user/managers in chargeout group 5 participating in training is higher than the user/managers in chargeout group 1. Compared to chargeout groups 2, 3, and 4, there is a 100% chance. For further discussion on this methodology, see [20, Sec. 4].

References

1. Delehanty, G.E. Computers and organization structures in life insurance firms: the external and internal economic environment. In *The Impact of Computers in Management*, C.A. Myers, Ed., M.I.T. Press, Cambridge, Mass., 1967.
2. Guest, R.H. *Organizational Change: The Effect of Successful Leadership*. The Dorsey Press, Homewood, Ill., 1962.
3. Guion, R.M. A note on organizational climate. *Organizational Behavior and Human Performance 9* (1973), 467–476.
4. Guthrie, A. Attitudes of middle managers toward management information systems. Ph.D. Diss., U. of Washington, Seattle, Wash., 1971, pp. 61–67.
5. Hand, H., Richards, M., and Slocum, J.W. Jr. Organizational climate and the effectiveness of a human relations training program. *Academy of Management J. 16* (1973), 185–195.
6. Holloman, R. A problem-solving approach to changing organizational climate. Presented at Tenth Annual Meeting, Eastern Academy of Management, Philadelphia, Pa., May 1973.
7. Katz, D. The functional approach to the study of attitudes. *Public Opinion Quart., 24* (Summer 1960), 163–204.
8. Kmenta, J. *Elements of Econometrics*. Macmillan, New York, 1971.
9. Lawler, E.E. III, Job attitudes and employee motivation: Theory, research and practice. *Personnel Psychology 23* (Summer 1970), 223–237.
10. Leavitt, H. Applied organizational change in industry: Structural, technological, and humanistic approaches. In *Handbook of Organizations*, Rand McNally, Chicago, Ill., March ed., 1965, pp. 1144–1170.
11. Lewin, K. Group Decision and Social Change. *Readings in Social Psychology*, E.E. Maccoby, T.M. Newcomb, and E.L. Hartley, Eds., Holt, Rinehart, and Winston, New York, 3rd ed., 1958.
12. Lucas, H.C. Jr. A user oriented approach to systems design, Proc. ACM 1971 Annual Conf., pp. 325–338.
13. Lucas, H.C. Jr. Performance and the use of an information system. *Management Sci., 21*, 8 (April 1975), 908–919.
14. Miller, N., and Dollard, J. *Personality and Psychotherapy*. McGraw-Hill, New York, 1950.
15. Nolan, R.L. Managing the computer resource: a stage hypothesis. *Comm. ACM 16*, 7 (July 1973), pp. 399–405.
16. Nolan, R.L. Plight of the EDP manager. *Harvard Business Review 51*, 3 (May–June 1973), 143–152.
17. Nolan, R.L. and Seward, H.H. Measuring user satisfaction to evaluate information systems. In *Managing the Data Resource Function*, R.L. Nolan, Ed., West Pub. Co., St. Paul, Minn., 1973, pp. 253–275.
18. Porter, L.W. A study of perceived need satisfaction in bottom and middle management jobs. *J. Applied Psychology 45*, 1 (Feb. 1961), 1–10.
19. Porter, L.W. and Lawler, E.E. Properties of organization structure in relation to job attitudes and job behavior. *Psychological Bull. 41* (1965), 23–51.
20. Schlaifer, R. User's Guide to the AQD Collection. Grad. School Bus. Admin., Harvard U., Boston, Mass., 1975.
21. Simpson, D.B. Leadership behavior need satisfaction, and role perceptions of labor leaders: A behavioral analysis. Ph.D. Diss. U. of Washington, Seattle, Wash., 1971.
22. Sorcher, M. and Danzig, A. Charting and changing the organizational climate. *Personnel 46* (1969), 16–28.

Management Applications — H. Morgan Editor

Cost/Utilization: A Measure of System Performance

Israel Borovits and Phillip Ein-Dor
Tel-Aviv University

A method is presented for evaluating computer system performance in terms of a cost/utilization factor and a measure of imbalance. These coefficients indicate the extent to which the total system cost is effectively utilized. The method includes a technique for the visual representation of system performance.

Key Words and Phrases: computer system, performance evaluation, cost/utilization, system balance
CR Categories: 2.44

Introduction

A considerable number of criteria have been developed for evaluating the performance of computer systems and many papers devoted to them [8]. These may be divided into four main groups:

(a) criteria relating to system reliability, such as mean time between failures or percentage of downtime (e.g. [4]);

(b) criteria relating to amount of work done, including throughput and job time ([6, 10]);

(c) criteria relating to user satisfaction, which may be objective measures such as response time and

Authors' address: Faculty of Management, Leon Recanti Graduate School of Business Administration, Tel-Aviv University, University Campus, Tel-Aviv, Israel.

turnaround or subjective evaluations by users ([3]);

(d) criteria relating to economic effectiveness.

The last set of criteria may be further subdivided into two subsets. The first of these addresses the question, "Is the computer doing jobs which are cost effective?" A number of studies have approached this issue (e.g., [1, 3, 9]). The second subset should answer the question, "Is the system being used with economic efficiency?" In other words, is it doing whatever it does as cheaply as possible? The accent here is on "cheaply," as measured in dollars and cents, not on "efficiently," as measured in terms of cycles or milliseconds (group (b) above). This is the major point of contact between DP managers and staff management. To the best of our knowledge, virtually nothing has been published on this issue, and in this paper we address ourselves to it.

Before beginning the analysis, it may be useful to present a framework within which one may utilize yet another measure of performance. Ideally, one would like a measure of system performance which would integrate all the multidimensional criteria mentioned above and provide one inclusive index. As a minimum, we require a tool permitting the various criteria to be considered simultaneously, and tradeoffs between them evaluated.

A significant advance in this regard, in the opinion of the authors, is the development by Philip Kiviat of "Kiviat charts," which relate the various aspects of system performance, allowing one to see at a glance whether, in general, the system under consideration is well utilized or not [7]. Such charts are prepared by first listing system parameters and the actual and ideal percentage utilization of the system with respect to each parameter. For an example see Table I. Note that the number of parameters for which the ideal is 0% should equal the number for which it is 100%. Next, the parameter values are plotted on the radii of a circle, the center of which represents 0% and the circumference 100%. Each radius represents one parameter. Parameters with 0% and 100% ideal values alternate around the circle. The example in Table I yields the chart of Figure 1(a).

In a well utilized system, the shaded area is star shaped, with points close to the circle, e.g. Figure 1(b). In the ideal case, the shaded area is reduced to alternate radii of the circle, Figure 1(c).

One of the problems with this method of presentation is that parameters are not scaled according to their economic significance. Thus it is possible to configure a system such that, technically, utilization will be satisfactory on most dimensions, at the expense of grossly underutilizing some other dimension which may be extremely significant in terms of its cost. In general, when systematic analysis has been employed in analyzing system performance, it has almost invariably been limited to hardware and software performance and there are few reported attempts to reach economic conclusions from performance monitoring data. (One notable exception to this rule is Gold [3], albeit in a somewhat different context.)

This paper is an attempt to expand the technological aspect of system performance evaluation to cover some of the economic aspects in the form of cost/utilization evaluation. The authors do not claim to have found an all-inclusive answer to the problem of

Table I. System Parameter Utilization.

Parameters	Percent utilization ideal	actual
I. Memory used for control programs	0	10
II. Memory used for production programs	100	90
III. Bulk storage units inactive	0	20
IV. Channel utilization	100	50
V. CPU wait time	0	30
VI. CPU active	100	.0
VII. CPU active on control programs	0	20
VIII. CPU active on production programs	100	80

Fig. 1. Examples of Kiviat charts.

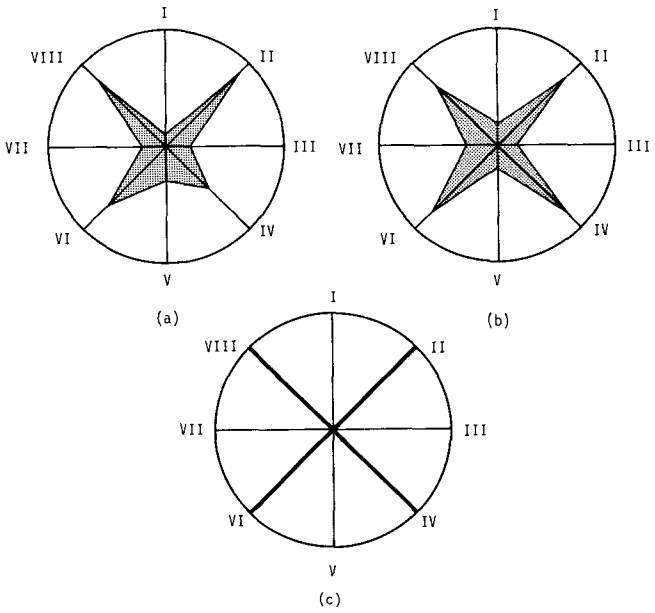


Table II. Percent of System Cost Contribution by Various Components.

Component	Percent of cost
1. CPU	41
2. Memory	23
3. Disks	13
4. Tapes	12
5. Card reader	2
6. Printer	9
	100

economically evaluating system performance; they do feel, however, that this is a first step. This technique bears a dual relationship to the Kiviat chart. On the one hand, it provides two relative, unscaled measures, which can be incorporated in such charts; on the other hand, it is a variant of the underlying concept of visual presentation developed by Kiviat.

The technique suggested in this paper assumes the existence of performance monitoring devices to provide the basic data. At a minimum these include system accounting procedures, either manufacturer provided, such as IBM's MSF [5] or CDC's Scope 3.4 [2] or commercial packages such as Value Computing's Comput-a-charge. In principle the basic model can be expanded to integrate cost-expressible data developed from any hardware or software monitor.

Components of Economic Efficiency and the Concept of Cost/Utilization

One may think of the attainment of economic efficiency in a system as a two-step process. First, the system must be applied in an area in which its benefits outweigh its costs. Second, for *maximum* efficiency to be achieved in a given situation, it must be applied in such a way that the costs are minimal. In many cases, the first stage of the analysis is made, especially before system installation; the second stage is often ignored. But obviously if a system is cost-effective in a given application, the same system doing more work, or a less expensive system doing the same things, will be even more profitable.[1]

Our analysis begins with two propositions. First, in order for a system to perform at minimal cost it must be utilized to a considerable extent—ideally, at as high a rate as is consistent with availability and user satisfaction criteria. Second, a high rate of utilization at minimum cost is attainable for the system as a whole only if it is balanced—i.e. if all components are utilized to about the same extent. If this is not the case, those components which are most heavily utilized will tend to form bottlenecks. These bottlenecks will preclude the attainment of higher rates of utilization for the underutilized components.

The first problem encountered is that of determining the level of utilization for a whole system consisting of a number of different components. It appears to the authors that the cost of any unit should weight the significance with which one reacts to information on its utilization. Thus the knowledge that a card reader, whose cost is 2% of total system cost, is utilized only 30% of the time is of different import than knowledge of the same degree of utilization with respect to the core memory, whose cost is 25% of total system cost.

This leads us to the concept of cost/utilization, which applies utilization figures indirectly to the costs of physical units rather than to the units themselves. Cost is a common dimension which allows us to integrate utilization data for all the components of a system. We can then develop a single measure of cost/utilization for an entire system. While it is impossible to develop meaningful figures of total system performance directly from physical utilization, the common dimension of cost makes a single measure of merit both feasible and meaningful.

The cost/utilization factor measures the extent to which the outlay on the total system is actually utilized. It is computed as $F = \sum_i P_i U_i$ where P_i is the cost of the ith component in the system as a percentage of total cost and U_i is the percentage utilization of the ith component.

Thus F can vary from zero, in a system not utilized at all, to 1 in a perfectly utilized system. The method described in the next section includes both computation of the cost utilization factor and its visual representation.

We can now use these same data to develop a measure of system imbalance. Such a measure is

$$B = 2[\sum_i (F - U_i)^2 P_i]^{\frac{1}{2}}$$

where B is the measure of imbalance and F, U_i, and P_i are as previously defined.

The expression included inside the square brackets measures the variance of the degree of utilization of individual components (U_i) around their weighted mean (F). Multiplying these squared deviations by the relative cost of the components gives the variance of utilization of units of cost rather than of physical components.

The scaling factor, 2, normalizes B so that it varies between 0, for perfectly balanced systems, and 1, for maximally unbalanced systems. A perfectly balanced system is defined as one for which

$$U_i = F, \quad \text{for all } i.$$

A maximally unbalanced system is defined as one for which

$$U_i = 1, \quad \sum_{i=1}^{m} P_i = 0.5,$$

and

$$U_i = 0, \quad \sum_{i=m+1}^{n} P_i = 0.5,$$

so that

$$F = \sum_{i=1}^{m} U_i P_i + \sum_{i=m+1}^{n} U_i P_i$$
$$= 0.5 + 0 = 0.5.$$

[1] One of the authors, a reformed computer salesman, has often been guilty of trying to persuade customers that "It doesn't really matter how much the system costs, so long as it's cheaper than what you have been doing to date."

Then

$$B = 2[(0.5 - 1)^2 \sum_{i=1}^{m} P_i + (0.5 - 0)^2 \sum_{i=m+1}^{n} P_i]^{\frac{1}{2}}$$
$$= 2(.25*.5) + (.25*.5)$$
$$= 1$$

The authors are currently working on an analysis of the characteristics of this measure of imbalance.

Cost/Utilization: Computation and Representation

The method is to construct a rectangular graph representing the maximum possible cost/utilization in the system. The horizontal axes of the graph represent percentage of total system cost contributed by each type of hardware component. The vertical axes represent percentage utilization of hardware components. A histogram is constructed within this graph, the bars of which represent utilization of system cost and the area above the bars represents slack in the system. Once this histogram has been constructed, the cost/utilization factor is calculated, relating cost utilized to total system cost.

The following are the steps to be taken in this analysis:[2]

1. List system components and their cost.
2. Compute the percentage of total cost represented by the cost of each component.
3. On the horizontal axis, mark off cumulatively the percentage of total cost represented by each type of component. This is the base of the histogram.
4. On this base construct a bar for each type of component, the height of which represents the percentage utilization of that component.
5. Compute the cost/utilization factor F, as described in the previous section.
6. Compute the measure of imbalance B, as described in the previous section.

Example

In order to illustrate this method of cost/utilization analysis a typical computer system has been chosen that includes: 500K memory, four disk drives, eight tape drives, card reader, and line printer. The distribution of the total cost among the various components (in percent) is given in Table II (p. 186).

[2] The authors may appear to have taken a rather cavalier attitude to measurement problems. This is not because we are unaware of them but because the major objective of this paper is to develop the concept of cost/utilization analysis rather than the techniques. Three of the major problems we have encountered are (a) the choice of criterion for disk utilization—storage space or access time, (b) the determination of cost for package-priced multifunctional units—e.g. channels or peripheral processors priced together with the CPU, and (c) the use of historical or replacement costs—especially in installations with equipment from various generations.

Fig. 2. Illustrative cost/utilization histograms.

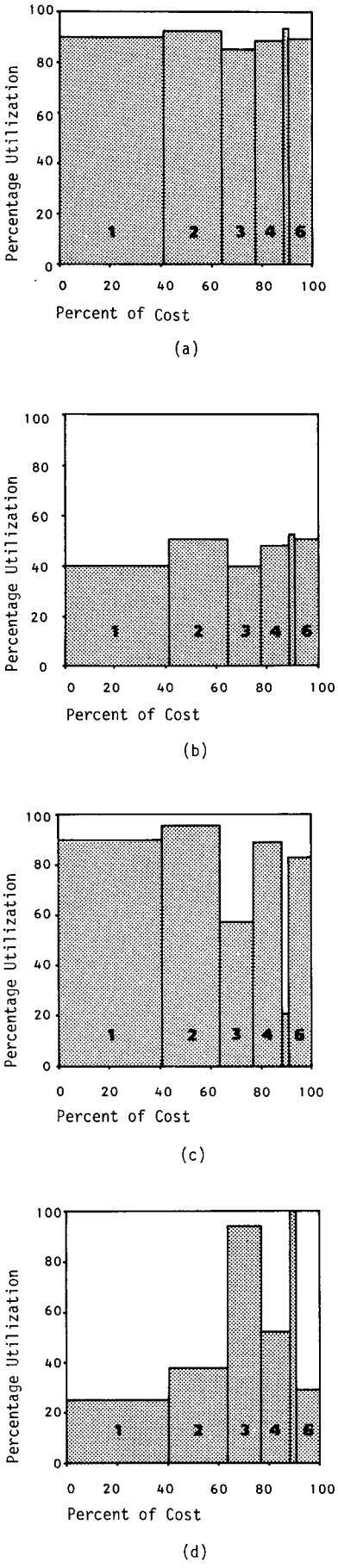


Fig. 3. Cost/utilization histogram with additional data incorporated.

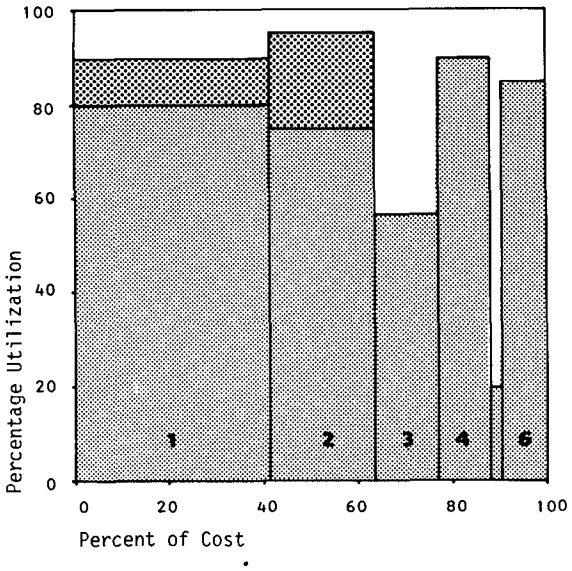


Fig. 4. Illustration of trace of cost/utilization criteria.

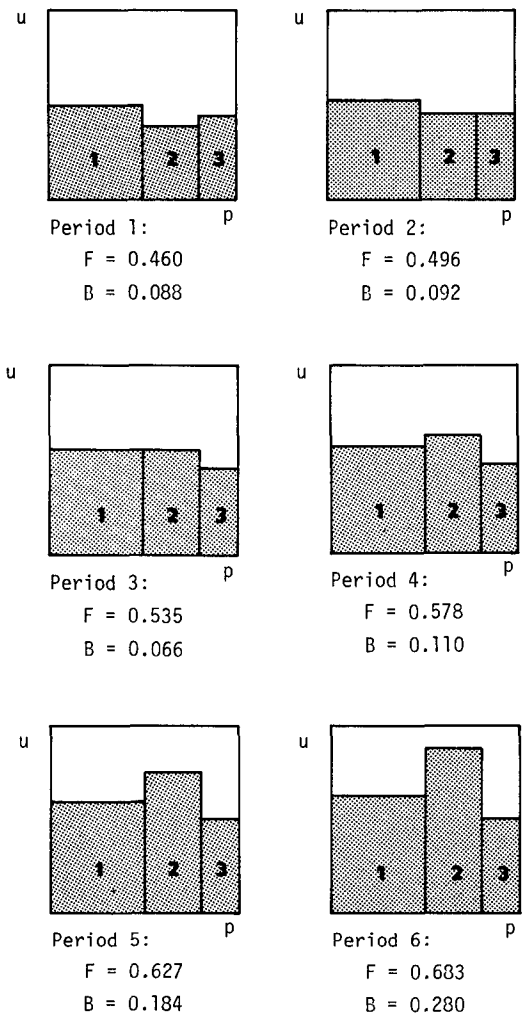


Figure 2 represents four classes of systems:

(a) The system is well balanced and well utilized. Balance is indicated by the fact that all bars in the histogram are about the same height, i.e. all components are utilized to about the same extent. Near full utilization is indicated by the fact that all bars are close to their upper bound, i.e. all components are utilized almost to the full. The cost utilization factor here is 0.903, and the measure of imbalance is 0.062.

(b) The system is well balanced but underutilized. The cost/utilization factor is 0.443, and the measure of imbalance 0.1.

(c) In this case the system is relatively well utilized but is somewhat unbalanced. The cost/utilization factor is 0.853, and the measure of imbalance is 0.3.

(d) This system is both unbalanced and poorly utilized. The cost/utilization factor is, at present, 0.432 and the measure of imbalance 0.51.

It is possible to extend this analysis to include additional information, if available. Thus if the operating system absorbs 20% of the core memory and 10% of CPU time, these facts can be incorporated in the histogram, as shown in Figure 3.

The supervisory system overhead is represented by the shaded area in the diagram. System utilization can now be broken down as follows:

Cost utilization:		
Production	.772	
System overhead	.081	
Cost/utilization factor		.853
System slack		.147
		1.000

As additional information relevant to the cost/utilization ratio becomes known, it can be incorporated into this scheme of analysis. It should be noted that this analysis depends on the additivity property of the cost/utilization factor. If $(U_R)_i$, $(U_0)_i$, and $(U_S)_i$ represent the percentage utilization of component i for production, overhead, and slack, respectively, then

$$(U_R)_i + (U_0)_i + (U_S)_i = 1.$$

Furthermore, since $\sum_i P_i = 1$, then

$$\sum_i (U_R)_i P_i + \sum_i (U_0)_i P_i + \sum_i (U_S)_i P_i = 1$$

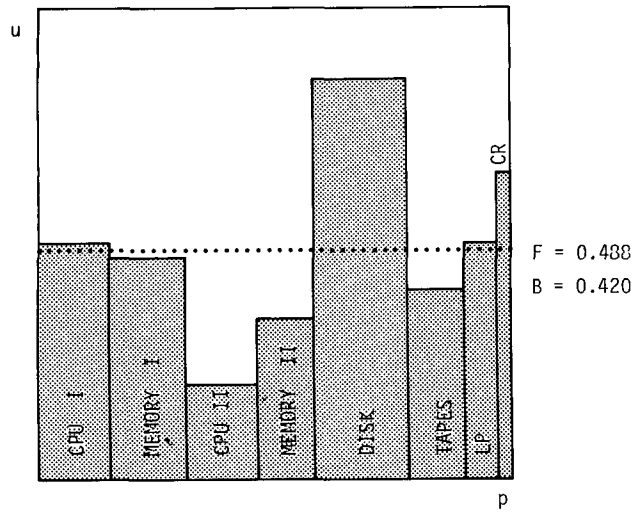
i.e. $F_R + F_0 + F_S = 1$.

The additivity property does not hold, however, for the measure of imbalance. Thus one cannot compute directly the contribution of each of the uses of the system to total system imbalance.

Use of Cost/Utilization Criterion

Having suggested two additional criteria of system performance, it is now incumbent upon us to demonstrate their uses. In the course of so doing, we may

Fig. 5. Composite cost/utilization histogram for two real systems.



also answer some of the questions which have come to the reader's mind to this point.

The first use is as a control device. Tracing the trends in the cost/utilization and balance factors can indicate where potential bottlenecks are developing, and at what rate the system is approaching saturation. Consider the example in Figure 4 relating to six periods for a hypothetical system composed of three components. It is clear that the system will be fully loaded within two or three periods because of a bottleneck developing in component 2. An increase in the capacity of this component could defer saturation for some time. The utilization of component 3 is increasing at a lower rate than F, indicating that it is a potential source of system underutilization.

This example is in fact based on two systems actually studied with the cost/utilization criterion. Figure 5 is a composite of the cost/utilization histograms containing some of the interesting features of the two systems mentioned. Both systems ran three full shifts. In both cases, F was about 0.5, and in both there was a bottleneck in disk drives preventing higher utilization. Furthermore, in both systems, the tape drives were grossly underutilized.

This brings us to a second use of the method—as a guide to improving system configuration. In the cases cited, one obvious improvement would be to increase disk capacity and decrease tape-drive capacity. A somewhat more subtle improvement emerges from a comparison of CPU and memory utilizations. There are two identical CPUs with different sizes of core memory. It is obvious at a glance that, for the mix of jobs involved, the larger core permits a more even balance between CPU and memory utilization.

[3] Our recommendation to reduce tape drive capacity also raises the issue of sunk costs. If it turns out that the excess capacity is purchased, it may not be economic to reduce it. This aspect is not considered in the cost utilization analysis, but does not, we believe, detract from its validity. The implication is that it may not always be economically feasible to make the adjustments indicated by the analysis, but the analysis itself is no less valid for that reason.

The recommendation to add disk drives gives rise to questions concerning interactions between units.[3] Is it true, in the example considered, that a doubling of disk capacity will lead to a doubling of system capacity, or will some new bottlenecks emerge, say in channels? This kind of interaction is not treated explicitly in the cost/utilization analysis, but it may, nevertheless, provide some insight. Since one can increase the level of detail in the histogram at will, we could separate the channels from the CPU in which they are currently incorporated. This would provide at least a first approximation to the effect of adding the disks. If the channels turn out to be highly utilized, it is reasonable to believe that adding additional disk drives may saturate them. If, on the other hand, they are very much underutilized, this danger is remote. An additional example of the way interactions are handled is in the discussion in the preceding paragraph of the balance between CPU and memory.

From the preceding examples it should now be clear why the concept of balance is so important. If a system is being operated three shifts daily, this may be because the whole system is highly utilized, or it may be because the system is badly balanced, and a bottleneck has developed, which extends job times beyond what would be necessary if the system were balanced. Removing the bottleneck may significantly improve throughput and turnaround. The concept of balance becomes more important, of course, as the degree of utilization increases and the system approaches saturation. If unrecognized, a state of imbalance may lead to acquisition of a larger system, when removing a bottleneck would solve the problem. Furthermore, since bottlenecks reduce throughput, lack of balance implies not only excess outlays on underutilized equipment but also excess outlays for personnel and utilities.

Finally, the cost/utilization histogram is a point of contact between data processing managers and staff functions. Probably one of the more difficult problems of DP managers is the lack of confidence engendered in top management and staff by their inability to evaluate the efficiency of data processing operations. Typically, this problem arises when approval is required for the purchase of new equipment or when the level of service appears to be inadequate. A difficulty in this respect is that peak cost/utilization for computer systems seems to be about 0.75. When the question arises as to why this should be acceptable as full utilization, the Kiviat chart can be of considerable help in explaining how pushing one point of the star out too far may be deleterious in causing other points to be withdrawn.

Summary

The concept of cost/utilization is an additional tool in the bag of the system performance evaluator.

Its principal contribution is in the inclusion of the economic dimension in an area which has been dominated by purely technological considerations, and its integration with those considerations by means of the Kiviat chart. Its simplicity and ease of visual representation should help bridge the communication gap between data processing staff and top management.

Received October 1974; revised November 1975

References

1. Ackoff, R.C. Management Misinformation Systems. *Manage. Sci. 14*, 4 (Dec. 1967), 147–156.
2. *CDC—Scope User's Guide*, Software Documentation Div., Control Data Corp., Sunnyvale, Calif.
3. Gold, M. Time sharing and batch processing: An experimental comparison of their value in a problem solving situation. *Comm. ACM 12*, 5 (May 1969), 249–259.
4. Hughes, J. Performance evaluation techniques and system reliability—a practical approach. ACM/NBS Performance Evaluation Workshop, San Diego, 1973, NBS Special Pub. 406, NBS, Washington, D.C., 1975, pp. 87–97.
5. IBM System/360 Operating System: Planning for System Management Facilities. C28-6712-1, IBM Data Processing Div., White Plains, N.Y.
6. Lucas, H.C. Jr. Performance evaluation and monitoring, *ACM Computing Surveys 3*, 3 (Sept. 1971), 79–92.
7. Morris, M.F. Kiviat graphs—conventions and figures of merit. Performance Evaluation Rev. (ACM SIGMETRICS newsletter) *3*, 3 (Oct. 1974), 2–8.
8. Performance Evaluation Bibliography. Performance Evaluation Rev. (ACM SIGMETRICS newsletter) *2*, 2 (June 1973), 37–49.
9. *Unlocking the Computer's Profit Potential*. McKinsey & Co., New York, 1968.
10. Watson, R. Computer performance analysis: Applications of accounting data. Rep. No. R-573-NASA/PR, Rand Corp., Santa Monica, Calif., May 1971.

Short Communications
Operating Systems

A Comparison of Next-fit, First-fit, and Best-fit

Carter Bays
University of South Carolina

Key Words and Phrases: memory allocation, first-fit, best-fit, next-fit
CR Categories: 4.32, 4.35

"Next-fit" allocation differs from first-fit in that a first-fit allocator commences its search for free space at a fixed end of memory, whereas a next-fit allocator commences its search wherever it previously stopped searching. This strategy is called "modified first-fit" by Shore [2] and is significantly faster than the first-fit allocator. To evaluate the relative efficiency of next-fit (as well as to confirm Shore's results) a simulation was written in Basic Plus on the PDP-11, using doubly linked lists to emulate the memory structure of the simulated computer. The simulation was designed to perform essentially in the manner described in [2]. The results of the simulation of the three methods show that the efficiency of next-fit is decidedly inferior to first-fit and best-fit when the mean size of the block requested is less than about $\frac{1}{16}$ the total memory available. Beyond this point all three allocation schemes have similar efficiencies.

Author's address: Department of Mathematics and Computer Science, University of South Carolina, Columbia, SC 29208.