Data Collection and Transmission

By E. P. G. Wright

The paper makes a proposal for evaluating the effectiveness of different error-detection and error-correction arrangements to be used for data-transmission systems. The proposal is based on measurements taken on national and international connections. The paper also describes some data-collection proposals which may be of interest to those seeking a very simple equipment.

Introduction

Not all users of data-transmission equipment will be interested in the same performance aspects of the systems which are appearing on the market. For applications in which the technical requirements are relatively simple the emphasis will assuredly be placed on cost; in other cases the users may judge a system almost wholly from the point of view of the effective message speed, due allowance being made for the handling of redundancy and the retransmission of those parts of the message which need to be corrected; and in yet other cases the most important factor may be the liability to fail to correct errors that occur as a consequence of interference during the process of transmission.

Several methods of correcting errors are being studied, and, although the amount of apparatus and the effective speed do not differ very much, the extent to which errors can be detected and corrected does vary greatly. These differences can be examined more closely by reference to Fig. 1 which shows an approximate comparison between four different error-correction arrangements in respect of

(1) the probability of undetected errors,
(2) the percentage of redundancy employed and the consequential effective message speed.*
(3) the amount of terminal apparatus involved.

A fourth parameter, which is not compared, is the ability with which a system, which is designed to correct without using repetition, can deal with real-time applications.

An accurate comparison of the ability of different redundancy codes to detect errors will naturally depend on a number of conditions such as the amount of redundancy used and the type of circuit which is employed to connect the sending station to the receiving station. It also needs to be recognized that, if retransmission of faulty blocks is employed, the size of the block influences the amount of repetition (see Fig. 2). The block size also has an influence on the liability to undetected errors.

It is clear from the differences shown in Fig. 1 that for system-design purposes it is advisable to find some method by which the effectiveness of different detection codes can be measured. It is also clear that, if it is desired to compare systems which can be expected to detect something like 99,990 out of 100,000 errors, the study will involve a large volume of statistics.

* The redundancy percentage is taken as $\frac{\text{redundancy} \times 100}{\text{data plus redundancy}}$

Comparison of Methods

In order to overcome the practical difficulty of undertaking measurements which would provide sufficient material for an accurate comparison, a two-process method is proposed. The first part would consist of a theoretical approximation which would assume a random distribution of "errors per block," and the second part would involve the determination of an adjustment factor which would depend on the ability of the detection code to recognize the actual distribution of errors rather than the assumed random distribution. By this method the amount of measurement necessary would be limited to that required to obtain the adjustment factors.

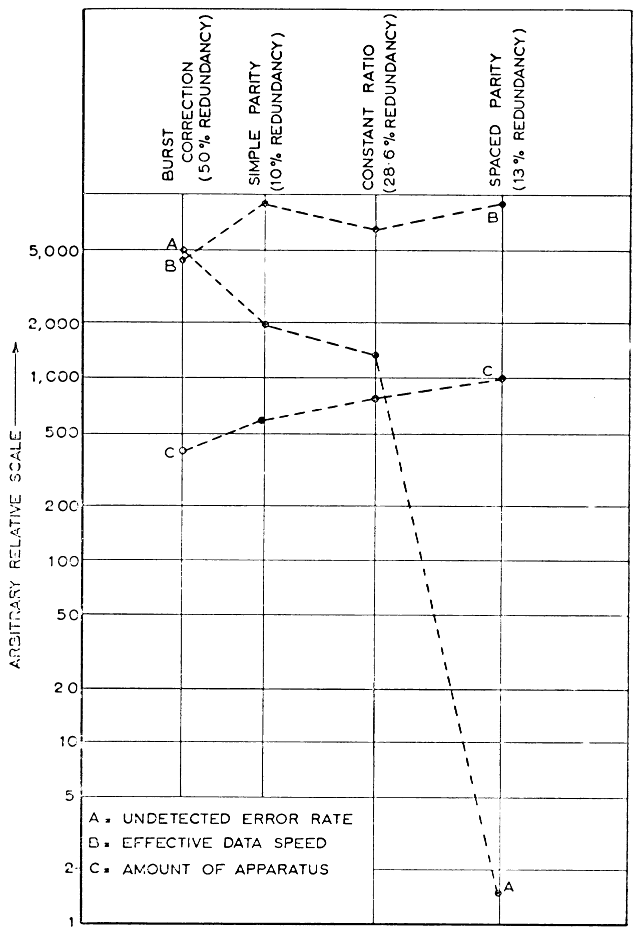

Fig. 1.—Comparison of error-correction systems (medium-speed data transmission)

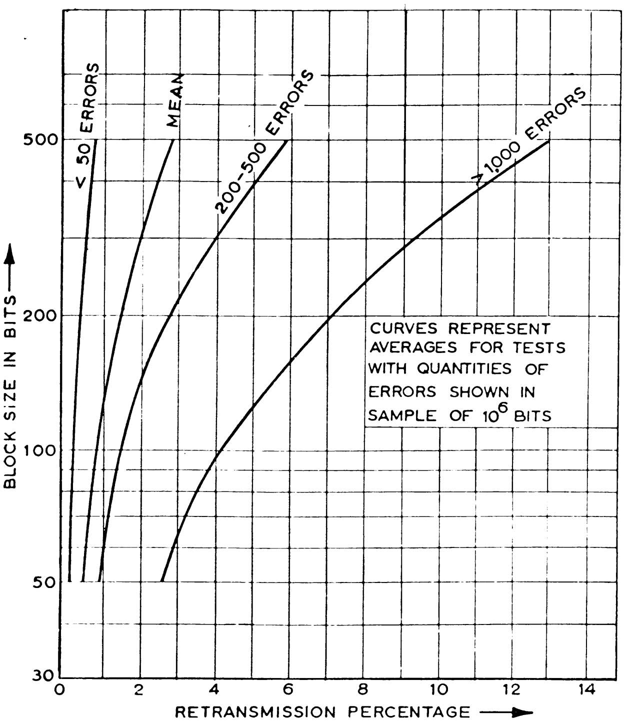

Fig. 2.—Percentage of blocks requiring retransmission (all levels). Note: Retransmission always involves two blocks—the block in error and the following block

It is known that for the simplest parity-detection codes the adjustment factor will indicate that the recognition of errors is much less effective for the measured distribution of actual errors than for a random distribution. On the other hand, it is believed that for certain spaced parity-detection codes the error recognition is better for the actual than for the random distribution.

Some idea of the extent of the difference between the measured distribution of errors and a random distribution is given by Fig. 3. In this diagram the numbers of bits in error are shown for a large series of tests, together with the probabilities for blocks of different sizes to contain errors for actual and random distributions. The random-error distribution shows a greater proportion of blocks in error than the measured distribution. This is explained by the fact that the average number of errors in the erroneous blocks differs for the two cases.

The theoretical ability of a code to detect errors depends to some extent on the number of "errors per block." Fig. 4 gives examples of theoretical detection values for parity arrangements employing single or double sets of regularly spaced parity words. The lines corresponding to each code show the values for different numbers of "errors per block."*

In practice it is not possible to evaluate the overall performance of a detection code without reference to

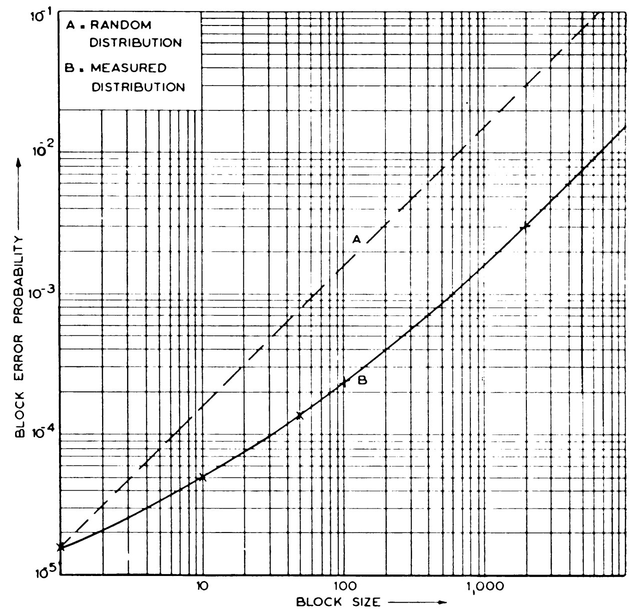

Fig. 3.—800-baud data transmission on leased circuits

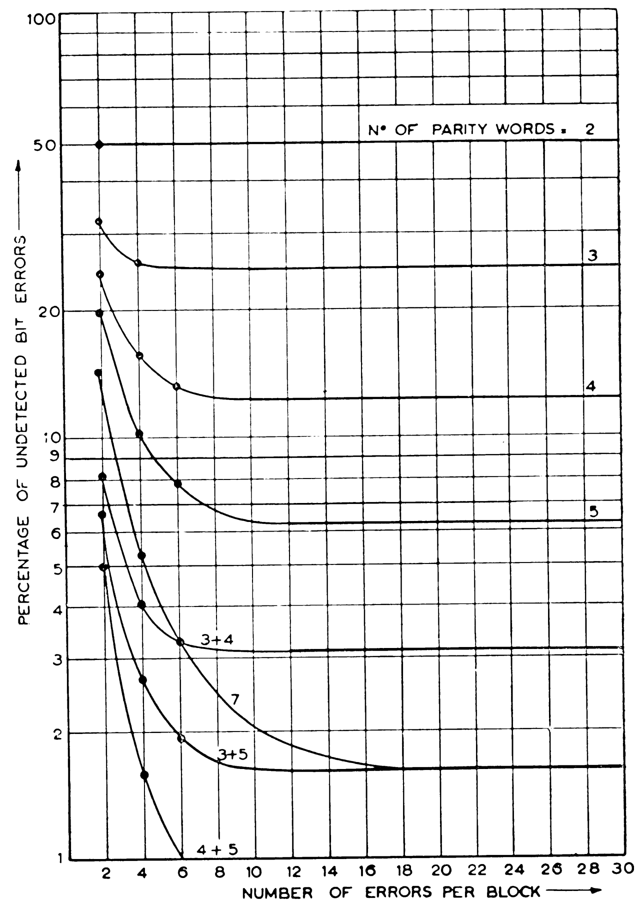

Fig. 4.—Undetected error percentage for large blocks with different numbers of spaced parity words (assuming random distribution of errors)

* In Fig. 4 it is assumed that the number of bits in the block exceeds the number of parity words with a single set (or the product of the numbers with a double set) and that the number of errors is less than the number of bits in the block.

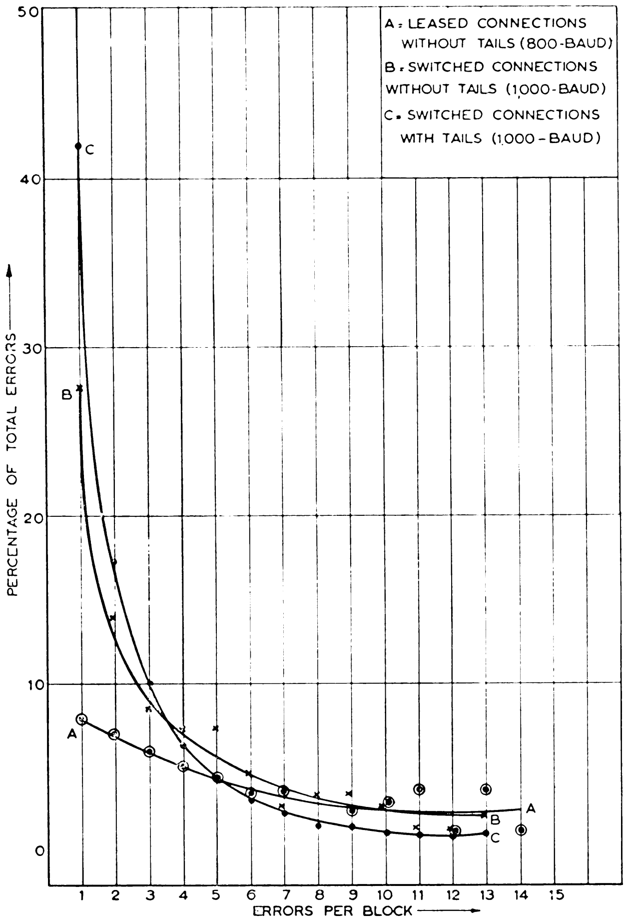

Fig. 5.—Error distribution for 50-bit blocks

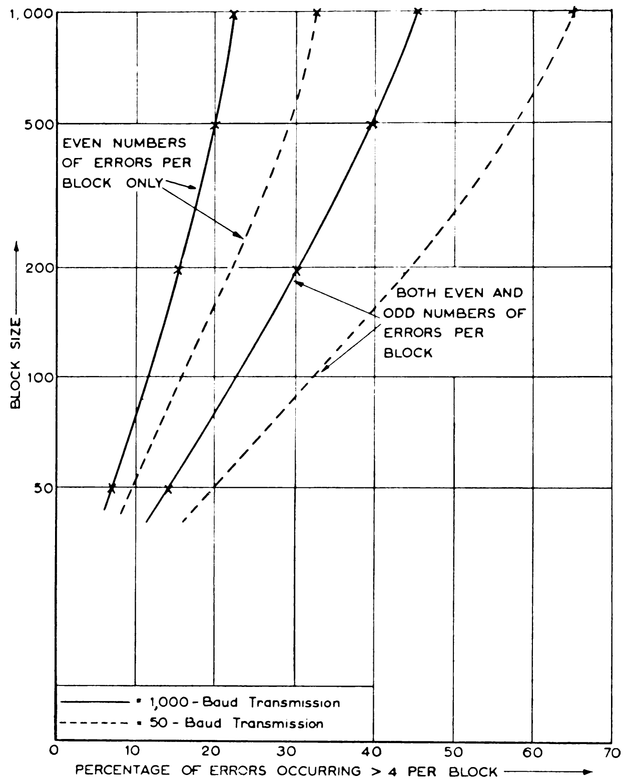

Fig. 6.—Percentage of errors occurring more than four per block for different block sizes and for transmission speeds of 1,000 and 50 bauds

some statistics in respect of the characteristic distribution of "errors per block." This is a characteristic which varies with the size of block and the conditions of transmission. A typical set of distribution curves is shown in Fig. 5 for blocks of 50 bits, while Fig. 6 shows some typical curves for all cases of more than four errors per block, for different speeds and block sizes.

As already stated, the process of evaluating the effectiveness of a detection code with actual errors is a formidable task in view of the very large amount of measurement and processing which becomes necessary. A reasonable approximation can be achieved, however, by sampling a few examples of different numbers of "errors per block" (e.g. two, four, and eight "errors per block") and comparing these values with the corresponding theoretical probabilities. If this work is undertaken for several block sizes, a large programme of measuring and processing is still involved.

Error characteristics for typical line interferences can be recorded by measuring the separation between errors and the correlation between adjacent separations. Fig. 7 shows a typical error-separation distribution. The separations are dependent on error rates and on the distribution of "errors per block." The block size has

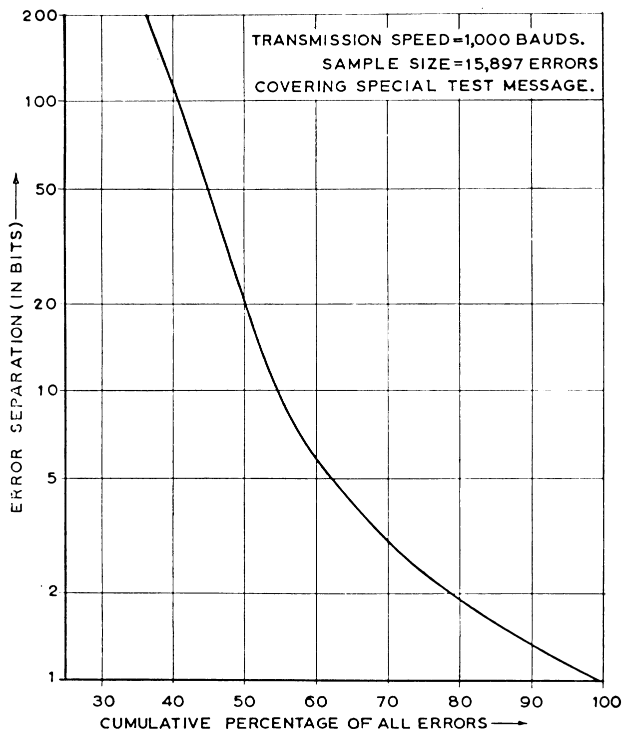

Fig. 7.—Distribution of error separations

no influence, but the pattern of errors is related to the message content.

Use of a Computer

A computer can be programmed to generate random numbers which can be used with a series of reference tables to determine the correlation factor and subsequently an error-separation value which will reproduce a succession of errors corresponding to those actually measured. A great advantage of such a computer program is that unwanted information such as odd numbers of "errors per block" can be disregarded at the time of generation so that the output result provides a concentrated record of spaced errors. This conservation of time and space is important because it is advisable to produce separate error tapes for each block size being studied.

A further computer program examines the artificial error tapes in conjunction with one or a number of error-detection codes and picks out the error patterns remaining undetected. With this information and the characteristic distribution of "errors per block" it is possible to determine the error-recognition ability of different detection codes. Fig. 8 compares some measured results with the corresponding theoretical values for blocks with random errors and a measured "errors per block" distribution.

There is evidence that the ability to recognize errors varies for the theoretical and the actual distributions according to the arrangement of the detection code. Some codes are much more effective with random errors than with actual errors, while other codes are more effective with actual errors; and this difference emphasizes the necessity for finding an accurate method of evaluation.

The proposal which has been described for the comparison of error-detection codes assumes that for many applications of data transmission there will be a demand for much more accuracy in the future. This assumption is based on the expectancy that data messages will frequently contain a large percentage of figures or codes which will not include the redundancy present in speech. Furthermore, it is to be expected that in some cases the daily routine transmissions will contain so much data that it will be impractical to provide operators or clerks to read through the incoming messages in order to search for errors. For transmissions over the switched telephone network it has to be expected that interference will be considerably greater than with the telegraph network, and it is likely to be much less expensive to use error correction than to attempt to reduce the sources of interference.

Data Collection

In telecommunication systems it has been the practice for many years for machines to enter into question-and-answer exchanges of information, and there is little doubt that such operations will become more and more extensive. The general concept of data transmission is

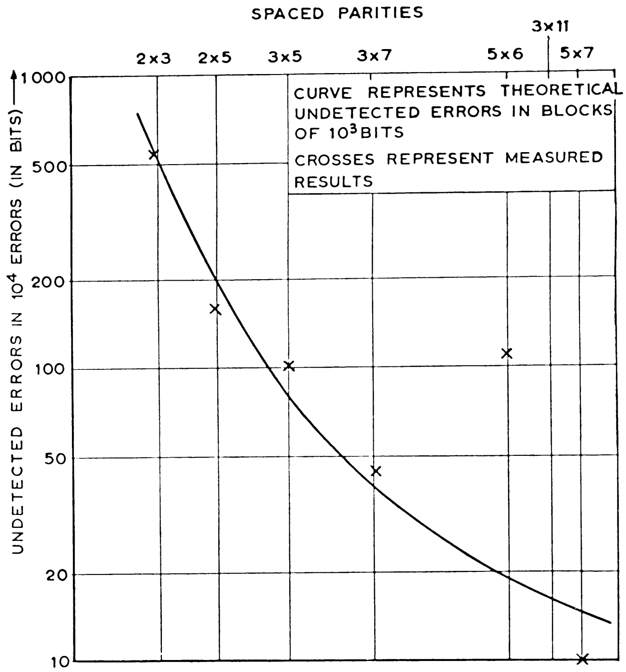

Fig. 8.—A comparison of theoretical and measured undetected error rates

one in which a sending station establishes a connection to some remote receiving station and, after going through the preliminaries, proceeds to dispatch the message, applying any corrections that are necessary. The expression "data collection" seems more appropriate than "data transmission" when it refers to a process involving a machine which sets up one or a sequence of connections and asks each station in turn to transmit any waiting message or messages—it might perhaps be a computer which has been programmed to obtain a subroutine from another computer, or even an off-course totalizator asking for the winner of the 3.30. The idea of such operations is not altogether far-fetched, because there are many advantages in having a centralized machine designed to collect waiting data by applying for it either at predetermined intervals or when previous messages have been processed, rather than having a machine designed to receive data whenever it is offered by any of a number of stations. A collection system can operate effectively over a switched telegraph or telephone network, and the expense is naturally less if the lines and terminal apparatus have already been installed for other telecommunication traffic.

Data-collection systems are not necessarily restricted to any one transmission rate. If the amount of traffic from one source is such that the transmission time would be excessive with a telegraph channel, it is possible to employ two or more telegraph channels or even a telephone channel. In any case the processing machine will require the messages to be suitably buffered so that it can pick them up at its own speed rather than at the line speed determined by the sending station. To provide for collection at both low and medium speeds the processing machine may be connected by several lines to both telegraph and telephone exchanges, enabling it to receive several messages simultaneously.

Having provided in some such manner for outstations which have too much traffic for a telegraph line, it is also desirable to provide for other outstations with too little traffic to justify a telegraph line. For such cases it is possible to provide much simpler outstation equipment, which can be connected to an ordinary telephone line. The data to be transmitted may be initiated by push-button (or even dial) operation, resulting at the receiving station in the production of a record which may subsequently be read by a machine. A simple form of error detection and correction can be provided by the use of speech-announcing apparatus at the receiving station to repeat back each section of the messages. The outstation is naturally given facilities for cancelling any block of information which is repeated back incorrectly. Data transmitted at such very low speed can be passed into the processing centre either directly or via a concentrator station.

The outstation portion of an equipment using low-speed, push-button operation can be demonstrated rather easily because it is quite a simple matter to establish a connection from any lecture hall to a remote receiving station by means of a normal telephone line. Although there is no intrinsic difficulty in constructing a single subscriber's set which could be used either for ordinary speech or for data transmission, it is possible to incorporate an additional apparatus unit for use in conjunction with the normal telephone. The demonstration about to be given employs such a unit having ten digit buttons and three auxiliary buttons signifying "BLOCK END," "AGREE" and "CANCEL." (See Fig. 9.) After the operation of the appropriate digit buttons, the BLOCK END button is pressed, and this causes the receiving station to announce over a loudspeaker, incorporated in the unit, the digits forming the block; no handset being necessary, the operator's hands are left free. If a repetition of the announcement is required, a second operation of the BLOCK END button will obtain this. Operation of the AGREE button causes a permanent record to be made at the receiving station, and operation of the CANCEL button results in the clearing of the receiver temporary store; in each case the receiver acknowledges these signals with a short burst of tone.

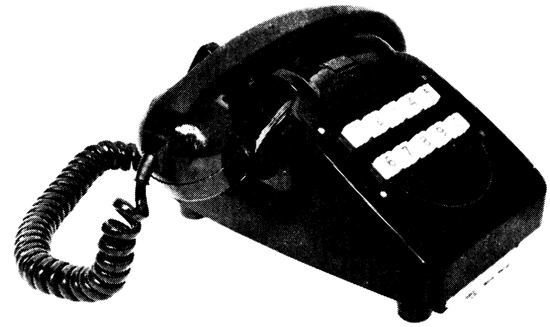

Fig. 9.—Outstation equipment

With additional buttons it would be possible to transmit alpha-numerical signals. The time taken to transmit four characters is about one second, and about the same time is needed for the check-back. No mains supply is necessary as the whole unit is powered from the exchange battery. The line signals consist of different combinations of two frequencies; such signals could, if desired, be generated from a card reader.

This process of passing intelligence from a person to a machine is analogous to the normal dialling process in telecommunications, but it also opens the way for the machine to talk to a subscriber by establishing a connection and transmitting a message from a vocabulary of pre-recorded words. The machine can ask the called party to operate push-buttons in order to provide a check-back or to obtain a repetition of the message. This allows the machine to make sure that the human being is in proper operating condition and responding intelligently.

Acknowledgements

The author wishes to thank the Engineer-in-Chief of the British Post Office for permission to use information concerning Post Office line plant.

Acknowledgement is also made to Standard Telephones and Cables Limited and to Standard Telecommunication Laboratories Limited for permission to publish this paper.

References

Croisdale, A. C. (1961). "Error Rates and Error Detection on Telegraph Circuits" (paper presented at the International Symposium on Data Transmission organized by the Benelux Section of the Institute of Radio Engineers at Delft in September 1960), *I.R.E. Transactions on Communications Systems*, Vol. CS-9, No. 1, March 1961, p. 28.

Enticknap, R. G. (1961). "Errors in Data Transmission Systems" (Delft Symposium paper), *I.R.E. Transactions on Communications Systems*, Vol. CS-9, No. 1, March 1961, p. 15.

Maguire, T. A., and Wright, E. P. G. (1961). "The Examination of Error Distributions for the Evaluation of Error-Detection and Error-Correction Procedures" (Delft Symposium paper), to be published in *I.R.E. Transactions on Communications Systems*, Vol. CS-9, No. 2, June 1961.

Wright, E. P. G. (1961). "Error Rates and Error Distributions on Data Transmitted over Switched Telephone Connections" (Delft Symposium paper), *I.R.E. Transactions on Communications Systems*, Vol. CS-9, No. 1, March 1961, p. 12.

Summary of Discussion

The Chairman (Dr. M. V. Wilkes, F.R.S.) asked Mr. Wright whether he foresaw a great use for the system of automatic special answering back. It seemed a very neat idea and might well have applications. Could the author add anything to what he had said about the type of application envisaged?

Mr. Wright replied that he saw some quite interesting possibilities in the system. It needed some development, no doubt, to make it suitable for all the possible applications. The basic idea was to have a very cheap outstation, a number of which could work in conjunction with one central station; the processing could then be carried out for a large number of small users who did not wish to bear, or could not afford, the expense of operating independent installations of their own, but who were willing to accept the reduced handling speed necessary with push-button or card-reading operations.

Mr. A. P. Clark (*British Telecommunications Research*), who recalled the author's diagram showing the effectiveness of various error-detection and -correction systems (Fig. 1), said that the fourth of these showed a marked increase in efficiency compared with the other three. Could Mr. Wright give some indication of the precise nature of this coding system?

Mr. Wright stated that a complete explanation of an error-detection system would occupy more time than was available. Very briefly, it could be said that the arrangement shown in the fourth column was one in which there were three sets of spaced (interleaved) parity words, with spacings of 3, 5 and 7, and that such an arrangement would therefore involve 15 parity bits per block. Apparently the block size could be increased to several hundred without much change to the undetected-error rate. There were other detection codes which gave similar results on a theoretical basis. The code mentioned was specially interesting as it gave better results with the actual distribution of errors than with a random distribution.

There was a general feeling that the number of occasions when more than two errors occurred in a block of 500 was quite low. On Fig. 5 it was shown that even for a block of 50 bits a considerable percentage of the errors occurred with more than two errors per block. This distribution made some of the detection codes more prone to undetected errors than would otherwise be the case.

The Chairman invited the author to enlarge upon the difference between a rented line and a switched line, in view of the point which he had made about there being a different error pattern.

Mr. Wright answered that it was an interesting fact that, although there was more security with a private telephone circuit than with a switched circuit, the errors appeared to be more in the form of bursts with the private circuit. The total number of errors on the private circuit was about one order less than that on the switched circuit, the difference being attributable to microphonic contacts, very short discharges, and the consequence of maintenance operations on the distributing frames and switch racks. Many of these faults produced independent errors which were separated from adjacent ones by more than 100 milliseconds.

In addition to these switching errors, there was little doubt that the switched connections also suffered from the bursts of errors experienced on the private circuits. Such faults would presumably be due to maintenance work on frames, to adverse weather conditions, or in some cases to the overloading of circuits. Such results on the switched connections were clearly swamped by the errors resulting from the switching operations.

Mr. C. C. Dilloway (*Ford Motor Company Ltd.*) said that, when speaking about the outstation equipment, the author had mentioned the possibility of this equipment being used in a situation where a number of equipments would be interrogated at the central station and each in turn would have the possibility of feeding information as the information became available.

Would it be possible to operate the equipment in the following way? At an appropriate time, the central station interrogated the outstations in turn and a list of information was "read out" to the outstation, this information being received at the outstation as voice signals. In answer to these interrogations, quantitative information, or data, could be transmitted back by a keyboard. When the block-end signal was given, the outstation would hear the indicative information and the numerical information just transmitted. That having been done, the data would be recorded at the central station as information suitable for processing and correctly transmitted. In this way, a considerable amount of detailed information of a repetitive nature could be collected fairly rapidly in a short space of time.

Secondly, Mr. Wright had mentioned that the equipment was quite inexpensive. Could he give an idea of "quite inexpensive"?

Mr. Wright replied that the type of application which Mr. Dilloway had proposed had been considered. There is a need for some simple record to be kept of the outstations which are found to be engaged, so that further attempts may be made. With magnetic tape storage the announcements can come direct from the store, whereas with paper tape storage it will be necessary to translate the telegraph language into speech which has been previously recorded.

In the demonstration which had been given, a connection had been established by dialling over the switched network to a central installation at Southgate. At the receiving station the signals were identified by voice-frequency telegraph detectors which responded, in combination, to the tones transmitted. The received data was stored on capacitors, and, when the block-end signal was received, the store was examined to enable relays to operate and control switching operations in order to select one of a number of tracks on a magnetic drum. Suitable speech recordings on the different tracks were therefore used to provide the check-back signals. The capacitors were re-charged when read in order to keep a record of the message to be used, after acceptance, for the preparation of the final punched-card or paper-tape record.

The equipment is experimental for the time being, and before placing it on the market it would be necessary to reach agreement with Administrations concerning the terms and conditions. Administrations have established rates for subscribers' lines which include the subset, the line and use of the central battery.

The additional apparatus included push-buttons and a loud-speaker, and these could be included in a box about the same size as the subset. There was no need for an additional line or an independent power supply. For the time being it would seem reasonable to assume that the outstation was likely to be about the same price as the subscriber's telephone instrument, but the price would vary to some extent with the number of push-buttons specified.

Mr. A. C. Croisdale (*Post Office*), who expressed his appreciation of the enormous amount of work done by Mr. Wright in preparing some of his "simple" curves, said that anyone who dealt with the business of analysis of data transmission errors immediately got bogged down in an enormous amount of work. To provide the curves shown by the author represented a tremendous job. It was an outstanding effort.

In Mr. Wright's Fig. 1, were the comparative results taken on actual line results?

Mr. Wright confirmed that the results were all taken from line measurements, but he was not able to say that all the results were drawn from exactly the same tests. Measurements had been proceeding for more than a year. As time went by, more and more statistics became available; for some of the studies all these statistics were needed, whereas for other studies a smaller sample seemed adequate.

In order to collect sufficient statistics to compare different error-detection codes which might be expected to have a very small undetected-error rate, it was impractical to depend on daily measurements as it would take too long to reach conclusions. By feeding into a computer a program which included information on the distribution of error separations, an output equal to several years of measurement could be obtained over one weekend. Even with a small computer quite a number of different error-detection codes could be checked simultaneously.

SABER: A Real Time Problem in Tele-Processing*

By K. S. Hope

SABER was the research code name given to IBM's Airline Reservation System project; it is now formally known as the IBM 9090 Airlines Reservation System. It should be emphasized that this is not a fixed configuration of machines other than a minimum power requirement of the computer. In fact any of the IBM "7000" range of computers can be used at the Data Processing Centre (D.P.C.).

After 1945, air passenger traffic built up at such a rate that some of the larger airlines were soon concerned at the increasing volume of records to be handled and controlled. Moreover, faster aircraft, with the consequent reduction in flight times, demanded up-to-the-minute and more accurate records to be maintained. With these problems in view IBM started a research project, the object of which was to combine all the latest techniques of telecommunications and computer design into a fully automatic reservation system.

System Requirements

One airline's requirements from such a system were as follows:

Inventory. This should cover details of the numbers of seats available in each class over each section of all flights, for every day that these flights are currently scheduled.

Availability. The first stage of seat booking is to find out if there are any available seats, possibly on a selection of flights. To examine each complete inventory record would be a lengthy and wasteful operation; therefore an abbreviated record should be maintained showing merely the availability of 1, 2, 3, 4, or more than 4 seats in each class, so that information on a number of similar flights could be retrieved and displayed together.

* An IBM trademark.

Fig. 1.—Agent's set

Flight Information. Details of departure times, delays, alterations, etc., should be immediately available to the agents on request.

Passenger Name Records. Full passenger details should be kept within the system, including passenger flight data, telephone numbers, connection information, etc.

Message Processing. Teletype connections should exist between the D.P.C. and other airlines for automatic transmission of messages.

Ticketing Data. A record of each transaction should be printed to provide for subsequent accounting purposes.

Communications. The system would employ seven pairs of high-speed telephone lines, five of which would be 1,500–2,000 miles long and the other two about 3,500 miles long.