OPTIMIZATION OF A CORPORATE PHONE SYSTEM VIA COMPUTER CONTROL

A Modeling Case Study

R. E. Buten
M. E. Doherty

Procter & Gamble Co.
Cincinnati, Ohio

ABSTRACT

A large corporate long distance telephone network is studied. A computer control system is proposed, and WATS and toll call records are obtained form the local telephone company. Analysis of this call data revealed some surprises regarding applicability of the standard telephone company engineering guidelines.

Call trace records together with specifications of the existing equipment plus the proposed computer control system are used to develop a trace-driven model of the phone system. This model is validated against actual system behavior. The validation process revealed that callers used toll facilities when WATS capacity was apparently available. This phenomena was attributed to access constraints within CENTREX, line outages, and user education problems.

After validation, the characteristics of the proposed computer control system are investigated and operating strategies are developed. Savings projections are developed and used to justify the acquisition of the system.

The system was subsequently acquired and installed and performance measurements taken. These measurements support the earlier modeling projections.

INTRODUCTION

The business needs of a large corporation today require the rapid movement of large volumes of information. This information comes in many forms from letters and pictures to computer readable data, and most common of all, conversation. The requirement of speed in all of these cases produces a common focus: the telephone.

In response to these needs business finds a large and often confusing assortment of service offerings available from an interconnecting network of phone companies. Among these, several are of interest here:

CENTREX - Switching facilities dedicated to a specific customer, provides direct inward dial as well as dial access other phones and facilities attached to it.

FP - Full Period line; a type of long distance, voice grade leased line which connects two user owned switching facilities.

FX - Foreign exchange line; a type of long distance, voice grade leased line which connects a user's switching facility to a distant telephone company exchange. The effect is to users of the switching facility access to any number locally serviced by that exchange.

Toll - A long distance call which uses the U.S. public dial telephone network. The acronym DDD (Direct Distance Dial) is frequently used interchangeably with the term toll.

WATS - Wide Area Telephone Service; essentially a billing arrangement for bulk buying of toll service. WATS lines are available in five interstate categories or bands which correspond to geographic areas roughly concentric to a given state of origin and proportionally priced according to distance from it. Each successive higher numbered WATS band serves an area which includes that of all lower numbered bands, excluding the state of origin. To service this area there is an additional type of service called intrastate WATS.

To assist a business in evaluating their offerings the local phone company will do a study and make recommendations for leased lines which will generate savings over the exclusive use of toll. Still these questions remain in the manager's mind: "Is this the best I can do?"; "Am I getting the proper return of these lines?"; and "How can I control the growth of these lines?". To these issues add the confusion of new and different service offerings by other common carriers and we have a significant managerial problem.

SYSTEM DESCRIPTION

The Cincinnati portion of Procter & Gamble's phone system is centered about a large "step-by-step" CENTREX switching facility which spans two local phone exchanges. This type of equipment is so named because during dialing a call signal is channeled digit by digit through successive banks of electro-mechanical rotary switches.

In addition to "0" for operator and "9" for outside lines, many other first single digit accesses are taken by other facilities such as conference lines, dial dictation, etc. Because of this and the large number of different long distance facilities, multi-digit access schemes are used.

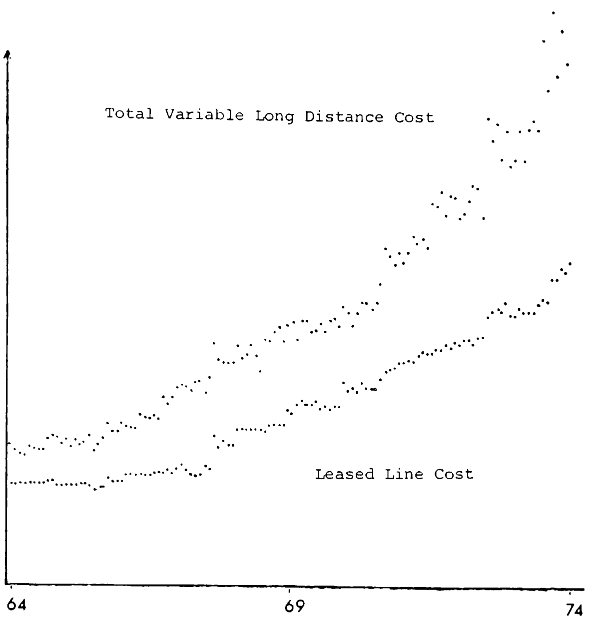
Many private phone systems employ operators to access leased lines. At this point, human intervention can provide a number of services including priority placement of calls, queueing, and record keeping useful for charging out the system cost. The call volume of our system makes this approach impractical and complete dial access is used instead.

Users dial 1-n to access WATS lines of band n. Two digit access codes are also provided for groups of FX lines to Chicago and to New York. A three-digit system provides access to the multitude of FP lines extending to factories and offices scattered across the U.S.

The complete dial access to the system, while providing a measure of convenience to the phone users, leaves the system manager without vital usage data with which to cross-charge other departments. Such charges provide a measure of justification to management for the size and cost of the system.

Efforts were made to remedy this situation. Operators queried callers for a department code on an intercept basis. This sample data was used to allocate system charges proportionately to using departments. This procedure had two problems. The accuracy of the data was questionable since some callers habitually avoided the intercept by placing toll calls during the surveys. Also, the number of operators just for sampling had become a cost item hard to justify. These efforts were discontinued in 1971, and leased line costs were charged out via a flat rate applied to each phone.

The cost of providing long distance phone service has grown at a truly alarming rate (see Figure 1), having quadrupled in the last 10 years.



GROWTH IN TELEPHONE EXPENSE
Figure 1

At the same time as the sampling mechanism was removed, we experienced additional increases in growth. The concern about the alarming telephone cost growth rate, together with the lack of accountability to these cost, convinced the management to form a study group.

BACKGROUND INVESTIGATION AND PROBLEM DEFINITION

Available Data

Two sources of data were found to be available from the local telephone company. One was our monthly toll bill, which because of volume, arrives on magnetic tape. This is processed by a telephone accounting system which provides a detailed listing of long distance charges to each department. Department Managers can then check the phone usage of their area for amount and kind of usage, as well as for accuracy of the bill. This is sometimes necessary since our phone equipment is ONI (Operator Number Identification), as opposed to the new ANI (Automatic Number Identification).

We found that another magnetic tape, containing a record of each WATS call, could also be purchased form the local telephone company. Each call record contained time of origin, duration, destination number, and facility used. Because no operator breaks into a WATS call asking "Number please?", the tape lacked information necessary to cross-charge.

Computer Control System

The next option investigated by the study team was the use of a computer based data capture system. This system would interface between the CENTREX and the long distance facilities. Users would dial additional identifying digits in placing their calls. The computer system would write a record for each call including time, duration, facility used, and identification, which could be used to charge back calls to the originating department. In addition, the timely availability of this information would be of value in optimizing the network facilities. With the exception of additional digits required, this type of system is essentially passive in the telephone network.

Some vendors of this kind of equipment have extended the role of the computer in such a system to perform various active functions to increase line utilization or aid users in placing calls. Some of these features are:

Auto Call Placement - the system collects and stores the digits as the caller dials them. When the number is completely specified, the system searches for a line. If available, the system then "dials" the number with touch tone pulses in about 2 seconds. The net effect of this is to shorten the duration of all calls by the time needed by a user to dial 10 digits.

Optimal Facility Selection - the computer system will select the least cost facility to place a call. If all lines of this type are busy, the system has the ability to select alternate facilites to satisfy the request. In some systems, the system manager may specify which and in what order additional facilities are to be tried.

Call Queuing - while an operator administered system could manually queue calls for lines, a dial access system such as ours and the Public Phone System have discrete attempts. A computer control system can queue calls for lines or, like an operator system, "notify" users when lines become free.

Priority Scheduling - an intelligent phone control can also allocate facilities on a discipline other than a first come first serve basis. It can execute priority queuing for calls deemed to be more important. The study team was faced with the task of evaluating what type system, if any, should be installed. And, because of the inavailability of data on load other than WATS and toll, as well as the sheer size of the system, it was decided to limit the investigation to WATS traffic alone.

Engineering Methodologies

Telephone Company. The local telephone company will assist in designing a leased line network much in the same way as they design the public system. Their engineering guidelines[5] use tables which are based on the Erlang Loss Equation [2]. The Erlang Model assumes poisson distribution of call arrivals with exponential holding times.

Checking this assumption with statiscal measures calculated from the purchased WATS tape, we find exponential holding times a poor fit. The coefficient of variation calculated for call lengths was greater than three. This is most likely do to the presence of data calls on the network. In addition, the arrival rate of calls changes dramatically as a function of time of day (see Figure 2).

Even if the call traffic met the assumptions of the Erlang Model, the telephone company would use their tables to engineer the circuits to a P-grade of service. This "P" refers to the number of losses to the system in the next higher decimal order of magnitude of calls. For example, P05 means only 5 calls in 100 attempts are turned away without service; P001 means 1 in 1000. Such callers turned away may represent lost revenue to a phone company.

Unlike a telephone company, we have no interest in setting a specific service level on our leased line system. The objective is to provide adequate service to the corporation at the lowest total cost. And since any caller can always place a toll call, adequate service is essentially guaranteed. The charges from these toll calls form the variable portion of the total cost equation to be minimized.

Analytic Efforts. While there has been much work done in analyzing queuing models, there has been little analytic progress against the minimization of total phone cost. One effort [4] used an LP model to obtain upper and lower bounds on the optimal WATS configuration. This method made use of an ad hoc "probability of a toll call" as a function of service level of the WATS system. The actual optimization equation themselves were non-linear and a solution was obtained via Monte Carlo simulation. This is where we started.

Simulation. As stated above, the call traffic does not meet the criterion of exponential holding time. Moreover, to use poisson arrivals one would need to alter the mean of the arrival time distribution for each WATS band, for many time period per day. In addition, there are known interdependencies between call lengths of data calls and time of day.

While none of these complications represent insurmountable obstacles to Monte Carlo generation of the load, they were for our purposes superfluous. That is, the purpose of our investigation was to determine the effect of a system change - given the current system load. To accomplish this with minimal difficulty, we turned to a technique more commonly found in the modeling of computer systems, and for the same reason. This is trace-driven modeling. [1, 3]

MODEL

Description

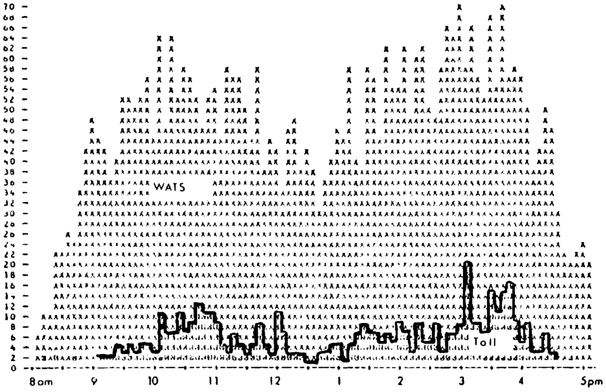
The model is approximately 400 statements written in SIM PL/1, an IBM PL/1 based high level discrete event simulation language. This language was chosen because of the complete access to the system I/O facilities and optimization of the model's computer program not available with GPSS or SIMSCRIPT. The mdoel runs in 180K bytes of core and requires about 40 sec. of IBM 370/168 CPU time to simulate one day's call activity.

A preprocessor module merges the purchased WATS tape with our monthly toll bill which also arrives on magnetic tape. The preprocessor also reformats the data, and makes selections based on date and time and deletes irrelevant data from the toll tape. The result of this operation is a call trace file ready for input to the model.

The model reads a call trace file, a description of the phone facilities, which features, if any, of the computer control system are to be used, and what reporting is desired.

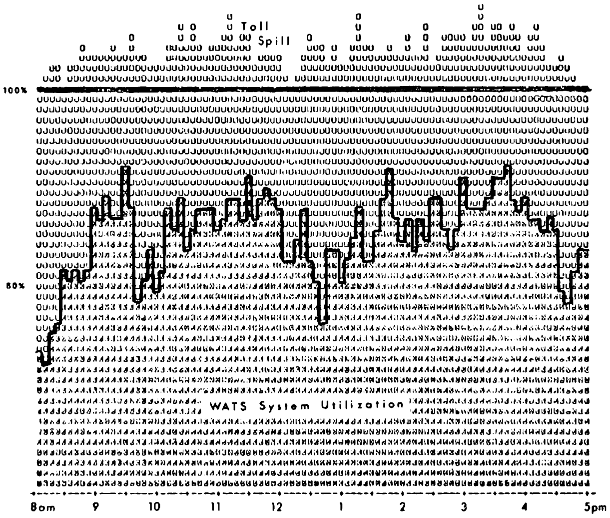
The model attempts to "place" each call from the call trace file in much the same way the phone system does, with one notable exception. The model treats a group of lines on a rotary selection basis regardless of size, ignoring the limitations of 10-position selector switches. In addition, the model artificially segregates toll calls according to which WATS band it had been destined. By collecting call counts and utilizations in this way, we see a plot of the traffic concurrency required by our system by geographic area.

One of the model outputs is a profile of the day's phone activities in terms of arrivals is shown in Figure 2.



CALL ARRIVALS VS. TIME OF DAY
FOR A REPRESENTATIVE BAND
Figure 2

Another output, shown in Figure 3, snows a load profile.

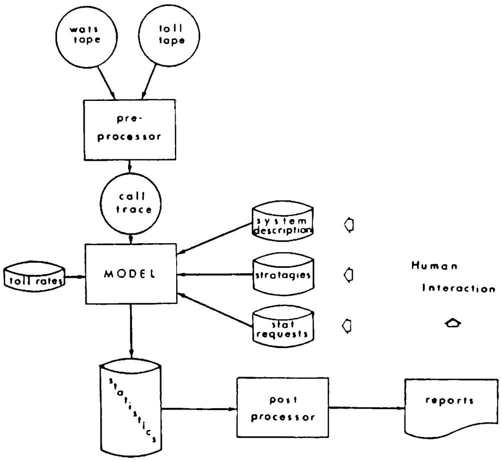


WATS AND TOLL LOAD FOR A SINGLE
DAY FOR REPRESENTATIVE BAND
Figure 3

A post processor module reads a file written by the model during execution and produces the desired reports. The complete analysis system is shown in Figure 4.

Validation

The most important step in any modeling effort is validation. With a trace-driven approach the model execution itself is deterministic rather than stochastic. Having captured the load data itself, the validation process reduces to comparing the action of the model to the record of what actually happened.



ANALYSIS SYSTEM OVERVIEW
Figure 4

One of the benefits of the deterministic nature of the model was obtained from the following report.

WATS LINE NUMBER	PRIME TIME UTILIZATION
1	99%
2	95%
3	91%
4	81%
5	74%
6	65%
7	58%
8	39%
9	0%

TOLL LINE NUMBER	PRIME TIME UTILIZATION
1	52%
2	38%
3	25%
4	15%
5	10%
6	4%
7	3%

For this particular WATS band our records showed 9 lines. Yet the model, using actual call traffic, found a WATS concurrency requirement of only 8 lines coincident with enough demand to generate a non-trivial toll spillover. Upon searching for causes, we discovered that installation activity had taken place in the same time frame. Our suspicion is that the lines were not working properly at this time. This suspicion was further strengthened by subsequent model runs for a time period several days later which showed the appearance of these lines.

This presented us with another source of tangible, but hard to estimate, savings - line outages. A computer control system could assist a user in getting full serviability from the lines in place.

Once the model was shown to correctly duplicate the performance of the current system, the features of the proposed control system were tested. Without an actual system against which to compare, the validation of these features was limited to consistency and trace checking.

MODELING THE COMPUTER CONTROL SYSTEM

The next investigation area concerns modeling the active features of the computer control system in order to estimate potential savings. With the trace-driven model and the actual load data at hand, there are two major areas to investigate: (1) What active features should be used and how can they be managed to best advantage?, and (2) How many lines of each type should be connected to the control system? The first question concerns such strategical decisions as queue discipline, queue time limit, and line selection schemes; the second question asks for the optimal "configuration". Of course these questions are not independent, and it should be emphasized that the criterion of optimality used to judge the various strategies and configurations was minimum total cost. The cost associated with a particular simulation run is composed of two parts: the leased line cost which can be computed with perfect accuracy, and the toll cost which is estimated from the lengths of the calls and the average toll rates for calls into each area code called, obtained from actual toll bills.

The "total cost" criterion for optimality was motivated by the consideration that there was no a priori level of grade of service which had to be met. Instead, the following simple argument shows how an optimal grade of service is attained. When the configuration is too tight so that the grade of service is not acceptable to the user, he will make toll calls. The toll cost will rise to the point where additional lines are justified economically. At this point the lines are added and the grade of service improves. In essence, the user is a feedback element that signals the grade of service through toll usage.

Strategies

Queue Time. Utilization of lines can be increased by increasing the maximum time for which a caller will be held in queue. However, when the time limit is set too high the access to the system can become clogged, that is, serviceable calls may be refused access while unserviceable calls occupy system access ports. Also, a great utilization advantage is obtained with a modest amount of queuing (about 30 to 40 seconds) and extensions beyond this limit have greatly diminished additional benefit. A human engineering factor taken into consideration was the amount of time a caller would have to spend if he were to hang up and dial 9 + 1 + area code + number to place his toll call outside the system. To dial 12 digits would require a minimum of about 25 seconds. Hence, if the caller can be guaranteed that his call could be placed (even if a toll line is required) by the computer control system, then he should not object to waiting at least as long as he would need to re-dial the number he has already entered into the system. As a result of our modeling studies a 40 second queue time maximum was chosen. With this value in the model the average time spent in queue by queued calls was about 25 seconds.

Queue Discipline. Alternatives considered here included ranking the queue by priorities assigned to each caller and ranking the queue by priority based on the destination (e.g., WATS band) of the call. The former was not adopted as a corporate policy, and the latter did not reveal any measurable operational advantage. Hence, a FIFO (First-In, First-Out) queue discipline was adopted.

Line Selection Schemes. Since any interstate call can be made over a WATS line for any band encompassing the destination, a typical call has more than one type of WATS line that can carry it. For example, a call to a location in band 2 could be placed on a band 2 line, a band 3 line, a band 4 line, or a band 5 line. A line selection scheme is a specification of which bands should be searched for a free line, in what order they should be searched, and which bands should be searched while the call is in queue. Several schemes can be defined and the appropriate one selected based on the area code to which the call is destined. For example, we might specify that all

Band 2, Band 3, Queue, Band 4, DDD.

This is interpreted as follows. Try for a free band 2 line; if all are busy then try for a band 3 line; if all are busy then go into queue. While in queue constantly search for a band 2 or a band 3 line to become free. If the queue time limit runs out, then try for a band 4 line; if none are available try for a DDD line. Note that this scheme does not permit the call to use a band 5 line. Also, note that if a band 3 line frees before a band 2 line while the call is in queue, the call will take the band 3 line.

A natural organization of schemes is to have six distinct schemes: one for each band where an interstate call may be placed and one for intrastate calls (for which tariffing requires a unique and exclusive type of WATS line). However, the WATS tariff structure, which is different for each state, shows that for Ohio a band 4 line is about only 2% more expensive that a band 3 line. Hence, no band 3 lines were used; instead all calls into a band 3 or band 4 were grouped together and treated by a single scheme. We call a group of schemes, one for each band a strategy for call placement.

In order to design strategies for testing in the model, some "principles" of management were developed. One such principle defines the "cost" of a WATS call. By this we mean the cost to telecommunications operations, not a billing amount. Several confusing and contradictory definitions were presented. Among them were the following: the real cost is zero because the flat monthly amount has already been paid and cannot be recouped; the real cost is a fraction of the flat rate determine by the line utilization and the length of the call. Both of these and other definitions were rejected and instead the following definition was adopted: the cost of a WATS call is a random variable whose expectation value is equal to the product of (1) the expected cost of a toll call, and (2) the differential probability that a toll call will be made, due to the WATS line being busy. For example, suppose we want to make a 10 minute call over a WATS line. At the time of day we plan to make the call let the probability that another call will arrive and have to go toll during this 10 minutes be $P=0.15$. Since we will be using a WATS line this probability is increased to $P'=0.25$, for example. Then the differential probability of a toll call occurrence is $P'-P=0.10$. If the expected cost of a toll call is \$3.20, then the cost of our WATS call is \$0.32. Consequently, a WATS call made at 1:00 a.m. is essentially "free" because the chance that it will cause another call to go toll is practically zero. However, a call at 2:30 p.m. on Friday afternoon has a significant cost because of the high competition for lines at that time. Our definition simply states that the value of the resource depends on the demand for it.

Another principle which is strongly intuitive is that a call should not be placed on a more expensive facility if a cheaper facility is available. This indicates that a scheme should search bands in order when trying to find a line. Since DDD is the most expensive facility to use it should be a last resort in any scheme. Corporate policy does not restrict anyone from making a toll call. So, for any call, a scheme should examine all possible WATS bands in order and then DDD, the major question being at which point queuing should be done. Further, queuing should be done before searching for a DDD line since there is no desire to force traffic on these lines.

We have named a scheme which, before queuing, examines k bands beyond the primary band required to place the call a "hunt-k" scheme. For band 5, a hunt-0 scheme is the only possibility, but for the other bands there are multiple possibilities. A strategy (set of schemes) is called a hunt-k strategy if each scheme is a hunt-p scheme, where p is the minimum of k and the number of existing bands for hunting up. Examples of a hunt-2 strategy and a hunt-0 strategy are shown in Figure 5, where "B-j" abbreviates "Band j". Recall that no band 3 lines are used at all.

HUNT-2 STRATEGY

(Band 1): B-1, B-2, B-4, QUEUE, B-5, DDD
(Band 2): B-2, B-4, B-5, QUEUE, DDD
(Band 3-4): B-4, B-5, QUEUE, DDD
(Band 5): B-5, QUEUE, DDD

HUNT-0 STRATEGY

(Band 1): B-1, QUEUE, B-2, B-4, B-5, DDD
(Band 2): B-2, QUEUE, B-4, B-5, DDD
(Band 3-4): B-4, QUEUE, B-5, DDD
(Band 5): B-5, QUEUE, DDD

THE DETAIL SCHEMES FOR TWO STRATEGIES
Figure 5

Considerable simulation experiments were performed with various hunting strategies. These experiments indicated that a hunt-0 strategy was best.

Some analytic approximations, however, indicaed that a hunt-1 strategy should outperform a hunt-0 strategy when the system was not very busy. It is exactly at such times when toll usage is almost zero anyway, so that there is no significant impact on costs depending on whatever hunt-k strategy is used.

The hunt-0 strategy was adopted as the best alternative, and there is a strong intuitive explanation of its merit: calls should be "route advanced" to higher bands when the cost is not too high. By our first principle this means that assigning the call to the higher band will not likely force out a toll call. This means that the line in the higher band is relatively "free". The hunt-0 strategy ensures that this is the case because it does not queue on the higher bands, but only searches them "on the fly" after the queue limit. Hence, if the higher bands are busy there is a proportionately reduced probability that a line will be stolen by a route-advancing call.

The hunt-0 strategy tries to force calls to be placed in their primary bands because of the queuing on those bands. In Figure 6 we see the percent of calls which have route-advanced to higher bands. Too much route-advancing, which was the result of more liberal hunt-k strategies leads to a cascading effect which causes calls in higher bands to be forced out to toll at more expensive rates.

Line Configuration

During the process of testing and evaluating hunting strategies, it became necessary to evaluate the cost of the strategy under optimal configuration of the leased lines. Within the model the lines were grouped by type, such as band 2 WATS lines, and a line of that type was selected by a call in a rotary selection method. This means that a search for a free line always starts with the first line and proceeds in

DESTINATION BAND	PERCENTAGES BAND USED 1	2	3	4	5	DDD
1	91	4	—	2	2	0
2	—	87	—	7	5	1
3	—	—	—	95	4	1
4	—	—	—	96	3	1
5	—	—	—	—	98	2

CALL DISTRIBUTION PATTERN
RESULTING FROM HUNT-0 STRATEGY
WITH 40 SECOND QUEUE
Figure 6

order. Whether or not this is the selection used in the real world or not, the capacity of the group of lines is unchanged. However, it does permit the modeler to see the effect of removing one or more lines because it "packs" the load onto the lines in order. Consequently, the traffic on the last line would be overflow if one line were removed.

In determining the optimal configuration the following scheme was used. For each line in the model, statistics were kept which gave the monthly projections of what would be monthly toll cost of the calls carried on that line. In the model this was done for both WATS lines and toll lines, which were artificially segregated by WATS band. Hence, the modeler could compare the toll "carried" by the last WATS line and see if it justified its flat monthly cost. Similarly the toll cost carried by the first toll line could be tested to see whether it warranted being converted to a WATS line. By this scheme WATS lines could be added or deleted. This was not a perfectly clear-cut case of result prediction because of the route advancing feature which allows interaction between bands with regard to load and resources. Hence, more simulation runs were needed to validate the economic benefits of suggested configuration changes.

When a satisfactory WATS configuration was reached, an additional model run was made in which there was only one type of toll line, as opposed to the artificial segregation mentioned above. The lines were a common pool for all calls to access, and hence the required number of lines could be determined. Essentially a grade-of-service measure was used here because of the expense involved in connections made to the computer control system, even though toll lines themselves are inexpensive.

Using the above techniques, we decided to use a hunt-0 strategy for line selection, a queue time of 40 seconds, and how many lines of each type to be connected to the computer control systems.

Savings Estimates

In view of the changing load, absolute savings amounts were less important than percentage of system expense saved. The only measure available which had sufficient historical data was total variable long distance cost. This consists of the combined costs of all leased lines (WATS, FX, and FP) and all toll costs, and is abbreviated TVLDC.

The savings forecasted by the study effort described above are shown below.

ITEM	SAVINGS (% of TVLDC)
Controlled access to system	3.20
Configuration Optimization	6.65
Call routing strategies	2.45
Auto call placement	1.25
Queuing	.75
TOTAL	14.36

IMPLEMENTATION

The savings estimates and technical feasibility studies done convinced management to approve the acquisition and implementation of the computer control system as a measurable benefit savings project.

Implementation of the computer control system was done in two stages, the first of which cut-over about 40% of the telephone calling load. As of this writing, the system has been carrying the entire planned load for about one month. However, we recognize that there are several factors which tend to support a transient type of behavior. Among these are: user education, residual system "bugs", and the impact on users behavior due to billing for WATS calls. Furthermore, although the routing strategies used were those designed in the modeling study, the line configuration was deliberately padded to provide the users with very good service upon initial exposure to the system and to provide contingency through the early stages of system shakedown and problem solving.

Recycling

One particularly interesting type of behavior was discovered during the implementation process. This behavior, which we term recycling, consists in a refusal by a user to let his call go toll. Physically this is triggered by a computer control system feature which provides the caller with an audible tone when his call is about to go over a toll line (after all hunting and queuing). If the user wants to permit the toll call, he must dial the digit '1', after which the call is immediately made. If the user does not respond to the tone within five seconds (or if he simply hangs up), he is disconnected from the system. We term this a recycle call. If he never (within the same day) calls that number back again, we say that he "recycled zero times"; and if he repeats the attempt k times before completing or abandoning his call, we say he "recycled k times".

Bell engineering guidelines assume that all calls denied service are "cleared" from the system and do not reenter. Our data clearly shows that our load does not conform to this assumption.

Analysis of Recycling

Although recycling was known to exist before the computer control system was installed, there was no way to measure it because telephone company WATS and toll tapes contained data for completed calls only, and we could not tell how many times a user made unsuccessful dial accesses to WATS before dialing '9' and making a toll call. However, since the computer control system keeps detailed records on all attempted calls we can quantatively analyze this behavior.

The control system activity tape was sorted by time-of-day within number-called within caller-authorization-number within day. A computer program was written to analyze this data and Figure 7 shows typical results.

Exact Number of Recycles	Percent of All Recycle Calls	Final Disposition of Call		
		%WATS	%ABANDONED	%TOLL
0	23.9	0.	100.0	0.
1	60.2	90.4	6.8	2.8
2	11.6	75.7	16.2	8.1
3	2.7	70.6	23.5	5.9
4	.9	83.3	16.7	0.
5	.2	100.0	0.	0.
6	.2	100.0	0.	0.
7	.3	100.0	0.	0.
8	.2	100.0	0.	0.

Percent of Eligible Calls Which Recycled: 51%

ACTUAL RECYCLE STATISTICS
Figure 7

The percent of eligible calls (i.e., calls offered a toll line) which recycle has a dramatic effect on the method used for line configuration. With zero recycling, a WATS line would be justified if it carried 100% or more of its cost, in estimated toll charges for calls over it. With recycling present, this percentage threshold is significantly higher, as can be seen by the following argument. Suppose the last line in band 5 carried 120% of its cost. If this line were removed we cannot expect to incur toll charges equivalent to the charges carried by this line since some percentage of those calls will refuse to go toll and will try to place their call again later. If the recycling is high enough, we may incur only 90% of the carried traffic as toll charges. In this case, removing the line would actually yield a net savings. A similar argument applies to deciding to add a WATS line. Suppose the first toll line carried 130% of the cost of a WATS line. Depending on the degree of recycling, it may not be wise to add a WATS line because such an added line will not carry all the traffic that appeared on the toll line. This is so because it will absorb recycling calls which went on existing WATS lines but will now be trying to seize it before recycling. Figure 8 shows examples of these cases.

WATS Line Number	% of WATS Line Cost Carried	Recycle Percent	
		0.	50.
.	.	.	.
.	.	.	.
.	.	.	.
n-2	130	justified	justified
n-1	110	justified	unjustified
n	95	unjustified	unjustified

Toll Line Number	% of WATS Line Cost Carried	Recycle Percent	
		0.	50.
1	125	add WATS line	add WATS line
2	110	add WATS line	(no change)
3	95	(no change)	(no change)
.	.	.	.
.	.	.	.

EFFECTS OF RECYCLING
ON LINE JUSTIFICATION
Figure 8

As a tool in helping the modeler judge whether to remove WATS lines when recycling is present, the following mathematical model was developed and implemented through an interactive computer program.

Let the following parameters be defined for the particular band of interest:

B = probability that the system is busy (i.e., that a call won't get a WATS line)

Nw = number of completed WATS calls

Nt = Number of completed toll calls

N = Nw + Nt

Nr = Number of recycle calls

CLL = Number of calls carried on the last WATS line (i.e., the line being tested for removal)

Then the probability that a call goes WATS is:

$$Pw = Nw / N.$$

The fraction of calls offered a toll line which recycled is given by:

$$R = Nr / (Nr + Nt).$$

Assuming that it is unlikely that a call will have to recycle many times, we have:

$$Pw = (1 - B)\ (1 + (B*R) + (B*R)^2 + \ldots)$$

$$\text{or } Pw = (1 - B) / (1 - B*R).$$

Solving for B,

$$B = (1 - Pw) / (1 - Pw * R).$$

Now, if the last WATS line is removed, these parameters have revised values given by the following relations:

$$B' = (Nr + Nt + CLL) / (Nr + N).$$

$$Pw' = (1 - B') / (1 - B' * R).$$

Hence, the revised number of tolls calls is

$$Nt' = N * (1 - Pw').$$

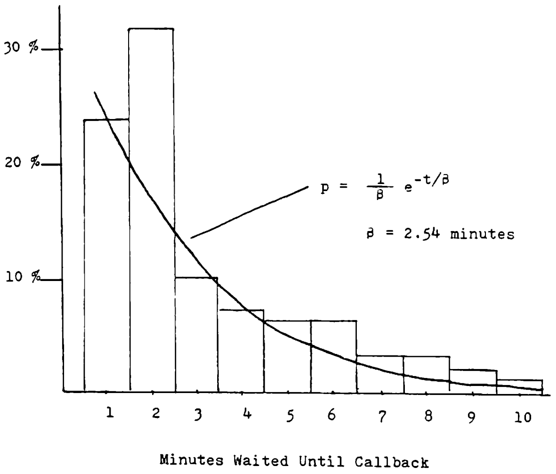
Overall we can expect a net increase in the number of toll calls of

$$\Delta Nt = N * (Pw - Pw').$$

Based on the average cost of a toll call, we can then judge whether the removal of a WATS line is economically justified.

Although this simplified approach has on the average been only about 15-20% accurate, and is inapplicable to large scale changes (several lines at a time), it has proved valuable as an aid to the modeler.

The actual recycling behavior has been implemented in the model itself, and two parameters are used to specify the behavior. They are the recycle probability R, and the mean of an exponential distribution used to approximate the average time that a recycling call waits before calling back. This distribution is determined from the type of data shown in Figure 9, which resulted form a computer program analysis of actual call log records.



DISTRIBUTION OF TIMES
BETWEEN RECYCLES
Figure 9

REOPTIMIZATION

An ongoing procedure is underway which measures system performance and reoptimizes system parameters by use of the model. The first step in this process is the running of programs which read the log tape from the computer control system and print various performance statistics reports. Based on the information in these reports, a telecommunications analyst decides if a configuration adjustment warrants investigation. If so, a model run is made with the suggested modification to see the expected effects. The model run can have such up-to-date statistics as the percent of recycling input since it is computer in the performance reports. Since the control system dumps activity records to the log tape daily, very fresh information can be used for a reoptimization.

RESULTS

The reoptimization process is currently taking place with the philosophy of gradually trimming the system configuration to an approximately optimal set of values. Hence, the total potential savings have not yet been realized. The first reoptimization study has been done and the model projections are shown below.

Since the feasibility study was done, there have been significant increases in rates in the AT&T tariff. Taking these unforeseen adjustments into account and computing the cost per minute per WATS and toll calls, we arrive at the following table:

	COST/MIN INDEX
(Jan.) Existing Base	1.00
(Jan.) Original Model Projection	0.86
(Nov.) Current Actual*	0.93
(Dec.) Reoptimization Model* Projection	0.83

(*) -- Rate Adjusted

With a 40 sec. maximum queue time and a hunt-0 strategy we have found the following statistics for judging the model's predictive accuracy:

	Mean Queue Time	% Queued	% Calls Toll
(May) Model Projection	24.9 sec.	21.8 %	2.3 %
(Nov.) Actual	24.4 sec.	22.4 %	0.9 %
(Dec.) Reoptimization	25.5 sec.	40.3 %	2.6 %

The November actual percentage of calls that went toll is low because the configuration is padded with extra lines.

The November actual percentage of calls that went toll is low because the configuration is padded with extra lines. The reoptimization model projects a much large percent of calls queued due to the discovered recycle phenomenon.

FUTURE

Extensive plans are being made for further development of the use of the computer control system. These include the installation of a second, identical system to be used with the current one. This will provide capacity for projected growth and for the management of FX and FP lines, which are not currently on the system. Spare capacity on the existing system is under 10%.

Although a second system would be installed in the same location as the first one, there are challenging networking problems to be analyzed due to the highly flexible methods possible for interconnecting the two systems. Further model enhancements to analyze these problems have recently been implemented and are being used to design and evaluate alternatives for this system.

ACKNOWLEDGEMENT

The authors wish to express appreciation to Gail Vories who prepared this manuscript.

REFERENCES

1. Cheng, P. S., "Trace-Driven System Modeling," IBM Systems Journal (8,4), 1969, pp. 280-289.
2. Cox, D. R. and Smith, W. L.,"Queues," Methuens Monographs on Applied Probability and Statistics, London, 1961.
3. Schwetman, H.D., "Simulation of Computer Systems Using Automatically Generated Load Descriptions," Proceeding Winter Simulation Conference, 1974, pp. 699-704.
4. Klitz, J. K. and Mecklenburg, A. W., "A Method of Optimum WATS Line Selection," Proceedings of American Institute of Industrial Engineers Conference, 1970, pp. 285-293.
5. Traffic Engineering Practices, Division "G," Section 5-C, American Telephone & Telegraph Company, 1963.