Annual Processing

There are certain jobs which fall due only once a year.

Between November and February the Cumulative Contributions file and the now updated Name and Address file will be processed together to produce a tape containing data for printing contribution statements. The statements (20 million) will be printed off-line on the Xeronic printer.

In March each year the Cumulative Contributions file and the History file will be processed to produce the Starting Contributions file for next year, and an updated History file. At the same time, printing tapes will be prepared for printing history schedules.

We decided against referring to the magnetic-tape History file to answer inquiries relating to back periods, simply because of the time factor involved. Initially we expected to store the printed schedules, but soon discovered that this would involve too much paper to store: someone has said 45 miles of it; this may or may not be true. The only way out of this fantastic storage problem is to microfilm the schedules as they are being printed, and a flow camera to attach to the printer is being developed for this purpose.

Conclusion

This, then, is the story of the way we plan to run this new work of Graduated Contributions Records. Much of it is still theoretical, and no doubt there will be changes of approach between now and April 1962.

I think we can claim that this application is the largest data-processing job yet tackled in the United Kingdom. Like all A.D.P. applications, it will be launched mainly because of the enthusiasm, and indeed devotion, of the implementation team now sweating it out at Newcastle.

We have, I believe, another claim to make. So far as I can see, this is the first time a completely new job has been put on A.D.P.; there is no existing procedure to fall back on if we are late with commissioning.

I think it is also true to say that this is the first time that the date of an A.D.P. installation has been fixed by Regulation laid before Parliament!

Problems of the Introduction of Large Scale Data Processing into the Royal Army Pay Corps

By L. D. Slater

This paper surveys the preparatory work on the transfer of the British Army Pay records to an IBM 705 computer system. It was presented at the Harrogate Conference of The British Computer Society on 7 July 1960, some three months before the main equipment is due to be delivered to the Royal Army Pay Corps from the manufacturers.

Introduction

At the beginning of a survey of the problems of the introduction of large scale data processing for a specific application, it is probably essential to define:

(*a*) the specific application to which the problems are related,
(*b*) the equipment to be used,
(*c*) the degree of experience on which comments are based,

and to spend some time on explaining the envisaged processing flow.

The Application

The application is that of maintaining pay accounts for soldiers of the British Army. This task is at present performed partially manually and partially on simple accounting machines. The application is PAY, but it is not PAYROLL. The requirements of the system are as follows:

(*a*) That individual soldiers' entitlements can be assessed on an as-required basis.
(*b*) That the drawings of individuals can be brought to account against entitlement. This is an arrears operation and it should be mentioned that there is no requirement for an individual to draw his entitlement in full.
(*c*) A running account and periodic statement of account shall be maintained and produced for each soldier. This is to be done monthly.
(*d*) Payments to dependants shall be made and brought to account where entitlement exists or where otherwise authorized. The bulk of such payments are, as a matter of interest, of a recurring nature.
(*e*) Certain analyses of issues can be produced for purposes connected with the Army Appropriation Account.

In addition, there is an overall requirement that any system adopted should be sufficiently flexible to be able to cope with rapid expansion in the event of an emergency.

The ultimate number of accounts to be absorbed into the system is dependent upon many considerations. The immediate target is, however, to create a master file of

something of the order of 150,000 accounts. The size of the individual-account record will be variable between a range of the order of 800 to 2,000 characters.

Daily average activity is assessed as being in the region of 10%, though this will vary between a maximum of around 20% and a minimum of around 3%. Activity is random in type of occurrence.

Initially, the intention is daily processing of the Master File.

The Equipment

The equipment on order is an IBM 705 Model II system.

The available storage within the computer will be:

(i) 40,000 individually addressable magnetic cores, as quick-access store;

(ii) an immediate backing store consisting of a magnetic drum of 60,000 characters. This is more correctly visualized as 300 addressable sections, each of 200 characters;

(iii) magnetic tapes providing additional storage.

The particular machine configuration which will be used allows for up to ten magnetic-tape units on line. This will cover programs, input data, master file, output tapes and some additional backing storage.

It is perhaps desirable to point out that whilst the machine configuration applicable allows for one drum and ten tape units on line, the 705 Model II system is capable of expansion of up to 30 drums and 100 tape units, though this may perhaps be hard to visualize.

A card reader will be used off-line, though the facility exists for this to be switched over to on-line use if required.

A 150 line per minute printer will be available and will be used off-line, though again the facility will exist for this printer to be switched to on-line. Conventional E.A.M. (punched-card) equipment will also be available.

The 705 system will be supported by a 1401 Model D system. This is basically intended as a high-speed printer (up to 600 lines a minute) but, in the configuration ordered, certain other facilities are provided. For instance, there are available a 4,000 character magnetic core memory and two magnetic-tape units on line.

The 705 system is scheduled for delivery and installation at the beginning of October 1960, with the 1401 system following during the spring of 1961.

The Degree of Experience on which Comments are based

Planning of an A.D.P. system has been developed over the last five years. Of these the first three were devoted largely to the feasibility study, evaluation of various systems, and the initial clearance of a number of broad policy matters. Over the last two years the detailed planning of conversion and of the operation of the proposed system has been and is being developed. The targets are: to commence the conversion of accounts within two months of installation of the equipment; to combine production running tests with parallel running; to achieve a master file of approximately 150,000 accounts within a period of two years, and to achieve this without any interruption in the facilities provided to the customers.

The team engaged on the detailed work of planning the application has grown from a handful of individuals in the initial stages to just over 100 at the present time. Of these approximately 30% are engaged on work directly associated with programming. Approximately 50% have at least a basic knowledge of programming the 705, and approximately 60% have at least a reasonable knowledge of punched-card equipment.

To date, something in the region of 35 hours machine time has been used in testing programs and this will be doubled before the date of installation. As a matter of interest it may be mentioned that program testing is carried out on the Continent at 705 installations with comparable machine configurations. The position regarding program testing to date is that the logical correctness of the programs required within two months of installation has been broadly proved, and that most other programs have been subject to initial tests and are developing favourably. Many individual subroutines, however, still remain to be tested.

Processing Work Flow and Size of Programs

It is perhaps convenient at this stage to give an indication of the processing work flow which is visualized, together with an estimate of the size of the various main programs.

Initial action must naturally be to transfer accounts from their current visible form to magnetic-tape record form. Action in relation to this conversion is the same, broadly speaking, as will be required to assemble a new account prior to insertion into the master file. The program is in two phases, with automatic transference into Phase II on completion of Phase I for all input data.

Phase I imposes edit checks on the content of individual input cards, assembly checks in relation to the deck of cards appropriate to an individual soldier's account, and feasibility checks on the account overall. Phase II is a print routine providing an interpretive print of the assembled account content. The two phases together make a large program using both core memory and drum.

The remainder of the programs being developed fall into two main categories, those associated with the daily actioning of occurrences against the Master File, and those associated with periodic processing of the Master File.

The main programs for the daily actioning of occurrences are:

An Editing Program

This edits input-card content; establishes accepted input control totals; builds up a control block over multi casualties for any particular individual—the main processing program being geared to single-occurrence input;

and provides the facility of introducing accounts or data from other computer processings into the main casualty stream or on to the Master File.

It is a program of about 6,000 instructions using the core memory only.

An Updating Program

This processes the input occurrences received from the editing run against the Master File. It updates the individual records within the Master File and produces any information which is required for visual interpretation.

The program is an exceptionally large one and uses the core memory, the drum and a subroutine magnetic tape. Overall, the program is made up of approximately 130 subroutines, of which some 60 are on the subroutine tape. In the processing of any one individual occurrence, something in the region of twenty-five of the subroutines will be used. One interesting feature of the program is the automatic selection of the appropriate subroutines, coupled with the continual switch of subroutines from the drum and the subroutine tape into and out of the core memory by a sizeable Master Program.

Other features which are perhaps worthy of note are:

(i) That the larger subroutines have been sectionalized so that they can be brought down into the core memory a section at a time, thus reducing the size of the areas which have to be reserved in the core memory for them;

(ii) That the subroutines on the subroutine tape have been grouped in accordance with their frequency of usage, and groups are repeated down the tape in accordance with that frequency. The size of each group is, incidentally, limited by the size of the reserved area within the memory into which it must be brought down. This does in fact mean that the maximum number of subroutines in any one group is four. By working in this manner it has been possible to reduce the access time to the most frequently used routines to no more than $1\frac{1}{2}$ seconds. The access time to the very infrequently used ones works out at about $9\frac{1}{2}$ seconds.

A Distribution Program

The main purposes of this program are to provide for economical use of the printers by grouping the information from the updating run by type of print format, and to save storage space in the updating run. It is also used to meet certain other subsidiary requirements including:

(i) The provision of flexibility whereby, for example, data taken from the Master File can be re-inserted in different sequence without removal from tape. Re-entry would be through the editing run.

(ii) The facility of interpreting any record, and of assessing an up-to-the-minute balance (should this be required).

Incorporating the subsidiary features it is a large program using both the core memory and the drum.

The programs for periodic processings cover such matters as credit of pay, statement production, production of the statutory annual notification of taxable emoluments and tax deducted. For the most part these are again rather largish programs, using both the core memory and a portion of the drum. They will normally incorporate subsidiary tasks such as specific scrutinies: for example, automatic increments and the like.

The Problem of Introducing such an Application

What are the problems of introducing a computer application of this nature?

In dealing with them it is proposed to discuss only the broad problems arising once the equipment has been selected. Feasibility studies and problems of system selection have been covered many times.

In considering the comment offered it must be borne in mind that the application being considered is a large-scale magnetic-tape application, and that the changeover being planned is virtually from a manual system to a computer.

Probably the first essentials were to realize that it was a complete system that was being planned, and that large-scale data processing is office mass-production. The initial concept of how the application was to be built up had therefore to aim at being all-embracing. It had to include: the outline revision of existing procedures to allow work to flow to and from the system equipment; the lines on which computer programs were to be developed, and on how they were to be integrated one with the other; and an indication of how control of processing was to be established. It also envisaged the requirement for the preparation of the many detailed procedure instructions and operation manuals which would be required. Beyond this the initial concept had to visualize how the changeover from the existing to the future system was to be effected.

For the whole development of the project a timed schedule was created. This was ambitious and challenging, but the aim was to make it possible of attainment.

The concept created, together with the schedule of work, then became the basis for training of staff so that individuals with the right qualifications became available as required.

Inevitably, as the project developed, many problems not originally seen in their true perspective were revealed. There has, therefore, had to be constant revision of the initial planning, though it is fair to say that the outline framework has not changed. A typical example of the failure to see the problem in its entirety was that the need for an editing run and a distribution run was not appreciated until a comparatively late stage. This sort of thing does react on such matters as the estimates of usage timings of the main equipment. Another matter which was not fully clear initially was the volume and variety of output. This does emphasize the necessity, in my opinion, of a full and detailed review of the output requirement as one of the first phases of development of any application.

Division of responsibility is a must in a large application, but it brings in its train quite large problems of co-ordination, control, and the dissemination of decisions and information. These problems are not eased by the various changes which occur both in army pay regulations and in the programming concept over the period of development.

A factor which is not always evident at the beginning of a project is the degree of inter-relationship which will develop between all aspects of the overall operation. The nearer to operation date, the greater is likely to be the ripple effect of any change, however minor. These effects will not be restricted only to the relationship between the various programs, but can completely alter, for example, card layouts and output forms already agreed.

This is perhaps a convenient point to emphasize the importance of coding in an application of this nature. Codes are not only required for the identification of information within the machine and within the records on the master files. They are required for input and output information (including error rejections) and for the operation of changeover (or conversion) to the new system. Again, these codes are all inter-related and must in most instances be common to all requirements, at least in certain significant digits. They must also have the quality of being simple and readily intelligible to the human being who must insert them on input and interpret them on output. Coding is a task which must be tackled early, even if only to lay down the principles on which the code is to be built up.

Program testing is the next area to which I would quickly refer. A completion of a successful machine test is, of course, the individual programmer's first goal. The planning of program testing is, however, a matter which must be considered quite early on in the operation. The successful build-up on target dates of overall programs and computer runs is dependent upon the availability of tested individual subroutines as they are required. Work priorities for individual subroutines must therefore be established early, in order that the completion sequence is correct. There are also decisions to be made in relation to the degree of checking to be applied, the responsibilities for such checking, and the test-data requirements. There are numerous administrative arrangements in connection with test sessions when these are carried out at installations other than your own. If in the planning of program testing one should set up a special numeric test or testing control section, it is here that you will collect a mass of vital statistics relating to constants used, work areas, size of subroutines—matters which are vital when the time comes to plan storage allocation.

Testing must not be restricted to programs. It is equally important that wherever possible all work procedures should be tested. The culmination of all testing is in the production-running tests, during which the whole linkage of the system will be proved. Even where this is to be done in conjunction with normal operations, it will still be necessary to carry out limited preparatory tests, using pre-prepared data. Limited data of this nature can easily take a small team several months to prepare.

Problems are not, unfortunately, restricted to the planning of a system. They also occur during its installation. The actual Processing Division should therefore be formed in skeleton some time before operations are due to commence. This can conveniently be done by giving to this division such problems as the detailed planning of:

(*a*) The processing control system.
(*b*) The magnetic-tape library and control organization.
(*c*) Machine-utilization statistical records.
(*d*) The provision of any cost-data statistics that may be envisaged.
(*e*) Emergency drills—fire, power surges or failures, ancillary plant breakdowns, etc.

In addition, the Processing Division must be closely concerned with the programmer in the writing of the tape-labelling routine and the machine HALT and ERROR conditions programs.

The problem areas mentioned are by no means exhaustive, no mention has, for instance, been made of the problems of card layout for input information; to form design for both source and output documents; to the problems and relative merits of using pre-printed stationery or generated headings. The solution to all problems, however, stems from certain basic requirements. These are:

(*a*) A sound flexible organization.
(*b*) A first-class system of inter-communication between all members of the planning team.
(*c*) Good documentation of problems considered, decisions taken, and of the build-up of the system in all aspects.

On the organization required for planning such a project the following outline is suggested:

(*a*) An overall policy-making committee.
(*b*) A manager with the widest possible delegated powers, responsible for seeing the operation through from initial concept to finality.
(*c*) Three main teams responsible to the manager and dealing with Programming, Systems and Procedure Planning, Processing.
(*d*) A training organization.

This paper has deliberately been very broad in its approach. Large-scale data processing presents no really new management problems. It is basically an organizational problem with a factory concept. Within this there is, of course, the question of where the equipment is to be located. Frequently a specially designed computer building will be required in order to achieve maximum efficiency. There are also questions of the comparable merits of centralization and decentralization, but these are still matters within any organizational problem.