SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853-3801

TECHNICAL REPORT NO. 938

October 1990

STRATEGIC PLANNING IN FIELD SERVICE: TRADE-OFFS BETWEEN FIELD ENGINEERS AND INVENTORY

by

Alisha A. W. Waller

Strategic Planning in Field Service: Trade-offs Between Field Engineers and Inventory

In practice, budgets for staffing of field service support and for spare parts inventory investment are made without regard for the interrelationships that may exist among customer service levels, field engineer utilization, and spare parts fill rates. This paper presents a rough-cut optimization model used to explore qualitative insights into the trade-offs among these variables. Extensions to the model include multiple customer classes and multiple part types.

I. Introduction

The rapid rise in the technological complexity of manufacturing environments during the past several decades has shifted the burden of equipment repair from the user to the original manufacturer. A general "handy man" is no longer a sufficient maintenance and repair crew even for a small manufacturer. However, the cost of training a crew to repair the diverse pieces of equipment found in a modern factory is enormous. Therefore, equipment manufacturers are receiving increased pressure to provide repair services as part of the lease agreement, as part of the sale contract, or as a distinct, marketable product.

Since, in general, the equipment is too large, too heavy, and too expensive to ship to a repair facility, the original manufacturer must provide this maintenance and repair at the permanent location of the equipment. This on-site repair is called field service. Hence, the field engineer must travel to the customer's location to perform the field service.

Given a fixed budget to provide both a crew of field engineers and their spare parts kits, the strategic planner has a continuum of options, e.g.the more field engineers that are used, the less money available for spare parts kits. In this paper, we build a rough-cut optimization model to examine the trade-offs between field engineer staff size and

spare parts kit composition. Our model minimizes the expected total service time subject to a budget constraint. In addition to finding the optimal balance between inventory and field engineers, this model allows a marginal analysis of an increased budget.

The current literature which addresses strategic planning in field service support falls in three basic categories. The planning of the field service staff size (or equivalently, the area one field engineer can service) is studied by Hambleton[82], who develops a rough-cut model by equating the expected number of required service hours with the number of service hours provided by N field engineers, and by Smith[80], who determines the area one field engineer can service and achieve a desired average response time. The second category of literature is the spare parts inventory kitting problem, which has been studied extensively. The most influential papers to our work include those by Graves[82] and Mamer and Smith[82]. Each of these authors minimize cost subject to performance criteria constraints; however, we maximize the performance criteria subject to a budget constraint. The final category is the evaluation of customer service level, which is primarily done by Agnihothri[88]. Each of the authors maintained a separation

between the three categories. Our model, however, considers the simultaneous determination of the number of field engineers and the spare parts kitting. We want to formalize the following program:

minimize E(Total Service Time)

s.t.

E(field engineers cost) + E(inventory holding cost) < Budget.

In order to develop our model, we make some basic assumptions about the field service operation. First, we assume historical or hueristic estimates are available for the travel time between customer sites, which is assumed constant and independent of the number of field engineers. We assume that the repair time is constant once the necessary parts are procured and that an estimate of this repair time is available. The service request arrival rate, denoted by λ, is known and constant. Finally, we assume that the calls are handled in a first-in-first-out (FIFO) manner, equally divided among the field engineers.

II. Realizations of a service call

In order to understand the expected total service time, let us first examine the possible realizations of a service call.

As Figure [1] shows, the first activity initiated by a customer call is phone diagnosis by the hot-line attendant (D_1). This phone diagnosis may be as simple as questions such as "Is the machine plugged into the wall socket securely?" or it may be as complex as an expert system using a detailed knowledge base of the machine. The phone diagnosis may be sufficient to resolve the customer's problem, which would complete the service. Denote the probability of the diagnosis not being sufficient by p_1.

If the phone diagnosis is not sufficient, then the customer's request is added to the queue and the customer begins his wait for the field engineer to arrive (Wq1). After a random amount of time a field engineer is assigned to this call, and begins his travel to the customer's location (T).

Upon arrival, the field engineer begins the on-site diagnosis (D_2). With probability p_2 the field engineer finds that a replacement part is required. In this case, the field engineer may obtain the part from his spare parts kit, the local depot, or the central warehouse. If the part is in the kit, then the repair may start immediately (R_P). With probability ($1-p_2$), the field engineer finds that no part is needed and only machine adjustments are required (R_A).

When a part is required, let $p_3(K)$ be the probability that the needed part is not in the kit. Note that $p_3(K)$ is a function of the spare parts kit. If the part is not in the kit, the field engineer must decide whether to break the repair, leaving to service another customer while the part is being delivered, or to continue the repair, waiting at the current customer's site until the part arrives (L_{FE}) and then completing the repair (R_p). Let p_4 be the probability of breaking the repair given the needed part is not in the kit.

The decision of when to wait and when to break the repair is one of the many policy decisions which affect the customer service level. The handling of the call after the decision to break the repair has been made is also an important policy decision. One policy example is to break the repair only if the part delivery is expected to take more than two hours and once the part has arrived at the customer's site, complete the repair as soon as possible. In this policy example, the queue of calls may be thought of as having two classes of priority customers, where the high priority class are those customers whose repairs were broken while the parts were being delivered. A break-of-repair policy should attempt to

balance the extra travel time involved in leaving and returning to the customer with the part acquisition time.

If the repair is broken, the customer experiences additional waiting times for the part to be delivered (L), in the queue for a field engineer to be assigned(Wq2), for the field engineer to travel (T), and finally the repair time (R_P).

From this diagram we can calculate the expected total service time for a customer as

$$E(T.S.T.) = D_1 + p_1[Wq1 + T + D_2 + (1-p_2)R_A + p_2(1-p_3(K))R_P + p_2p_3(K)(1-p_4)(L_{FE}+R_P) + p_2p_3(K)p_4(L + Wq2 + T + RP)\ ]$$

where

Wq1 = E(Wait in queue for initial service request)

Wq2 = E(Wait in queue for return of field engineer after the repair is broken)

L_{FE} = E(Part Acquisition time when field engineer is waiting)

L = E(Part Acquisition time when repair is broken)

R_A = E(Repair time when no parts are required)

R_P = E(Repair time when a part is required)

T = E(Travel time between customers).

Since this model is concerned with the general trade-offs

between field engineers and inventory, let us assume that D_1, D_2, RA, R_p, and T are constant and independent of the number of field engineers and the inventory which they carry. Also, note that p_1 is dependent only on the phone diagnostics, p_2 is dependent only on the equipment repairability, and p_4 is a policy decision. Therefore, each of these probabilities are independent of the number of field engineers and the spare parts kits.

III. Model 1

For our first model, we assume that the repairs are never broken, that the customer calls are handled in FIFO order, and that the field engineers follow an (s-1,s) inventory policy. We also assume that the field engineers carry identical kits of one part type and have no interaction with each other. Additionally, the call dispatch policy is independent of the current inventory status of the field engineers. Therefore, $p_3(K)$ is the complementary Poisson distribution, denoted P(K), and Wq1 can be described with the standard M/M/x wait time in queue formula.

In order to use the wait time in queue formula, we need a service rate for the field engineers. In our model, the

service rate depends on the expected part acquisition time. So we approximate the service rate with

$$\mu = 1/(T+R+p_1p_2P(K)L_{FE}).$$

Let B denote the total available budget, C_{FE} denote the annual cost of a field engineer without a kit, and H denote the annual holding cost of one part. Assuming a simple budget constraint leads to the following model:

$$\min \frac{\left(\frac{\lambda}{\mu}\right)^x \mu}{(x-1)(x\mu-\lambda)^2}\left[\sum_{n=0}^{x-1}\frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \frac{1}{x!}\left(\frac{\lambda}{\mu}\right)^x\left(\frac{x\mu}{x\mu-\lambda}\right)\right]^{-1} + p_2(1-P(K))R_p + p_2P(K)(L_{FE}+R_p)$$

subject to: $xC_{FE} + xHK \leq B$

x and K are non-negative integers.

where

x = number of field engineers

K = number of parts to carry in a field engineer's kit

Since the objective function is non-increasing in the inventory variable for a constant number of field engineers and the number of field engineers is discrete, we solve this model by evaluating the objective function for various values of x with the inventory being set as high as possible under the constraint. Therefore the optimal value is the minimum objective function value over all the feasible values of x. Although we evaluate the function for various values of x,

this procedure is efficient since a lower bound on x is (λ/μ), an upper bound on x is (B/C_{FE}), and the waiting time in a queue falls rapidly as the number of channels is increased.

IV. Model 2

To enhance the basic model, we first assume that there is a positive probability that the repair will be broken. The customers now fall into two classes: the high priority class customers are those customers whose repairs have been broken and are waiting for a field engineer to return, while the low priority class customers are the customers who are waiting for a first visit from the field engineer. We now use the queueing theory results for an M/M/x queue with two priority classes. Both classes must have the same service rate in order for the derivation of the expected waiting time to be tractable. Therefore, we use a weighted average service rate:

$$\frac{1}{\mu} = \frac{1}{(1+p_1p_2p_3(K)p_4)} [p_1T+p_1(1-p_2)R_A + p_1p_2(1-p_3(K))R_p + p_1p_2p_3(K)(1-p_4)(R_p+L_{FE})]$$

$$+ \frac{p_1p_2p_3(K)p_4}{(1+p_1p_2p_3(K)p_4} (T + R_p)$$

In addition, we can extend the basic model to include different part types. Each part type has an associated annual holding cost, H_i, and an associated probability of being the required part, G[i]. Allowing multiple part types adds another dimension to the optimization of the model, which we handle by using dynamic programming to optimize an inventory allocation submodel. For this inventory allocation submodel, we have adapted Graves' model for spare parts kitting to suit our objective function and budget constraint. This model assumes that the kit is restocked between each job and therefore either zero or one part of each type is carried in the kit. Now the variable K denotes a vector of size NP, where NP is the number of different part types. Assuming independence between part types allows us to write our inventory allocation submodel as follows.

$$\max \quad p_3(K) = \prod_{i=1}^{NP} (1-G_i)^{1-K_i}$$

$$\text{s.t.} \quad \sum_{i=1}^{NP} xH_iK_i \leq B - xC_{FE}$$

$$K_i \in \{0,1\} \quad \text{for} \quad 1 \leq i \leq NP$$

The optimization of model 2 is done by repeating three basic steps: 1) set the number of field engineers, 2) optimally

allocate the remaining budget using dynamic programming, and 3) evaluate the objective function. Again, knowing upper and lower bounds on x makes this a reasonable procedure.

V. Trade-off Analysis for Urban Environments with a Local Depot

A. Breakdown of Total Time in System

In order to begin understanding the trade-offs between field engineers and inventory, we have developed a graph of the breakdown of the expected total time a customer spends in the system. This graph also shows the optimal number of field engineers and the optimal kit composition for a range of budget values. See Figures [2], [3] and [4]. The total expected time in the system is broken into three parts, S_1, S_2, S_3.

$$S_1 = p_1(Wq1 + T + D_2) + p_1(1-p_2)R_A + p_1p_2R_P$$

$$S_2 = p_1p_2p_3(K)(1-p_4)L_{FE}$$

$$S_3 = p_1p_2p_3(K)p_4(L + Wq2 + T)$$

S_1 includes the initial wait for a field engineer, the first travel time, and the repair time, which corresponds to the time which every customer must wait. S_2 is the expected time that the field engineer waits at the customer's site for the

part to be delivered, which is dependent on the kit composition and the break-of-repair policy. S_3 is the expected additional time incurred if the repair is broken, including the part delivery time, the second wait in queue, and the second travel time. In addition, the graph has a row of symbols at the bottom of the vertical scale. The size of the symbol indicates the optimal number of field engineers and the sections of the symbol represent the different parts in the kit. If a part is carried in the optimal kit composition, then its corresponding section is shaded.

The parameters used to develop Figures [2], [3] and [4] are listed in the table below. They correspond to an urban environment with a local depot. Note that the only difference between the three parameters sets is the break-of-repair policy.

Parameter:	T	R_A	R_P	L	L_{FE}	p_1	p_2
Value	2.0	1.0	4.0	1.5	1.0	1.0	0.8

Parameter:	C_{FE}	G_1	H_1	G_2	H_2	G_3	H_3	G_4	H_4	G_5	H_5
Value	55k	0.1	1k	0.10	4k	0.25	1k	0.25	3k	0.30	2k

Data Set:	20	21	22
p_4 value:	0.10	0.90	0.50

The first observation we make is that whenever full kits are carried, (i.e. the field engineer symbol is completely

shaded) the second and third times are zero. This is evidence of the power of a full kit. If all the parts are carried, the restocking assumption implies that the probability that the needed part is in the kit is one. Therefore, the upper two branches of Figure [1] do not affect the total expected time in the system.

The second observation is that whenever the number of field engineers is increased, there is a sharp decrease in the total expected time in the system. This decrease is due to the queueing theory embedded in the model. The expected wait time in the queue is highly sensitive to the number of servers. Also, the utilization of the field engineers decreases markedly with an increase in the number of field engineers.

The next observation is that whenever inventory is decreased in order to increase the number of field engineers, (e.g. Figure [2] at $660,000) then S_2 and S_3 increase. However, the additional field engineer causes such a decrease in the expected queue time that the total expected time decreases.

The final observations come from comparing the three graphs. For the particular data sets used, the break-of-

repair policy of $p_4 = 0.1$ results in a lower expected total time in system for all of the budget levels. This is due to the lead times being so much smaller than the travel and repair times. Also note that the optimal solutions differ substantially. In Figure [4] with $p_4 = 0.9$, the optimal solution is to use fewer field engineers and more inventory. This causes more waiting time in queue, but less part delivery time.

B. The Effect of p_4 on Various Budget Levels

The next graph which we use to understand the trade-offs is a plot of the optimal total time in system for various budget levels and various p_4 levels. In Figure [5], for each p_4 level, the budget increases from the highest point to the lowest point. In other words, higher budgets achieve lower total expected times in the system. In the figure, a sampling of budget levels is marked in order to show the decreasing marginal return of additional budget dollars.

As the graph indicates, at smaller budgets the parameter p_4 has a greater effect. This is because the optimal solutions are partial kits and therefore, the decision of whether to break the repair must be made more often. By contrast, at higher budgets, the optimal solutions are full kits and the

needed part is always available. Therefore, the field engineer never encounters the break-of-repair decision and p_4 has no effect.

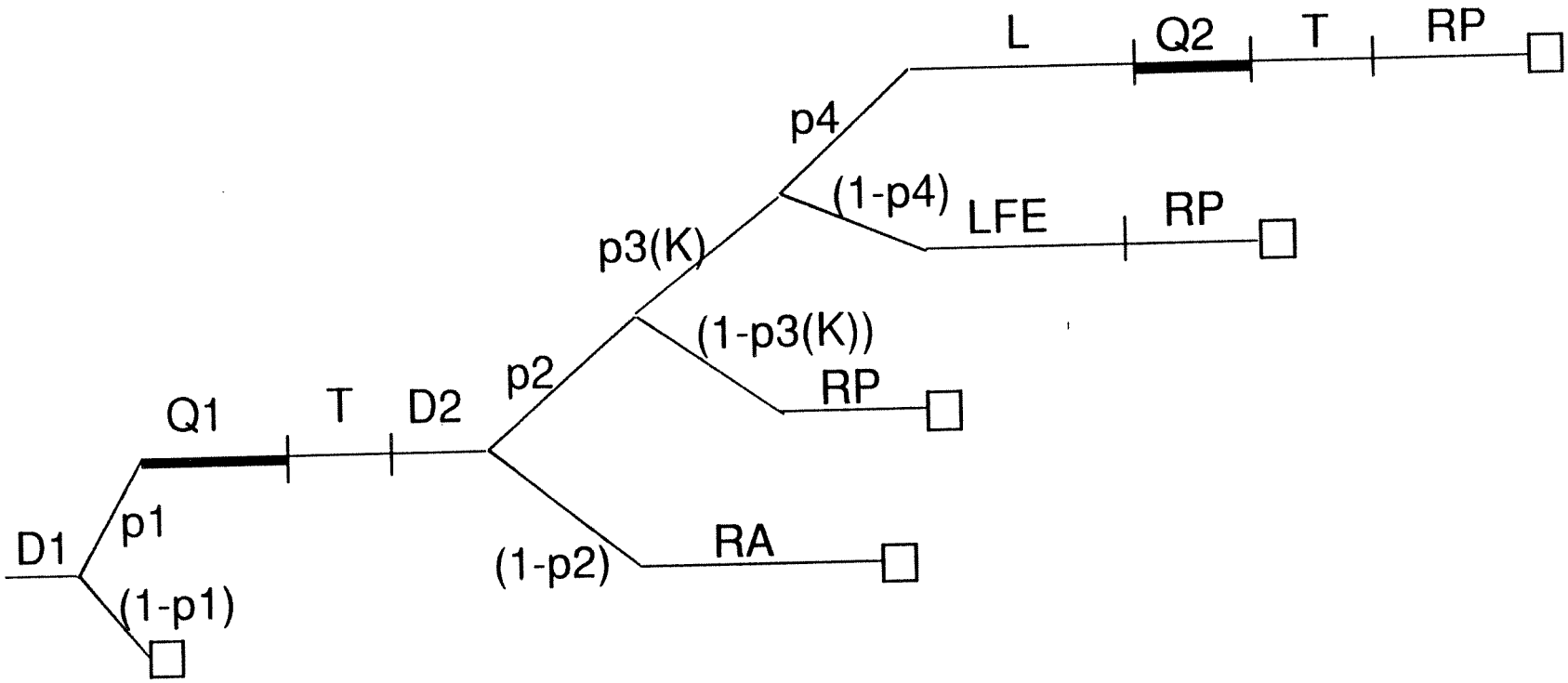
The other observation we can make by examining this graph is that small increases in the budget at $p_4 = 0.1$ can decrease the expected total time in the system, while at $p_4 = 0.9$, larger increases in the budget are needed to affect this time. The difference is due to the integrality of field engineers and parts. Since the optimal solution at $p_4 = 0.9$ usually involves fully kitted field engineers, it costs more money to add another fully kitted engineer than at $p_4 = 0.1$, where the field engineers are only partially kitted.

VI. Extensions and Further Research

At this time, we are developing several extensions to these models. The first extension relaxes the restocking assumption and uses a different inventory policy for the field engineers. The second extension allows a realistic number of parts, which is often too large to directly use the dynamic programming algorithm. The final extension allows random, rather than fixed travel times. We also believe the following research areas have great potential: non-

homogeneous field engineers, overlapping service territories, and alternatives to the FIFO service discipline. Non-homogeneous field engineers are field engineers whose skill level, training, or experience are different enough to result in service rates that are dependent on which particular field engineers is serving a particular customer. The overlapping service territories provide the opportunity to increase the average field engineer utilization and to decrease the queue length. Many companies who provide field service do not use a strictly FIFO service discipline, but instead combine FIFO, customer priority, and nearness to a field engineer's current location.

REALIZATIONS OF A SERVICE REQUEST



— known, constant time

▬ stochastic time

p1 = Prob{phone diagnosis not sufficient}

p2 = Prob{repair needs a part}

p3(K) = Prob{needed part is not in kit}

p4 = Prob{repair is broken when needed part is not in kit}

Figure [1]

BREAKDOWN OF TOTAL TIME IN SYSTEM

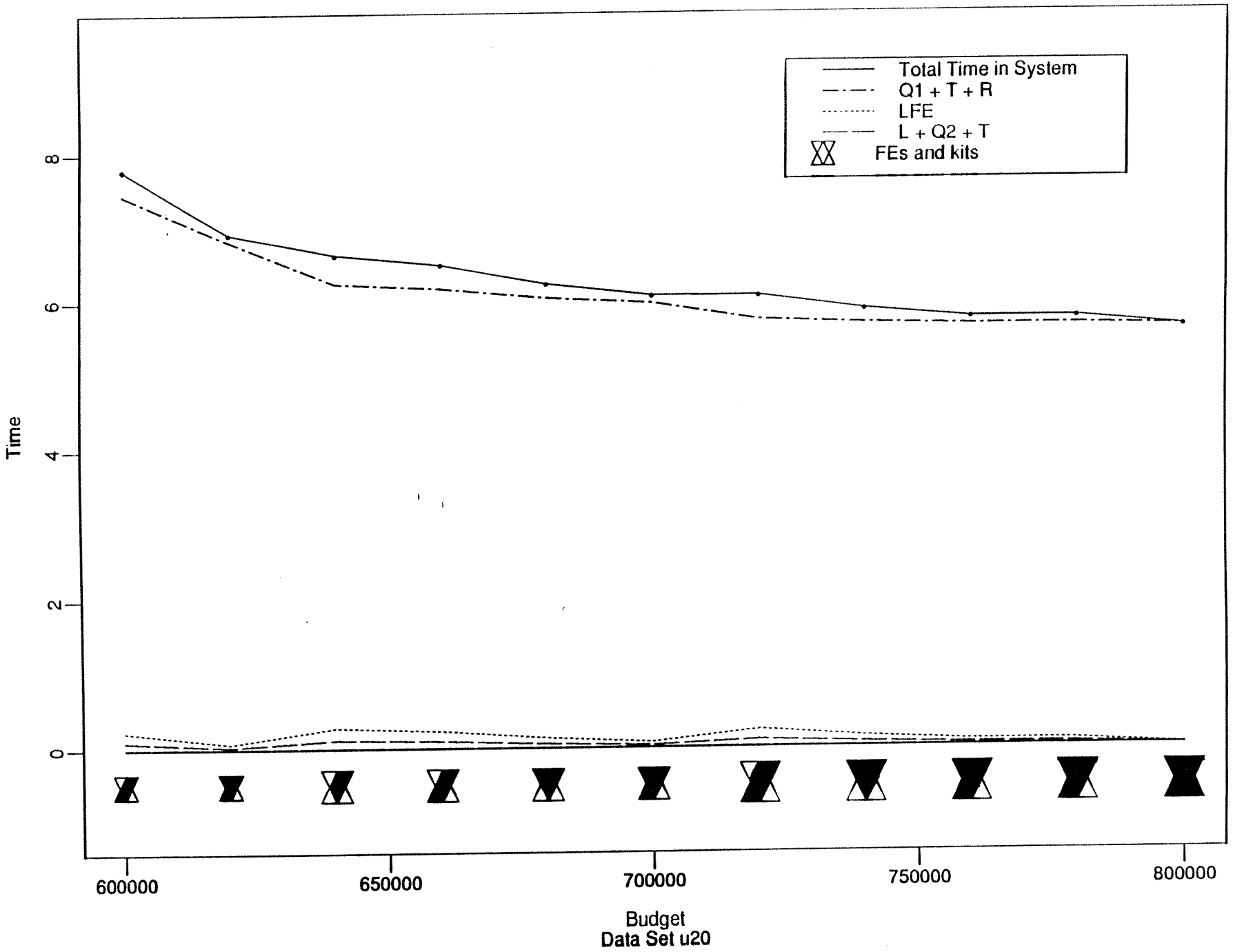


Figure [2]

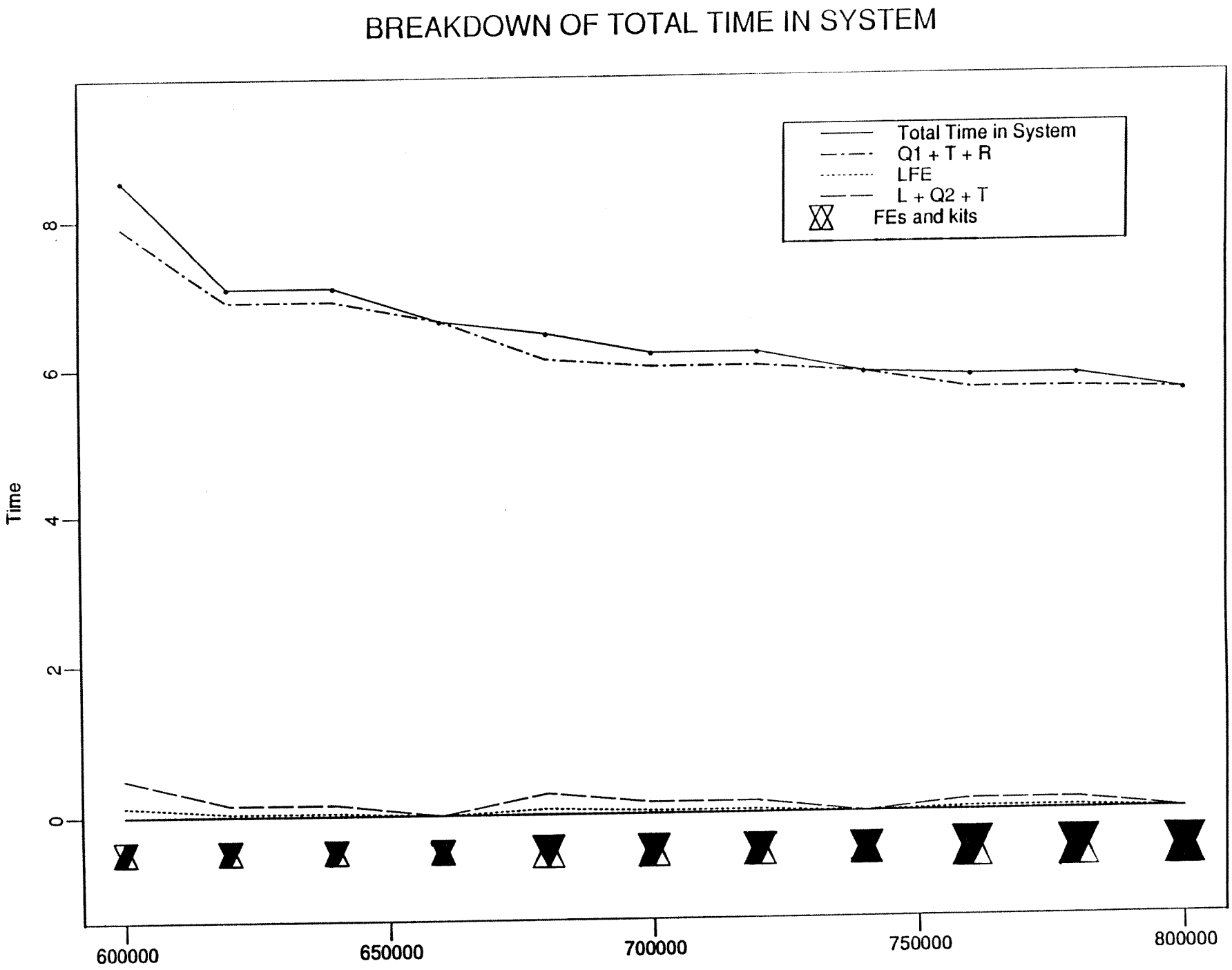

BREAKDOWN OF TOTAL TIME IN SYSTEM
Total Time in System
Q1 + T + R
LFE
L + Q2 + T
FEs and kits
Time
0
2
4
6
8
600000
650000
700000
750000
800000
Budget
Data Set u22

Figure [3]

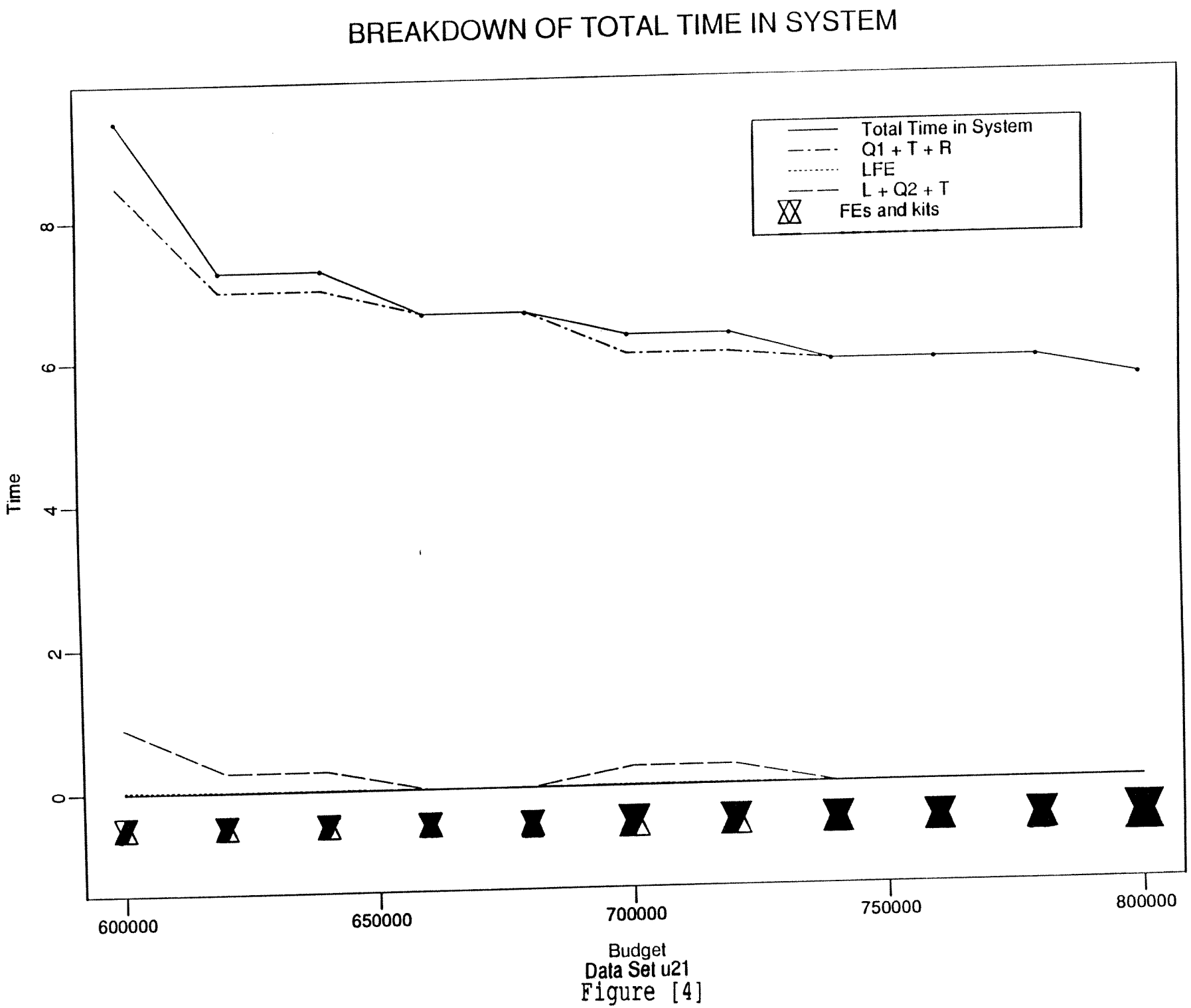

BREAKDOWN OF TOTAL TIME IN SYSTEM
Total Time in System
Q1 + T + R
LFE
L + Q2 + T
FEs and kits
Time
0
2
4
6
8
600000
650000
700000
750000
800000
Budget
Data Set u21

Figure [4]

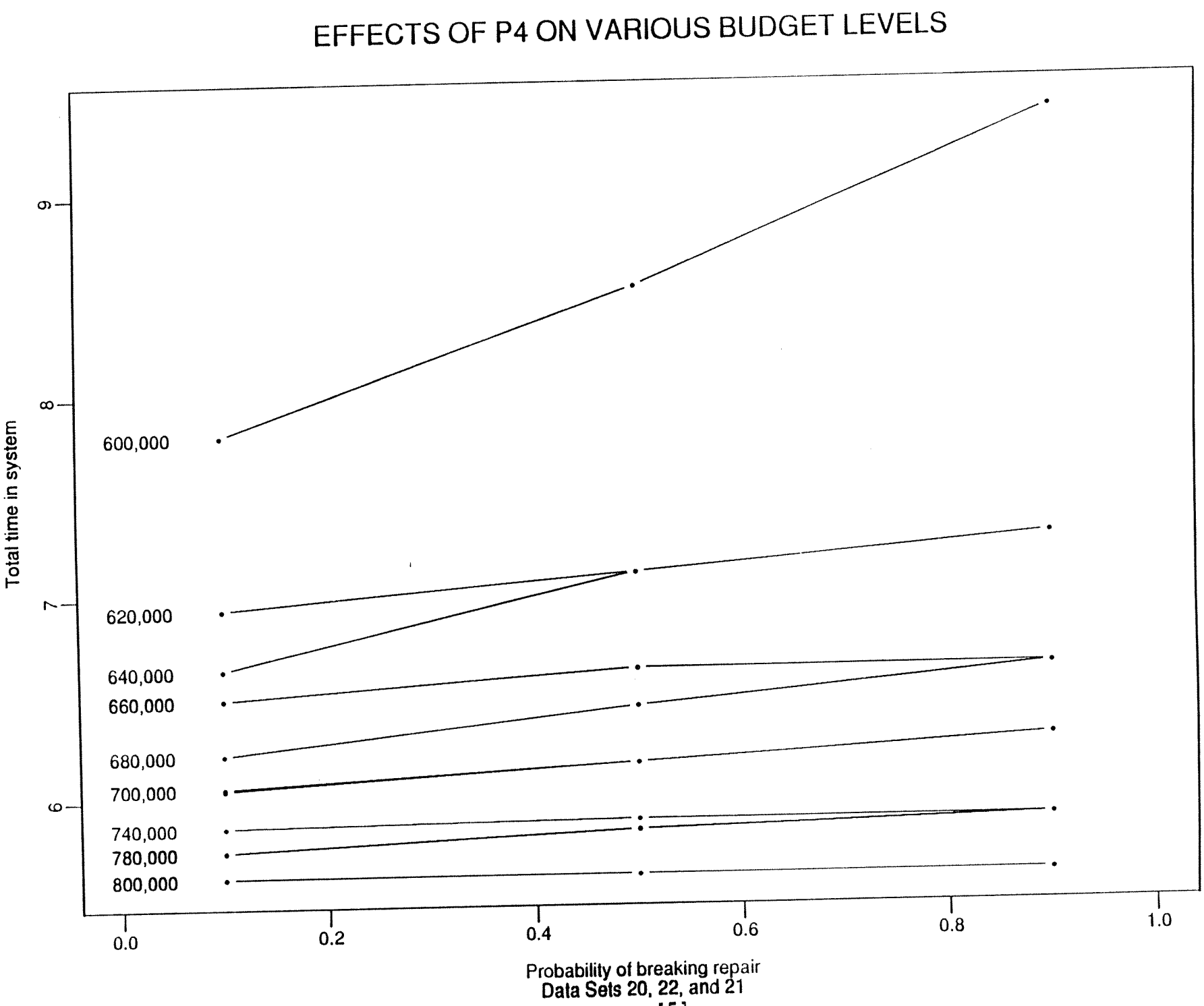

EFFECTS OF P4 ON VARIOUS BUDGET LEVELS
Total time in system
9
8
7
6
600,000
620,000
640,000
660,000
680,000
700,000
740,000
780,000
800,000
0.0
0.2
0.4
0.6
0.8
1.0
Probability of breaking repair
Data Sets 20, 22, and 21

Figure [5]

References

Agnihothri, Saligrama R., 1988, "A Mean Value Analysis of the Traveling Repairman Problem". IIE Transactions, Vol 20, No 2, 223-229.

Agnihothri, Saligrama R. and Uday S. Karmarkar, 1986, "Performance Evaluation of Service Territories". Working Paper Series No 86-107, School of Management, SUNY Center at Binghamton.

Avi-Itzhak, B. and P. Naor, 1963, "Some Queueing Problems with the Service Station Subject to Breakdown". Operations Research, Vol 11, No 3, 303-311.

Blumberg, Donald F., 1981, "Management System for Field Service Productivity Improvement". OMEGA, Vol 9, No 4, 419-428.

Carpenito, Thomas A. and John A. White,1976, "The Allocation of Non-identical Machines Among Non-identical Servers". International Journal of Production Research, Vol 14, No 4, 429-436.

Cohen, Morris, Paul Kleindorfer, and Hau Leung Lee, "Optimal Stocking Policies for Low Usage Items in Multi-Echelon Inventory Systems". Naval Research Logistics Quarterly, Vol 33, 17-38.

Elsayed, E. A., 1981, "An Optimum Repair Policy for the Machine Interference Problem". Journal of the Operational Research Society, Vol 13, 793-801.

Evers, W. H. and S. S. Thaker, 1969, " Programmed Automatic Customer Engineer (PACE) Dispatch". IBM Journal of Research and Development, Vol 13, 357-365.

Graves, Stephen, 1982, "A Multiple-Item Inventory Model with a Job Completion Criterion". Management Science, Vol 28, No 11, 1334.

Hambleton, R. S., 1982, "A Manpower Planning Model for Mobile Repairmen". Journal of the Operational Research Society, Vol 33, 621-627.

Kohlas, J. and J. Pasquier, "Optimization of Spare Parts for Hierarchically Decomposable Systems". European Journal of Operations Research, Vol 8, No 3, 294.

Mamer, J. and S. Smith, 1982, "Optimizing Field Repair Kits Based on Job Completion Rate". Management Science, Vol 28, No 11, 1328.

Mamer, J. and S. Smith, 1985, "Job Completion Based Invention Systems: Optimal Policies for Repair Kits and Spare Machines". Management Science, Vol 31, No 6, 703.

Miller, J. and W. L. Berry, 1974, "Heuristic Methods for Assigning Men to Machines". AIIE Transactions, 97.

Reynolds, B. H., 1975, "An M/M/m/n Queue for the Shortest Distance Priority Machine Interference Problem". Operations Research, Vol 23, No 2, 325.

Scholl, M. and L. Kleinrock, 1983, "On the M/G/1 Queue with Rest Periods and Certain Service-Independent Queueing Disciplines". Operations Research, Vol 31, No 4, 705.

Smith, S., 1979, "Estimating Service Territory Size". Management Science, Vol 25, No 4, 301-311.

Smith, S., 1980, "Optimal Inventories Based on Job Completion Rate for Repairs Requiring Multiple Items". Management Science, Vol 26, No 8.