Decision support systems: a practical application—Branch office structure

by JOHN R. WETMILLER
Digital Equipment Corporation
Maynard, Massachusetts

INTRODUCTION

As the decade of the 1980s approaches, it is generally agreed that computer companies will stress not technology, but rather customer service functions as a means to differentiate themselves from one another. Consequently, the need to understand the customer's service requirements and to plan for them is tantamount.

Simply stated, the purpose of a computer service organization is to provide maintenance services to customers with computer equipment. The primary objective of the organization is to minimize the down-time of that customer equipment and thereby minimize customer inconvenience. At the same time the service organization seeks to operate as cost effectively as possible which, of course, minimizes the service cost to the customer. To provide the maintenance services required, computer service organizations will generally establish a branch office to supply the needs of customers within a certain geographical region. In order to effectively structure that office three fundamental questions need to be answered:

1.) What types of service engineers should the office have (i.e., should the engineers be generalists, specialists, or some combination);
2.) How large should the branch office be (i.e., how many service engineers are required); and
3.) How should service requests or calls of different types be scheduled and which engineer types should be assigned (e.g., first-come-first-served, shortest-expected-service-time; generalist, specialist).

For our company the task of addressing these questions was given by our upper level management to our internal management science consulting group. In the discussions that follow in this paper, I will indicate our findings specifically with regard to question 1. Although some of the numerical results of our studies must remain proprietary, I will indicate the overall conclusions reached and note the pilot test plans that have resulted from our work. I will comment briefly about questions 2 and 3 throughout my discourse on question 1. The complete answers to those questions are still being determined; perhaps, in the future, they can be incorporated into papers similar to this one.

SUMMARY RESULTS

At the outset it was generally felt amongst the members of the analysis team that the proper use of engineer specialization could generate significant branch savings over an all engineer generalist environment. To define terms, an engineer specialist is an individual who can perform certain repair tasks in less time than the average engineer generalist who is able to perform any repair task. Our findings confirmed that, indeed, engineer specialization can result in considerable savings to a branch office. These savings may be expressed as lower response time or increased call handling (or call rate) capability, with no increase in personnel nor in cost. If the proper conditions are present at the branch, specialization can even possibly result in reduced personnel and lower costs.

Figure 1 shows two plots of response time versus call rate. The one to the left is for an all generalist office (no specialization) while the one to the right is for an office with some specialization. Note that the response time/call rate curve for the specialist office lies below and to the right of the non-specialist office.

The implications of this shift are summarized in Table I and discussed below.

<table>
<thead>
<tr>
<th>Response Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call Rate</td>
</tr>
</tbody>
</table>

![Figure 1](From the collection of the Computer History Museum (www.computerhistory.org))
Switching from an all generalist office at Point 1 on the generalist curve to an office with specialists at Point 2 on the specialist curve reduces response time without affecting the call arrival rate and without increasing the total number of engineers. Since specialists are used, overall mean-time-to-repair (MTTR) decreases and hence idle time increases. This last point could be considered as a loss to the office in that there are engineers available to take calls but there are no calls to take. However, this idle time increase over offices without specialization may be minimal and therefore insignificant when compared to the other benefits of specialized offices. Note also that engineer salaries are generally lower for specialist engineers and training requirements are also reduced.

Moving from an all generalist office at Point 1 on the generalist curve to an office with specialists at Point 3 on the specialist curve increases call rate capacity without increasing response time and without increasing the number of engineers in the branch office. Again, overall office MTTR decreases but since more calls are being taken engineer idle time decreases. Also, as was said for Point 2, total salary and training requirements are reduced over the all generalist office with its operating characteristics described by Point 1.

**METHODOLOGY**

The results of the study summarized above were determined from a computer simulation model written in GPSS (General Purpose Simulation System) to represent a typical branch environment. As stated earlier, minimizing customer system downtime becomes the goal of each and every branch office. In the simulation model system, downtime is divided into its three component parts: (1) waiting time, (2) travel time, and (3) repair time. Waiting time is that period of time from when the service request is received by the service organization until an engineer is dispatched to the call. Travel time is, of course, the time it takes the engineer assigned to reach the customer site. Taken together waiting time plus travel time is referred to as response time. Repair time is the time the engineer requires to correct the malfunction once on site. Waiting time is a function of engineer availability (which is, of course, dependent on many factors). Travel time is dependent on the geographical distribution of customers, and repair time is a function of component technology and engineer skill levels. The simulation model considers travel times and repair times as distributed system inputs and evaluates the variation in waiting time (also distributed) as a function of all system input variables.

The easiest way to describe the nature of the other system inputs is to consider Figures 2 and 3 below (Note: the repair time and priority rankings used in the Figures are strictly arbitrary).

**Figure 2** depicts how MTTRs for the various engineer groups can be specified to the model for varying types of arriving calls. This allows the user to create specialist groups if desired and to differentiate repair times for different types of calls (e.g., corrective maintenance, preventive maintenance, and installation calls).

**Figure 3** details how the priorities with which engineers are assigned to certain types of calls are inputed into the simulation model. This allows the user to determine the impact on the office operating characteristics of varying the engineer/call type priority assignments.

It is important to note at this point that fundamental textbook queuing theory analysis will not readily permit the use of such an extensive collection of input parameters. Basic queuing theory will consider only the total number of engineers, an average call rate, and an average service time as input parameters. The ability to specify engineer types, to use distributed call rate and service times, and to prioritize engineer assignments by call type is not available in the standard queuing equations.
From the GPSS simulation model it was possible to obtain the following output data:

1. Average waiting times and waiting time distributions by call type,
2. Average repair times and repair time distributions by engineer group,
3. Idle time by engineer group, and
4. The number of calls of each type taken by each engineer group.

Basic textbook queuing theory would only have allowed the determination of a composite waiting time and a composite engineer idle time.

In the process of performing the study three different types of specialists were considered—true, limited, and senior specialists. Typical MTTRs in hours (arbitrarily chosen here) for a generalist and for each of these specialist types are noted in Table II.

As noted a true specialist could repair his/her speciality more quickly than a generalist, but the true specialist could only repair a very small subset of devices. A true specialist would be paid considerably less than a generalist and would receive less training as well. A limited specialist could repair his/her specialty devices more rapidly than a generalist, but the MTTR on the non-specialty devices would be greater. Salary and training levels would be between the true specialist and the generalist. The senior specialist essentially resembles a generalist in repair times except that his/her repair time on specialty devices would be lower. The salary of a senior specialist would exceed that of a generalist, but training requirements would be equivalent.

**ANALYSIS**

When investigating the impact of specialization on offices of a given size (i.e., offices with a fixed number of engineers), a number of interesting results with respect to response time and engineer idle time were noted when the number and type of specialists in the office was varied while the percentage of specialist type calls remained the same. For example,
Figure 4 expresses the general response time and idle time phenomena for an office with an increasing number of true specialists. As the percentage of specialists in the office approaches the percentage of specialist type calls, response time decreases since all the specialists are busy. However, when the percentage of specialists exceeds the percentage of specialist type calls, some specialists must remain idle since they cannot take calls of other types. Hence, the office actually has fewer effective engineers, and both the response time and engineer idle time increase.

Figure 5 is a similar representation for limited specialists. Note that the response time curve is similar to that of true specialists. This is true because, when there are more specialists than specialist calls for them to handle, they accept non-specialist calls which require an MTTR greater than that for a generalist. Idle time consequently decreases under these circumstances.

Finally, Figure 6 represents the response time and idle time characteristics for an office utilizing varying numbers of senior specialists. As was true for both the true and limited specialist cases, response time initially decreases and idle time increases as the percentage of specialists approaches the percentage of specialist type calls. However, after that equality point, response time and idle time remain the same since there is no degradation of MTTR below the all generalist MTTR level when a senior specialist is working on non-specialty type calls.

The impact of these idle time and response time phenomena on the response time/call rate characteristic operating curve for a branch office is significant. This can be readily seen in Figure 7 which can be considered representative of a branch with a given total number of engineers, a variable number of true or limited specialists, and a fixed percentage of specialist type calls.

The operating curve for an all generalist office (G generalists and O specialists) initially moves to the right as the number of generalists is decreased by one (G-1) and the number of true or limited specialists is increased by one (1). That movement to the right will continue as long as the percentage of specialists in the office is less than the percentage of specialist type calls. Once it becomes greater the characteristic curve will move to the left approaching and eventually passing the all generalist curve (note the G-S,S curve to the left of the all generalist curve in Figure 7). It should be noted here that for the senior specialist case the characteristic curve will continue to move to the right regardless of the number of generalists replaced by senior specialists since any senior specialist MTTR is never higher than that of a generalist.

At this juncture it would appear that several rules-of-thumb can be stated regarding the proper use of the various specialist types so that the branch office can operate more effectively.

**Rule 1. True specialist**

In an office with true specialists, the percentage of specialists must not exceed the percentage of specialist type calls.

**Rule 2. Limited specialists**

In an office with limited specialists, the percentage of specialists should not exceed the percentage of specialist type calls.

For an office with senior specialists no rule can be stated explicitly because, although the operating curve will always be superior to the all generalist office, senior specialists are more costly in terms of salary and training. Hence, some type of performance versus cost tradeoff must be made in this case.

In the preceding analyses the percentage of specialist type calls has been held constant while the percentage of specialists has been varying. It is interesting to note the impact on response time of varying the percentage of specialist type calls while holding the percentage of specialists fixed. Figure 8 is a graphical representation of these phenomena for the three different types of specialists.

It clearly shows that the lower the percentage of specialist type calls the more sophisticated the specialist type must be.

**Table III**

<table>
<thead>
<tr>
<th>Device</th>
<th>% Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>53</td>
</tr>
<tr>
<td>B</td>
<td>55</td>
</tr>
<tr>
<td>C</td>
<td>27</td>
</tr>
<tr>
<td>D</td>
<td>22</td>
</tr>
<tr>
<td>E</td>
<td>50</td>
</tr>
<tr>
<td>F</td>
<td>40</td>
</tr>
</tbody>
</table>

From the collection of the Computer History Museum (www.computerhistory.org)
in order to keep response time within reasonable bounds. Point A represents the equality of the percentage of specialist type calls and the percentage of specialists in the office. Clearly, when the latter exceeds the former (i.e., to the left of Point A) the use of true specialists becomes inferior to the all generalist case. To the right of point A, however, the three curves converge since all specialists are kept busy on their specialty.

IMPLEMENTATION

Before committing to the implementation of an engineer specialization program it was first necessary to determine the amounts by which specialization could actually reduce MTTR. Although I am not at liberty to discuss the complete nature and extent of the tests that were conducted to make these determinations, I have noted in Table III some sample percentage reductions in repair times for specialists over generalists for six individual (but unnamed) devices.

With this and other supporting data at hand our company has decided to implement a field service (i.e., hardware) specialization program. Our management science consulting organization is presently helping individual branch offices determine when and if and what kinds of specialization should be utilized. By modifying the simulation model described in this paper to reflect individual branch office operating parameters various alternative specialization schemes can be evaluated for a given office. In essence, a decision support system is being utilized jointly by our consulting group and by our branch management personnel to determine the office structures that will both maximize customer satisfaction and be cost effective. Given our belief that customer service issues will be the driving force in our industry in the 1980s, our dedication to the task at hand must be complete. Whereas, our emphasis is on the field service organization today, it will most certainly be on the software services business tomorrow. Modification of the simulation model (or the decision support system, if you will) to that end is a certainty.