacity follow.

3. LOAD REALLOCATION METHOD

There are two steps to the reallocation algorithm. The first is to ensure that the independent load forecasts used for the domestic and international networks are commensurable, and the second is the reallocation equations themselves. Both steps are based on the conservation of load at an ISC that states that the total load between all the foreign countries routed through an ISC \( i \) to a switching region \( r \) in time period \( t \) in forecast year \( y \) must be identical for the international and domestic forecasts:

\[
TL^I_{iry} = TL^D_{iry}. \tag{1}
\]

\( TL^I_{iry} \) is obtained from the international load forecast \( L_{cly} \) between the U.S. and foreign country \( c \) in load set period \( t \) and forecast year \( y \) by:

\[
TL^I_{iry} = \sum_c f_{ icy} L_{cly}, \tag{2}
\]

where the sum is over all the foreign countries \( c \). \( f_{ icy} \) is the fraction for the load between foreign country \( c \) and the U.S. originating or terminating in switching region \( r \) that is routed through ISC \( i \) in year \( y \).

\( TL^D_{iry} \) is calculated from the domestic forecast of the international component of the load, \( L_{isy} \), between the ISC \( i \) and the other domestic switches \( s \) in load set period \( t \) and forecast year \( y \) by

\[
TL^D_{iry} = \sum_{s \in R_r, s \neq i} \kappa_{is} L_{isy} + \delta_{ir} L^D_{isy}, \tag{3}
\]

\[\delta_{ir} := 1 \text{ if } i \in R_r \text{ and } 0 \text{ otherwise.}\]

\( R_r \) is the set of switches in switching region \( r \). \( \kappa_{is} \) is equal to 1/2 if \( s \) is also an ISC and 1 otherwise. \( L^D_{isy} \) is the component of the access load that is routed directly through the ISC into the international domain. The reconciliation is accomplished by adjusting either the international or domestic load values or both to insure that the identity given in equation (1) is satisfied.

Changes to the switching region homings, fractions and the international load and topology is denoted by:

\[f_{ icy} \rightarrow f'_{ icy}, \quad L_{cly} \rightarrow L'_{cly} \]

The new international component of the loads on the domestic network are given by

\[
L_{isy}' = \begin{cases} \sum_c \sum_{r} \rho_{recy} f'_{ icy} L'_{cly}, & s \in R_r \setminus I, s \neq r, \ \ \\
\sum_c \sum_{r} [\rho_{recy} f'_{ icy} + \rho_{ icy} f'_{ recy}] L'_{cly}, & s \in R_r \cap I, s \neq r, \end{cases} \tag{4}
\]

\[
L^A_{isy} = \sum_c \sum_{r} \rho_{ icy} f'_{ icy} L_{cly}, \quad i \in R_r.
\]

1. Strictly speaking, the sum over \( r \) in these equations may be omitted as \( \rho_{recy} = 0 \) if \( s \) is not an element of \( R_r \).
where $I$ is the set of ISCs in year $y$ and $R_r \backslash I$ denotes the complement of $R_r$ relative to $I$. The quantities $\rho_{rc}$ are the fraction of the load between switching region $r$ and foreign country $c$ that originates or terminates at switch $s$ in load set period $t$ in forecast year $y$. The assumption that is critical for the load reallocation algorithm is that these fractions, $\rho_{rc}$, are invariant for the set of alternatives being considered. Furthermore, these fractions are independent of the particular ISC $i$ that the traffic to a country $c$ is being routed through for a given alternative, the switching region fraction and also the total load to this country. It is for this reason that equations (4) are valid for alternatives involving arbitrary changes to $f_{rc}$ and $L_{ct}$. The quantities $\rho_{rc}$ are determined from the fractions $\frac{L_{ct}}{TL_{ct}}$ defined by:

$$\rho_{rc} := \frac{L_{ct}}{TL_{ct}}.$$  

through the change of indices,

$$\rho_{rc} := \xi_{rc},$$  

where $i_r$ is the unique ISC on which switching region $r$ homes for country $c$ in year $y$. This apparently innocuous change of indices is the key to the algorithm.

4. AGGREGATION METHOD

Consider a network that consists of a set $S$ of switches. The point to point loads for this network may be represented by a symmetric matrix without a diagonal:

$$L_{sd}, \ s, d \in S, \ s \neq d.$$  

(In this section the load set period $t$ and forecast year $y$ subscripts are suppressed for simplicity.) The number of point to point loads is $\dim(S)(\dim(S)-1)/2$, where $\dim(S)$ denotes the dimension of $S$. For most hierarchical networks, the majority of these point to point loads are between the switches that are lowest in the hierarchy. In the the US network, these are referred to as Class 4 switches. Aggregation may be used to reduce the number of switches by subdividing the network into two disjoint sets:

$$S = S_0 + S_1,$$

$$N = \dim(S_0), \ M = \dim(S_1),$$

where $S_1$ are the set of lowest class switches (Class 4) and $S_0$ the remainder. See Figure 2. The sets of switches homing on each switch $s$ in $S_0$ partitions the switches in $S_1$ into disjoint sets of switches $H_s$, that is,

$$S_1 = \sum_{s \in S_0} H_s.$$  

Aggregation is accomplished by replacing each switch $s$ in $S_0$ and the set $H_s$ of switches homing on it with a single aggregate node. An aggregated node is composed of one switch from $S_0$ (type A) and one switch (type B) representing all the switches in $H_s$. We denote the set of type A switches as $S_A$ and the set of type B switches as $S_B$. This is illustrated in Figure 3. Note that there is a one to one correspondence between the three sets $S_0, S_A$ and $S_B$. The point to point loads between aggregated nodes can then be divided into four categories:

1. The loads $L_{sd}^{AA}$, $s \neq d$, $s, d \in S_A$ between type A switches in the aggregate nodes.
2. The loads $L_{sd}^{AB}$, $s \in S_A$, $d \in S_B$ between the type A and type B switches in different nodes.
3. The loads $L_{sd}^{BB}$, $s \neq d$, $s, d \in S_B$ between the type B switches in the aggregate nodes.
4. The loads $L_{sd}^{PP}$, $s \in S_A$, that are the estimates of the total offered loads between the switches in $H_s$ and the switch $s$ on which they home. This load is used to size the final between the type A and B switch in each aggregate node. There are $N$ loads of this type.
These loads are computed in terms of the original point to point loads $L_{sef}$ by the equations:

\[
L_{sef}^{AA} = \omega_s \omega_f L_{sef}^0, \quad s \neq s', \quad s, s' \in S_A,
\]

\[
L_{sef}^{AB} = \omega_s (1-\omega_f) L_{sef}^0, \quad s \in S_A, s' \in S_B,
\]

\[
L_{sef}^{BB} = (1-\omega_s)(1-\omega_f) L_{sef}^0, \quad s \neq s', \quad s, s' \in S_B,
\]

\[
L_{sef}^F = \sum_{h \in H_s} L_{sh} + OL_{sef}^1 + OL_{sef}^2, \quad s \in S_A.
\]

In these expressions, the loads $L_{sef}^0$ are the aggregate loads between node pairs defined by

\[
L_{sef}^0 = \sum_{k \in H_s \cup \{s\}} \sum_{k' \in H_s \cup \{s'\}} L_{kk}, \quad s \neq s', \quad s, s' \in S_0.
\]

The weights $\omega_s$ are given by:

\[
\omega_s = \frac{TL_{sef}^A + OL_{sef}^1}{TL_{sef}^A + TL_{sef}^B},
\]

where $TL_{sef}^A$ and $TL_{sef}^B$ are the loads from type A and B switches to the world, respectively. They are defined by:

\[
TL_{sef}^A = \sum_{a \in S \setminus H_s} L_{sa}, \quad s \in S_A.
\]

\[
TL_{sef}^B = \sum_{h \in H_s} \sum_{k \in S \setminus H_s} L_{kh}, \quad s \in S_B.
\]

The loads $OL_{sef}^1$ and $OL_{sef}^2$ are empirical estimates of the loads for high usage groups between B type switches that do not prove in and the overflow from those that do, respectively. They are given by:

\[
OL_{sef}^i = \sum_{h \in H_s} \sum_{k \in S \setminus (H_s \cup \{s\})} f^i(L_{hh}) L_{hh} \quad i = 1, 2, \quad s \in S_B,
\]

with

\[
f^1(L_{hh}) = \begin{cases} 1 & L_{hh} < \alpha \\ 0 & L_{hh} \geq \alpha \end{cases}
\]

and

\[
f^2(L_{hh}) = \begin{cases} 0 & L_{hh} < \alpha \\ \delta & L_{hh} \geq \alpha \end{cases}
\]
α is an empirical parameter estimating the load needed for a high usage group to prove in and δ is an estimate of the fraction of the load that overflows for those that do.

The hierarchical network is trunked based on these loads. Consider a pair of aggregated nodes with the type B switch representing \( x \) switches in one node and with the type B switch in the other node representing \( y \) switches. A maximum of \((1+\ x+\ y+\ xy)\) trunk groups could be engineered between the unaggregated switches. Instead only four average trunk groups are sized between the aggregated nodes: one to represent the \((xy)\) equal HU groups between the type B switches with average load \( L_{AB}^{xy} / (xy) \), one to represent the \( x \) HU groups between the type B switches and the type A switch with the average load of \( L_{AB}^{BA} / x \), one to represent the \( y \) HU groups between the type A switch and the type B switches with the average load of \( L_{BA}^{AB} / y \), and one group between the type A switches with load \( L_{BA}^{AA} \). The trunks between the type B switches in the aggregate nodes are not needed for termination counts but are computed to determine the correct overflow to the alternate paths between the aggregated nodes. The loads \( L_{f}^{y} \) are used to size a final between the type A and B switches within an aggregate node.

This aggregation dramatically reduces the number of point to point loads and trunk groups that need to be sized. The original network has a maximum of 
\[
\frac{(N+M)(N+M-1)}{2}
\]
point to point loads and trunk groups whereas the aggregated network has a maximum of 
\[
\frac{N(N+1)}{2} \quad \text{and} \quad N(3N-2)
\]
loads and trunk groups, respectively*. In cases that \( M \gg N \), this saving is dramatic.

The accuracy of the aggregation scheme is illustrated in Tables 1 and 2 by comparing the sizing of an aggregated network with the sizing of an unaggregated network. We identify the individual switches in the unaggregated network that form each aggregated node and sum the sizes of all the trunk groups between them to produce the correct total number of trunks between the aggregated nodes and the correct number of switch terminations for each aggregated node. The example is a four level hierarchical network with 271 switches of which 76 are class 4 switches. The aggregated network has 195 nodes. The number of load records are reduced from approximately 22,000 for the unaggregated case to approximately 9,000 for the aggregated case.

**TABLE 1.** Distribution of the differences in the trunk group sizes between aggregated nodes and the sum of trunk groups between unaggregated switches that constitute the aggregated node.

<table>
<thead>
<tr>
<th>Difference in Trunks</th>
<th>Percentage of Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>88.92</td>
</tr>
<tr>
<td>12</td>
<td>8.25</td>
</tr>
<tr>
<td>24</td>
<td>1.64</td>
</tr>
<tr>
<td>36</td>
<td>0.52</td>
</tr>
<tr>
<td>48</td>
<td>0.37</td>
</tr>
<tr>
<td>60</td>
<td>0.02</td>
</tr>
<tr>
<td>72</td>
<td>0.14</td>
</tr>
<tr>
<td>84</td>
<td>0.02</td>
</tr>
<tr>
<td>96</td>
<td>0.09</td>
</tr>
<tr>
<td>108</td>
<td>0.00</td>
</tr>
<tr>
<td>120</td>
<td>0.02</td>
</tr>
</tbody>
</table>

* The loads are the sum of \( N(N-1)/2 \) aggregated loads \( L_{ef}^{0} \) and \( N \) loads on the final \( L_{f}^{y} \). The trunk groups result is derived from the fact that four trunks groups exist between each pair of aggregated nodes plus \( N \) finals. The weight for each aggregated node must also be computed and stored.
TABLE 2. Percent differences in toll switch terminations between those at the type A switch of the aggregated nodes and those at the corresponding unaggregated switches.

<table>
<thead>
<tr>
<th>Percentage Difference in Terminations</th>
<th>Percent of Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10</td>
<td>94.5</td>
</tr>
<tr>
<td>10-20</td>
<td>5.4</td>
</tr>
</tbody>
</table>

Of the 4268 aggregated trunk groups, 97% of them are within 12 trunks or one U.S. trunk group module. In Table 2 only the toll switch terminations are considered. If the access terminations are also counted, 96% of the switches are within 4% of the correct number of terminations.

5. SWITCH CAPACITY ANALYSIS

A basic constraint on the capacity of a switch is the maximum number of trunks that can be physically terminated on it. For digital switches, this limit may never be attained due to additional constraints on the processing throughput of the CPU or on other constraints such as memory. Reliable models of these additional constraints are particularly important in the planning of ISDN gateway switch capacity as the handling of international calls is typically more processor and memory intensive than the domestic counterparts. In this section, the basis of the real time and memory models that may be used to determine the capacity of a 4ESS switch are presented. Similar models would be required for other electronic digital switches.

5.1 Real Time Processing Capacity

The 4ESS is a single processor electronic switch that controls call processing by a cyclic polling of the trunks terminating on it. If a request, such as a call setup or takedown, is detected, this basic polling cycle is suspended while the processor handles the task requested. The amount of processor time required to handle the task depends on its characteristics, for example, whether it is an MF or an ISDN CCS7 call setup.

The processor runs continuously in this cycle, the duration of which is determined by the number and type of tasks that require servicing during it. The real time processing capacity of the switch is determined by the maximum duration this cycle can have and still preserve the timing integrity of the switch.

The model for this is based on an equation for the fraction of time the processor is occupied, \( \rho \), given by:

\[
\rho = \sum_{s \in S} \lambda_s \tau_s + \frac{B}{T} + \rho_0, \tag{16}
\]

where

- \( \lambda_s \) is the rate service is requested for tasks of type \( s \)
- \( \tau_s \) is the average time to service a task of type \( s \)
- \( T \) is the duration of the base level cycle
- \( B \) is the time required to poll the trunks without service requests
- \( \rho_0 \) is the fraction of time the processor spends handling overhead tasks not directly related to call processing.

\( S \) is the set of types of tasks the switch can handle that includes various signaling protocols and services. An example of a task could be call setup for a CCITT international call or a CCS7 domestic ISDN call.

The processor runs continuously through the base level scheduling cycle, adjusting its duration, \( T \), so that the processor is continuously occupied: \( \rho = 1 \). With some algebra, this equation may be rewritten as

\[
1 = \lambda_{total} \tau_{A_0} + \frac{B}{T} + \rho_0, \tag{17}
\]
where

\[ \lambda_{Tot} = \sum_{s} \lambda_s, \quad \tau_{Av} = \sum_{s} f_s \tau_s. \] (18)

\( f_s \) is the fraction of the tasks that are of type \( s \) given by:

\[ f_s = \frac{\lambda_s}{\lambda_{Tot}}. \] (19)

The rate at which calls of a given type arrive, \( \lambda_s \), can be obtained by multiplying the number of terminations on the switch of this type times the average number of attempts per unit time for this termination type.

The equation may now be taken as giving the duration of the base cycle, \( T \), as a function of the total attempts, \( \lambda_{Tot} \), depending on the three parameters, \( B, \rho_0 \) and \( \tau_{Av} \). The switch exhausts on real time when the total number of attempts causes \( T \) to exceed the maximum base level cycle duration.

The three parameters are critical to the model. They may be determined by measurements on a laboratory switch and assumed to hold for all switches. Alternatively, the parameters may be determined by regression equations for actual data values of the quantity \( T \) versus \( \lambda_{Tot} \) taken from measurements on switches in the field. These equations are given in Appendix A.

5.2 Memory Capacity

The 4ESS™ has a limit to the amount of memory that may be used in the call setup and takedown process. The amount of memory required may be approximated by a linear model of the form:

\[ M = M_0 + M_T + M_{TG} + M_{CR} + M_F. \] (20)

where

- \( M_0 \) is a fixed amount of memory required for the switch to function
- \( M_T \) is memory allocated according to number of terminations on the switch.
- \( M_{TG} \) is memory allocate according to the number of trunk groups on the switch.
- \( M_{CR} \) is memory allocated according to the number of call registers required.
- \( M_F \) is memory allocated for features supported by the switch.

Expressions for \( M_T, M_{TG} \) and \( M_{CR} \) are of the form:

\[ M_T = \sum_t \mu_t T_t, \]
\[ M_{TG} = \sum_g \gamma_g TG_g, \]
\[ M_{CR} = \sum_s \sigma_s \lambda_s. \] (21)

where

- \( T_t \) are number of terminations of type \( t \)
- \( TG_g \) are number of trunk groups of type \( g \)
- \( \lambda_s \) are number of attempts of type \( s \)
- \( \mu_t, \gamma_g, \) and \( \sigma_s \) are modeling parameters.

The modeling parameters \( \mu_t, \gamma_g, \) and \( \sigma_s \) are determined from an analysis of the software requirements.

REFERENCES


APPENDIX A

Consider a function of \( n \) variables, \((x_1, \ldots, x_n)\), depending on \( m \) parameters, \((\alpha_1, \ldots, \alpha_m)\)

\[
f(x_1, \ldots, x_n; \alpha_1, \ldots, \alpha_m) = 0. \tag{A.1}
\]

The problem is to determine the optimal set of parameters to give a regression fit of this equation to \( N \) data points, \((x_1^i, \ldots, x_n^i)\), \( i = 1, 2, \ldots, N \). The least squares method is to minimize the sum of squares,

\[
S(\alpha_1, \ldots, \alpha_m) = \sum_{i=1}^{N} f^2(x_1^i, \ldots, x_n^i; \alpha_1, \ldots, \alpha_m), \tag{A.2}
\]

by setting the partial derivatives with respect to the parameters to zero:

\[
\frac{\partial S}{\partial \alpha_k}(\alpha_1, \ldots, \alpha_m) = 0, \quad k = 1, \ldots, m. \tag{A.3}
\]

In order for this to be a minimizing solution, the associated Hessian matrix of second partial derivatives must be negative definite which also is a sufficient condition for the system in equation (A.2) to be inverted to give the parameters \((\alpha_1, \ldots, \alpha_m)\).

For the case of interest, the system is linear with the solution

\[
\begin{bmatrix}
\tau_{Av} \\
B \\
\rho_0
\end{bmatrix} = 
\begin{bmatrix}
\sum_i (\lambda_{Tot}^i)^2 \\
\sum_i \lambda_{Tot}^i / T^i \\
\sum_i \lambda_{Tot}^i \\
\sum_i T^i \\
\sum_i T^i \\
\sum_i (T^i)^{-1} \\
\sum_i (T^i)^{-1} N \\
N
\end{bmatrix}^{-1} 
\begin{bmatrix}
\sum_i \lambda_{Tot}^i \\
\sum_i (T^i)^{-1} \\
\sum_i (T^i)^{-1} \\
N
\end{bmatrix}. \tag{A.4}
\]