CONTINUOUS PRODUCTION AND EMERGENT DEMAND

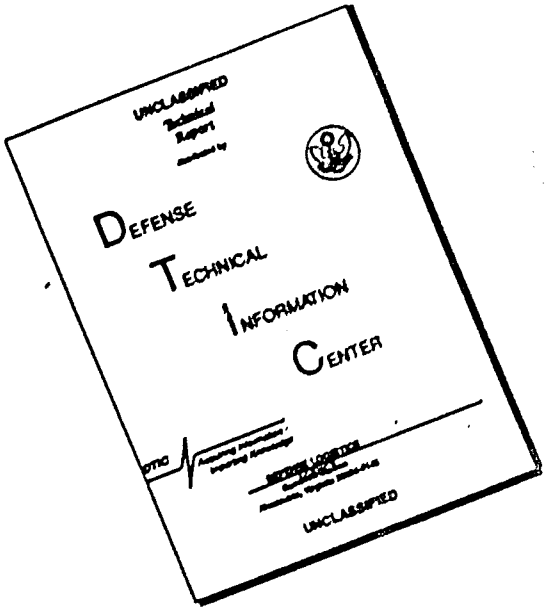
T. A. Goldman

P-1010

9 July 1957

The RAND Corporation
1700 MAIN ST • SANTA MONICA • CALIFORNIA

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

Summary

The manufacturer of machinery or other equipment will usually undertake to supply spare parts for such equipment to his customers. The demand for a spare part is a random event in time, which may be described by a demand probability function. If holding costs for spare parts are large, it may be desirable to avoid maintaining an inventory of such parts. One alternative is to supply spare parts, on demand, from the manufacturer's current production of parts for the assembly of new equipment, colloquially referred to as "robbing the production line."

A model of the real and monetary costs involved in supplying spare parts from current production is presented. The model leads to a cost function for each part individually depending on its production characteristics and its demand probability function, and on certain policy variables. By choosing appropriate values for the policy variables, the cost function for each part can be minimized independently of the others.

CONTINUOUS PRODUCTION AND EMERGENT DEMAND

The provisioning of spare parts by the Air Force for a new aircraft has many of the elements of a gamble about it. The record of Air Force procurement indicates that it is one of the most expensive games of chance ever devised. Outside observers are apt to blame bureaucratic inefficiency for the grandiose surpluses and dangerous shortages which usually arise, but it seems probable that this is not really a major factor in the situation. It is simply that the Air Force is constantly having to learn about stochastic processes the hard way.

One of the principal sources of difficulty has always been the necessity of placing a firm order for spare parts, in particular the most expensive ones, over a year before the first aircraft is delivered. This requirement arises from the well-known phenomenon of lead time, composed of administrative lead time and production lead time. It has long been realized that the spare parts used to repair an aircraft, or any other equipment, are in general identical with the parts used in producing such equipment on the assembly line. Where there is divergence, it usually appears at a relatively late stage in the manufacturing process. Hence the possibility arises of by-passing the lead-time requirement by diverting parts from the production line to meet the needs of maintenance in the field. Such procedures have been utilized by aircraft manufacturers to provide parts not only for military but also, and perhaps even more often, for commercial customers, in an emergency, but these procedures

have traditionally been regarded as a rather sneaky and undesirable way of getting spare parts. There are, in fact, a number of very good reasons for this attitude, but I shall not go into them right now. The characteristic phrase used to describe the procedure calls it "robbing the production line."

As a part of our research at the Rand Corporation on Air Force logistics, it occurred to us to inquire whether it might not be practicable to systematize some such procedure for supplying spare parts directly from the factory, in view of the apparently critical influence of the long lead times on the cost of provisioning a new weapons system.

One of the first questions that arose in investigating this problem was: "How much would it cost?" The answer turned out, not too surprisingly, to be: "That depends." We naturally asked, "What does it depend on, and in what way?" Such a question leads one right into a cost function. To construct a cost function, we needed a model of the way in which costs could be expected to arise, under such a system for providing spare parts out of current production.

At this point we narrowed the problem down considerably, in order to make it more tractable, by taking note of some of the fundamental facts of the production process. There is a basic dichotomy, with which many of you may be familiar, between parts manufactured in lots or batches and parts manufactured individually. The latter we found to be typically the major assemblies and sub-assemblies, and hence the most expensive items from the standpoint of spare-parts procurement.

Parts which are manufactured in lots are usually scheduled in such a way that the lots overlap in time, or they can readily be so scheduled. This provides a carry-over in the stocks, so that there is a constant minimum inventory on hand. Here we face problems of inventory control and optimum lot

size, which are certainly not unfamiliar and to which I have nothing to contribute at the moment.

The crux of the problem in which we were interested turned out to lie in the area of continuous production, that is, of those parts which are produced one at a time, and thus one after the other. The most expensive and most complicated parts, taking the longest time to manufacture, were found to belong to this category. Here the costs might be the highest, but the potential savings would also be highest.

Because these parts consist of complex assemblies of simpler components, they almost always can be, and of course are, repaired and returned to the spare-parts inventory if they fail or are damaged in any way. If the "reparable carcass," as it is called, that is, the part which has failed or been damaged, is replaced immediately by a part taken from current production at the manufacturer's plant, then the former part, after it has been repaired, constitutes, de facto, a one-item inventory of spare parts. The next demand which might arise for the same part could be met from the "inventory" thus provided, thereby relieving the manufacturer of the responsibility for meeting this second demand from current production.

The repaired part could have, incidentally, a further potential value as a backstop to the production line itself, in the occasional instance where a part is damaged in the factory and a replacement is not immediately available. This consideration was not, however, integral to our analysis.

A further simplification of the analysis resulted from the consideration that the system of supplying spare parts directly from the factory as an alternative to stocking them in the Air Force inventory would not, for a given model, be continued indefinitely with respect to the bulk of the parts. The

only exceptions would presumably be those parts classified as "not logical spares," for one reason or another, which are therefore not ordinarily procured by the Air Force in any case. Every other part would presumably be bought by the Air Force for spare-parts inventory at some time, either sooner or later. The question to be answered thus became, instead of <u>whether</u> to supply spare parts directly to the field out of current production at the factory, <u>how long</u> to do so. The procedure might, in the extreme case, be utilized over a time interval of zero length, meaning that <u>some</u> quantity of a particular spare part would be procured for inventory from the very beginning, as under current procedures. This case would apply in particular to parts for which a relatively high average demand rate could confidently be predicted.

Taking the time horizon as finite and variable in this way, and allowing it to terminate either on the occurrence of the first demand for each part or, no demand having occurred, at the discretion of the decision-maker made possible the construction of a relatively simple cost function.

In examining the process of continuous production, we concluded that a crucial element is the way in which each part is scheduled into the assembly process with respect to the particular aircraft of which it is to constitute a component. Obviously, the manufacture of the part itself must be completed before it can be assembled into the end item. The time which elapses between the completion of the part and its assembly into the end item is referred to by production personnel as the "cushion." The time which elapses between the assembly of a particular part into one end item and the assembly of the corresponding part into the next end item in the series may conveniently be referred to as the "production interval." There remains that part of the production interval which is not included in the cushion, and I have chosen to

call this the "forward gap."

In the system which I am attempting to describe, the assembly of a part into an end item at the factory is a scheduled event, whereas the occurrence of a demand from the field for a spare part is an unscheduled event. It is in fact a random event in time which we may refer to as t^0, with a probability distribution $F(t) = \text{Prob}(t^0 \leq t) = \int_{-\infty}^{t} f(t)dt$. Since $F(0) = 0$, we may disregard negative values of t in the following discussion. Because a demand from the field for a spare part has the nature of an emergency, in this framework, I have referred to "emergent demand" in my title (not without some feeling of word-play on the alternative definition of "emergent".)

It can thus be seen that the cost function will have a strongly stochastic element, depending as it must on the particular moment in time at which a demand from the field arises. I have mentioned briefly the basic elements of the model. The costs which can be incurred may be divided into two categories, which will be familiar to those acquainted with inventory theory. On the one hand are the costs incurred as a result of having an item when it is not needed, usually called "holding costs." On the other hand are the costs incurred as a result of needing an item and not having it. Terms such as "stock-out cost" or "depletion penalty," which are used in inventory theory, are not too descriptive in the present framework. It will be more suggestive to refer to these as "delay costs," the costs resulting from delay in meeting a requirement for a part.

An obvious but special feature of the system I am attempting to describe is that delay costs can be subdivided into two categories. If a demand arises in the field during what I have labelled the forward gap, when a completed part is not immediately available, delay costs of one type result. After a

part is sent from the factory to the field, that part is no longer available for assembly into the next end item on the production line, and delay costs of a second and characteristically different type are incurred at the factory.

In the cost function corresponding to this model, the length of each production interval is assumed given, although one production interval need not be the same as any other in length. The fundamental policy variable is the length of the cushion in any production interval. The entire set of cushion intervals may be thought of as a (policy) variable in vector form. Denote the length of the cushion interval within the i-th production interval as b_i, the length of the forward gap preceding that cushion, and thus within the same production interval, as a_i. We have $a_i + b_i = c_i$, where c_i is the length of the i-th production interval, and a vector $B \equiv (b_1, b_2, b_3, \ldots\ldots\ldots)$.

For an individual part, that is, a component of the end item being produced, some specific holding cost will be incurred during the i-th cushion, provided no maintenance demand from the field arises. Call the cost h_i. If no maintenance demand arises before the end of the k-th cushion, the total of holding costs for the part in question will be $H(k) = \sum_{i=1}^{k} h_i$. The increment in this cost element during an interval from t to $t + dt$ will be $H(t)dt$, a discontinuous function which has a positive (or at least non-negative) value on the cushions and is zero everywhere else. The expected value of this component of the cost of the system is: $E(H) = \int_0^\infty H(t) \left[1 - F(t)\right] dt$.

The cost resulting from delay in meeting a maintenance requirement for a part can be represented very simply as a function of the time at which the maintenance demand occurs. This is true *a fortiori* since we are only interested in the *first* maintenance demand for each part. It would still be a reasonable statement, however, if we were interested in succeeding demands as well. Let

this function be represented as $\phi(t^0)$, where t^0 is the time at which a maintenance demand occurs. The function takes on the value zero if t^0 lies within a cushion, and represents the cost of delay in meeting a maintenance requirement for a period from t^0 to the end of the forward gap, if t^0 lies within a forward gap.

The cost of keeping an end item waiting on the production line for a component during the i-th forward gap can be represented as g_i. If a maintenance demand occurs before the beginning of the k-th forward gap, the costs of this type will be, in the simplest case, the sum of the g_i over all forward gaps from k on: $G(k) = \sum_{i=k}^{\infty} g_i$. This assumes that the length of each forward gap is fixed once and for all by the original choice of a vector B of cushion intervals. As we shall see, these costs can in practice be reduced by the use of expediting procedures. The increment in G(k) during an interval from t to $t+dt$ can be represented by the function $g(t)dt$, which is zero over each cushion interval and has some non-negative value over each forward gap. Hence $G(t^0) = \int_{t^0}^{\infty} g(t)dt$.

In this form it is possible to combine the costs of the second and third types, since both depend on the occurrence of a maintenance demand in contrast to the holding costs, which depend on its non-occurrence, both result from non-availability of the part demanded, and hence both are zero over each cushion intervals. Call the combined function $W(t^0) = \int_{t^0}^{\infty} w(t)dt$, where w(t) is zero over each cushion interval, has the value $d\phi/dt$ from t^0 to the beginning of the next cushion interval, if t^0 lies within a forward gap, and the value g(t) over each forward gap which does not include t^0. The expected value of W(t) is $E(W) = \int_0^{\infty} W(t)f(t)dt$. To prevent possible confusion, it should be noted that this implies a double integration, and hence permits two alternative

verbalisations of this element of the cost function, depending on which integration one conceives of as taking place first.

The average cost of the system as a whole, meeting maintenance requirements from production, will be the sum of the two average cost functions, i.e., $E(H)+E(W)$. If the production intervals, c_i, are taken as given, the total will depend only on the choice of the b_i, and on the probability function, and may therefore be regarded as a function of the vector B.

Note that if all b_i are zero, the costs of the first type are necessarily zero. If the a_i are all zero, i.e. $b_i = c_i$ for all i, the costs of the second and third types are zero. The possibility of making these two statements is the principal reason for presenting this somewhat oversimplified formulation of the model, in which no expediting is assumed to take place.

Inclusion of expediting as an element of the model has little effect on the conceptual framework, but adds an important policy variable. The purpose (and the effect) of expediting is in essence to reduce the length of the forward gaps after a maintenance demand from the field has arisen. For a_i we substitute $a'_i = a_i - r_i$. The costs of the third type mentioned above are then to be taken over the shorter intervals a'_i, while a further cost is introduced, that of the expediting policy adopted. The vector R, whose elements are the r_i, is a policy variable, which is also a function of the time at which a maintenance demand from the field arises. This is an obvious consequence of the fact that expediting action is only initiated after a maintenance demand arises. Equally important, however, is the consideration that the cost of an expediting policy will usually depend in practice on the time at which expediting starts. Indeed, the expediting policy and the resulting costs of expediting will almost always be found to be a continuous function of time, or at the very

least piecewise continuous, with a relatively small number of discontinuities.

Expediting not only reduces the costs resulting from delays in the production process over each forward gap, but makes it possible to eliminate the forward gap completely at some point, i.e. to reduce the a'_i to zero for i greater than some value i*. Without expediting, the forward gaps once established by choice of the vector B would, in principle, be irrevocably established out to infinity. By means of expediting, the forward gap can be closed and some desired cushion reestablished. Expediting costs will continue to be incurred, therefore, until this state is attained. The expected value of the expediting costs incurred, when an expediting policy is chosen in advance is simply: $E(R) = \int_0^\infty R(t)f(t)dt$, where R(t) is the total expediting cost incurred if a maintenance demand occurs at time t.

It would be not at all difficult to repeat the cost computation for the possibility of maintenance demands after the first. As I have pointed out, we did not feel it necessary to do this. Clearly a somewhat different function would have to be incorporated to take account of the possible occurrence of a second demand while expediting is being carried on and before the forward gap has been entirely closed or before the first reparable carcass has been restored to serviceable condition. These should be in the nature of second-order effects, and, in view of the low probability densities with which we were concerned, did not appear likely to affect the conclusions significantly.

If a second demand occurred after repair of the first reparable carcass, and this were followed by a third within a short interval, the former would be covered by the original repaired item, but the latter demand would have to be met from the factory. The probability for this event should be extremely low in any situation where the policy of covering maintenance demands from current

production would be worth considering. Mention should, however, be made of the so-called "backstop" policy, a special case of the general policy I have been discussing, in which a very small inventory of spare parts is maintained with the understanding that any additional requirements will be supplied from the production line when the need arises. Since there is presumably some probability that this need will never arise, the probability function is somewhat different than when the inventory is zero and the eventual occurrence of a demand can be regarded as theoretically certain.

In conclusion, what advantages does this cost function provide in analysing the problem here considered? In the first place, it brings out quite clearly the stochastic nature of the costs incurred by a policy of providing spare parts from current production. If the shape of the probability function is significantly different for different parts, as was in fact the case in our problem, the influence of this fact on the costs of the policy, which may be quite considerable, can readily be determined. A clear presentation of the stochastic nature of the problem can be of particular importance to the operations research worker, who is called on to present his analysis to specialists from such fields as production, accounting, and maintenance, in whose thinking probability considerations are not likely to be uppermost.

In the second place, the cost-function analysis segregates the several cost elements and facilitates consideration of practical techniques for keeping each of them at a low level, for example, with regard to the availability of fabricating capacity on short notice, or by stocking a buffer inventory of some raw materials. In the third place, this analysis focusses attention on the two policy variables of cushion intervals and expediting policies. In both cases, the individual elements of the vector are quite likely in practice

to be functionally interrelated. In the aircraft industry, the learning-curve effect tends to reduce the length of each successive production interval, with obvious consequences for the relationships of forward gap and cushion. The length of the cushion in one production interval is usually related in some simple way to the length of the cushion in preceding production intervals. The use of analytic functions in the cost function may well be justified under these conditions, with resulting simplification of the computations. A similar conclusion is usually warranted for the expediting vector; the cost of reducing a specified forward gap by any given amount is likely to depend on the amount by which preceding gaps were reduced. Moreover, the choice of cushion and expediting policies for one part (or one general category of parts) can conveniently be made independently of the decision with respect to other parts, or categories of parts, if the cost functions can be shown to be significantly different with respect to the policies adopted. In this way the average cost of the spare-parts policy can be minimized separately for each type of part involved. Note that the use of the average cost as the basis for the decision is here warranted on the assumption that the number of parts involved is relatively large and that costs can be expressed in a single homogeneous metric.

If, finally, the cost of the policy of providing spare parts from current production is minimized for each part separately, this cost can be compared at each point in time with the costs of buying and stocking an inventory of the part in the field and the alternatives evaluated. As pointed out earlier, the assumption was made that at some time it would be found preferable to make the transition to procurement for inventory of each part involved.