THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 53, No. 1, January 1974
Printed in U.S.A.

Automatic Intercept System:

Administering the Intercept Data Base

By J. H. CARRAN, K. E. GREISEN, W. G. HALL, and D. J. WELLS

(Manuscript received October 5, 1971)

An Automatic Intercept System data base of up to a half-million changed or disconnected telephone numbers is updated, corrected, verified, abstracted, restructured, and restored through the actions of a collection of function-oriented subprograms. These subprograms run in the base-level main program loop under their own monitor which also controls interrupt-level accesses to the asychronous disc memory. The monitor together with the set of subprograms provides a file administration capability which responds to both machine stimuli, such as timed entries or trouble indications, and human requests initiated from teletypewriters.

I. INTRODUCTION

The Automatic Intercept System (AIS) assembles machine announcements for calls to telephone numbers which have been changed or disconnected. Such calls are switched to intercept trunks in many local offices connected to one Automatic Intercept Center (AIC). The dialed numbers are transmitted automatically to the AIC by local offices equipped to do so or by operators when local offices are not so equipped. The AIS also provides special handling for calls to numbers which have never been equipped and for calls to lines on which a trouble condition has been marked at the local office.[1]

The principal data base, containing as many as a half-million directory numbers, is stored in duplicated disc memory units.[2] Clerical personnel keep it current with additions, corrections, and deletions of numbers on intercept in all connecting offices.

A distinctive portion of the system program provides for these updating functions and for verifying, abstracting, restructuring, and restoring the data base while the AIC continues to process telephone calls. These actions, all referred to as file administration, respond to internal stimuli such as timed routine entries or trouble indications as well as to external requests. Interactive teletypewriter input/output (I/O) provides access for both clerical and maintenance requests through the No. 2 ESS processor[3] which is part of the AIC.

II. FILE ADMINISTRATION OPERATIONS

The information filed with each intercepted directory number in the data base includes a status code, a count of inquiries, and, if appropriate, a new number to which calls are referred by automatic announcements. The status and the new number are changed only by the administrative programs, which remove and insert the whole entry. The call count is incremented by hardware (to a maximum count of seven on each disc memory) every time an inquiry is made for the number.

Changes to the data base are made only by human intervention. An insertion is made when a number is disconnected, and a deletion when it is reassigned to an active line. Numbers in active service are not kept in the AIS file. Numbers which have never been in service are covered in the file as soon as the connecting central office is equipped to divert calls to the AIC, but may be noted in a single entry for a group of 100 or 1000 until individual assignments to active lines begin breaking up the group.

Most file administration functions are handled on a single-server basis; that is, only one action is undertaken at a time and additional overlapping requests are rejected. External requests are accepted through four different teletypewriter channels, three of which are intended primarily for various plant maintenance purposes. Only one, the file administration teletypewriter, is used for the routine clerical work of updating the data base. (See Fig. 1.)

2.1 Updating

A typical update message, though very brief and stylized, takes three seconds of teletypewriter transmission time for the order and an "OK" response plus a second or two of elapsed time for processing. The system is arranged to control a 10-character-per-second paper tape reader at the teletypewriter for batched clerical operation. An interface is under development which will also provide for an optional 2000-bit-per-second data link.

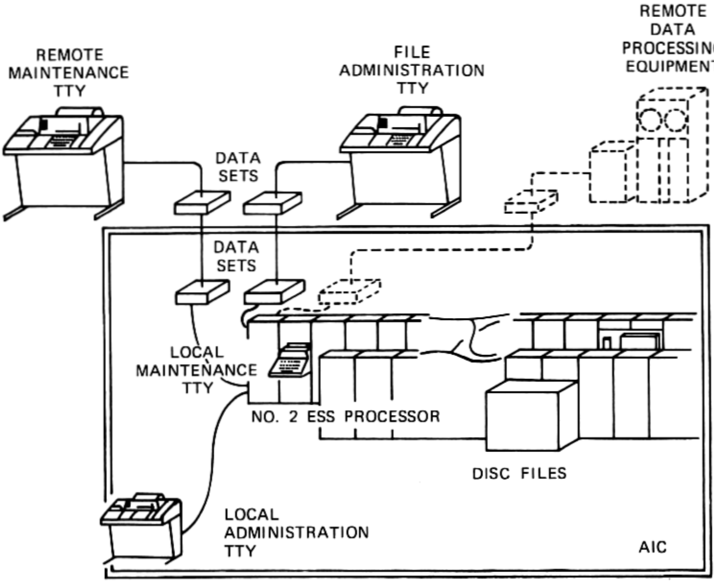


Fig. 1—External I/O channels for intercept data.

Routine updating transcribed by clerks from commercial service orders may average as many as 8000 original transactions per day, depending on the size of the data base and the mobility of line assignments in the telephone population served. Message originations and handling times are inflated by clerical errors and data inconsistencies, which result in retries and sometimes extensive response printing. In the presence of these anomalies, the system can handle 500 original transactions per hour. This requires efficient organization of the work flow feeding messages to the AIC and responses back to the clerks, as well as relegation of other file administration uses to separate hours.*

2.2 Other interactive uses

The other community of file administration users, the plant maintenance people, are called on occasionally to help the clerks correct any machine-related data anomalies. Plant also uses file administration functions to obtain data-related clues to machine troubles, using both interactive and internally stimulated messages. Hardware troubles

* A minicomputer-based File Administration System (FAS) is now under development which will assist the preparation of update messages and speed their flow, using the 2000-bit-per-second data link.

which automatically remove a disc memory and its controller from service preclude further routine updating, which must usually process both files in sequence. Thus, maintenance activity required to clear the trouble is preemptive. Call processing accesses continue in the duplicate file and the file administration monitor accesses the out-of-service file as requested by the diagnostic programs. In the event of a prolonged outage, a special condition can be instituted temporarily to permit updates to be done in one file only.

2.3 Out-of-hours activities

Some file administration functions are scheduled for light-hour operation. Routine tests of the validity of the intercept data are timed to start spontaneously at 1 a.m. each night. It takes from a quarter-hour to a half-hour to audit one-eighth of the file, plus time to print a record of all data anomalies found. When this is finished, the paper tape reader at the file administration teletypewriter is turned on by program. This provides a convenient means to obtain unmanned initiation of other actions in the middle of the night.

One routine job that lends itself to night turn-on is the abstracting of call count data from the file. This is done by printing directory numbers with counts less than a specified threshold for each of a desired list of central office designations, then resetting the counts to zero to start a new statistical period. Schedules for obtaining these statistics are set locally to provide data for reassignment of numbers to active lines. An hour of printing can provide call counts for about 1500 directory numbers.

2.4 Backup actions

Less frequently, a backup copy of the data base is created to guard against the remote possibility of loss of data from both of the duplicate disc memories. Depending on local practice, the backup may be maintained in an offline spare disc memory or in reels of paper tape or both. These are supplemented by paper tapes recording ensuing daily updating inputs. Eventually, other offline media will be accessed via the higher-speed data link. Copying the data base is done by the file administration programs while calling traffic is being handled, but must be scheduled to avoid routine updating work.

The file administration programs also provide for noninterfering system accesses to nongeneric office parameters[4] stored in the disc memory as a backup for call stores. This data base, three orders of magnitude smaller than the file of intercepted directory numbers, is maintained separately. System accesses to it have negligible effect on file availability for intercept record functions.

III. FILE ORGANIZATION

The AIS disc word has 42 usable bits, so that it is capable of storing ten binary-coded-decimal (BCD) digits. The formats used include:

(*i*) a seven-digit called number with two status digits and one call count digit,
(*ii*) a seven-digit referral number with three numbering plan area (NPA) digits, or
(*iii*) a four-digit or header number with associated machine address information.

Figure 2 shows the file organization.

The first two bits of the word form a tag which indicates which format has been used. A special pattern is recorded with a called number tag to mean that the location in which it resides is blank, i.e., available for storage of intercept record data. In a different context, the word can also be used in a free format, for example, for storing call store backup information on the special track reserved for that purpose.

The hardware environment has naturally induced two different means by which these numbers can be accessed and manipulated. On one hand, any number or group of numbers can be referred to on a machine address basis. That is, the user can retrieve a particular word, a block (18 or 20 words), an interlace[2] (16 blocks), or a track of data (5 interlaces totaling 1590 words). In addition, he has available operations which use many tracks up to and including the whole file (384 data tracks).

On the other hand, the implementation of a hardware search capability in the file complex[2] has induced a "thousands group" type of categorization. The clerical personnel can avail themselves of functions which operate on the data associated with one or more locator words. A locator contains four high-order digits of a telephone number, so that there are one thousand numbers which it can refer to. Each such group is called a thousands group.

The locator words on three specially accessed tracks act as a machine-address index to the intercepted directory numbers, so that the thousands groups form relocatable sets of data. Each thousands group is bounded by a header word and an end mark, and within this set of data all non-blank called number words are kept in ascending

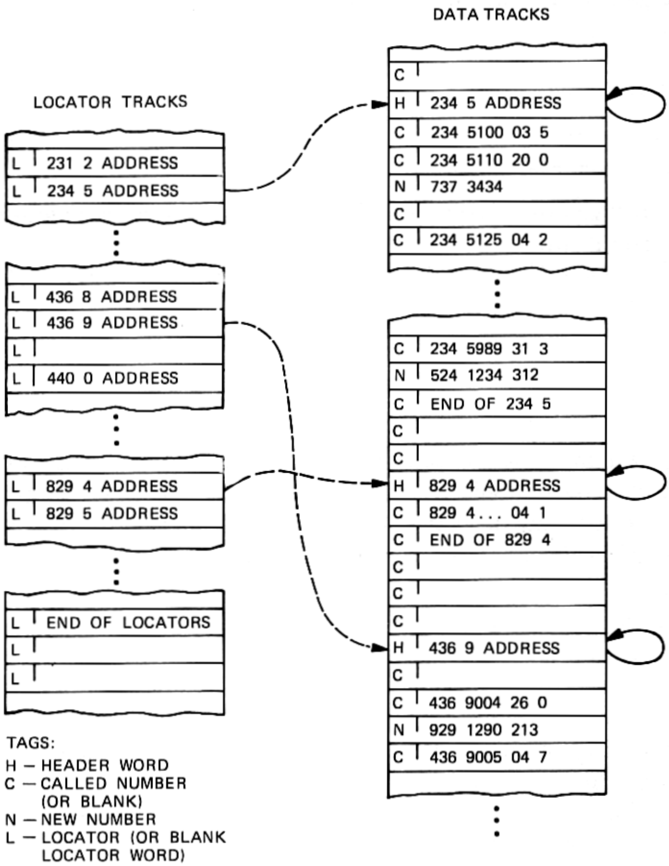


Fig. 2—File organization.

numerical sequence. New number (referral number) words are filed directly after the called numbers they pertain to. Blanks can be almost anywhere, though certain distributions are operationally desirable. Both the file administration programs and the hardware search features rely on these constraints.

IV. CONTROL PROGRAMS

4.1 File monitor

All file data handling except for call processing takes place under the auspices of the file administration monitor. This is basically a table-driven executive which is activated once each base-level loop.[4]

During each execution it updates an internal record of the state of the AIS file subsystems and either invokes an appropriate subprogram or, in case no work is to be done, returns control to the other base-level programs. This base-level execution may vary from a minimum on the order of a few hundred microseconds, when no work is to be done, to a maximum of about 3 milliseconds for the longest subprogram. In addition, the monitor must call for and manage a number of programs which operate in three interrupt levels, all higher than base level.

Every file administration function, including those executed on internal request, is defined by means of a table of addresses contained in program store. Each address is a transfer reference to a subprogram. A file administration subprogram is a program module designed to accomplish a simple operation. It is given control by the monitor, and when it has completed its task it returns control via one of several entry points to the monitor. Typically a subprogram can test or move data, request file I/O operations of either kind, or execute teletypewriter actions.

Under normal circumstances when a file function is in progress, the monitor accesses the function table once each base-level loop, and executes the associated subprogram. These table accesses are accomplished with the use of two call store words, one which indicates which function is in progress, and the other which is used as an index. Proper manipulation of these words achieves in a simple manner the ability to retrieve the subprogram addresses in consecutive sequence, to repeat a subprogram, to "branch" to other tables, or to loop on a sequence. It is also possible to execute a plurality of subprograms in one base-level execution. The choice, however, of performing loops, branches, etc., is made in the subprogram, and implemented through the use of the different entry points in a return to the monitor.

Requests for file administration activity come from several sources. They can be received from teletypewriter channels, from a 24-hour timer, from programs which audit and restore office parameter data, or from file diagnostic programs. The first two sources are referred to as data management sources; the same restrictions and priorities are applied to both. These data management requests are usually for relatively long and involved functions, while the call store audit and file diagnostic requests are always for highly specialized "single-shot" data transfers.

Nominally the monitor realizes three priority divisions established by the order it uses to examine request indicators. In order from first to last, they are requests from call store audit programs, from data management sources, and from file diagnostics.

Call store audit programs request only a transfer of a single block of office parameter data at one time. If such a request is received while another function of lower priority is in progress, the monitor suspends the function, processes the new request, and returns the previous function to an in-progress state. Neither data management nor file diagnostic requests can interrupt each other, e.g., if a file diagnostic request is in progress, it cannot be interrupted by a request from file administration.

4.2 File I/O

Two types of file I/O functions are available: lookups, which typically hold the file complex busy for 160 milliseconds, and block data transfers, which average 40 milliseconds. In file administration use, lookups are called on for the purpose of obtaining machine addresses, rather than referral information as in the call processing usage. The lookup I/O routines are capable of procuring the addresses of data words or of locator words on their respective tracks. The

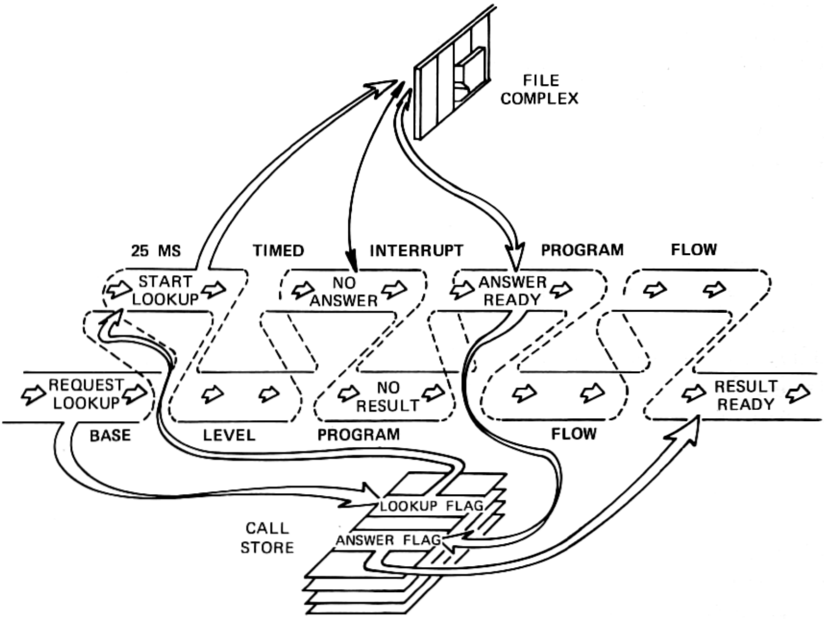


Fig. 3—Data lookups for file administration.

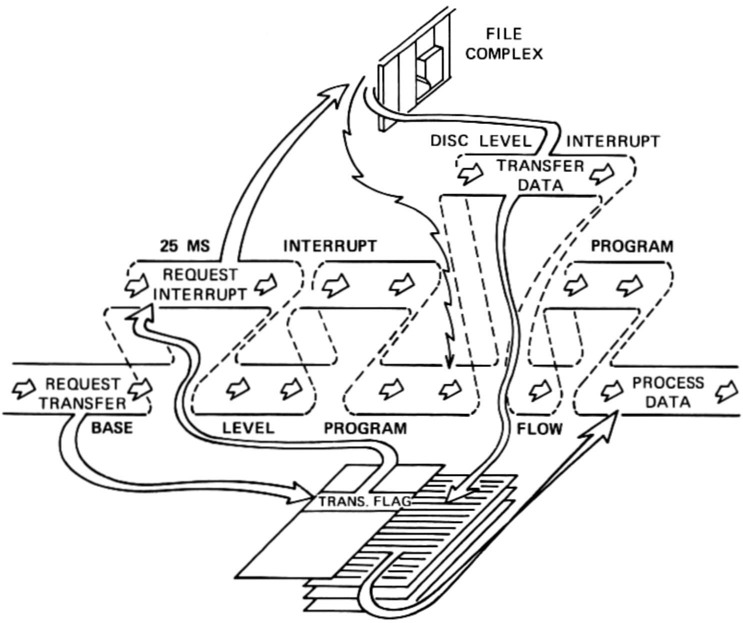


Fig. 4—Block transfers.

block transfer operation is used either to read a block of information from disc and load it in a call store buffer, or to do the reverse, i.e., write the file from call store. File usage for the most commonly requested functions entails a lookup followed by one or more block transfers.

Requests for I/O actions are executed in the same manner as data managing subprograms, in base level. All requests are recognized by a routine in the 25-millisecond timed interrupt level, and the order for action is then sent to the selected file control unit. (See Fig. 3.) If the request is for a lookup, another routine in the timed interrupt level scans the file status register in the file control[2] for an "answer ready" indication. When the indication is received, the address sought is retrieved from the track address register in the file control.

In the case of a block transfer, however, the file control generates an interrupt request in the processing unit when the proper disc memory address has been reached.[2] There are two disc interrupt levels which can be requested. Each is associated with only one file control, and both are of a higher level than that of the timed interrupt. (See Fig. 4.) The program affiliated with each disc interrupt level checks the validity of the interrupt request. If it is acceptable it transfers to a main I/O program which then scans registers in the file control until it is in close synchronism with the disc memory, and at this time it transfers the data in the desired direction.

The fact that the processor and the disc must be brought into synchronism for the duration of a data transfer introduces a restriction on file administration activity, i.e., only one file complex at a time can be used by the program in progress. Designing with this restriction (and applying it also to lookups) guarantees that a file will always be available to the call processing routines.

The high reliability of the AIS file subsystem[2] has warranted the use of an uncomplicated error-handling algorithm. If an I/O program detects a trouble indication and passes this result to the monitor, the I/O request is simply repeated. If no subsequent indication is detected on the second attempt, the function continues to progress normally. However, when the indication is repeated, the function in progress is aborted and teletypewriter output detailing the circumstances is generated. Basically, then, error handling consists of a single retry for all I/O functions.

4.3 File use under adverse conditions

The philosophy of changing data on the disc memories in normal circumstances is to modify one copy of the data base at a time. For example, when a number is added, a sequence of subprograms enters the number in one file complex, then the same insertion algorithm is repeated in entirety using the other file. This guarantees that if a file complex fails during either sequence, then at least one disc memory will contain completely valid data with or without the new entry, depending on how far execution proceeded.

If a function is aborted due to a persistent error, it will possibly produce a mismatch between disc memories: a number residing on one file complex and not on the other. In a case of this sort, manual intervention is required via teletypewriter-requested data maintenance functions (Section 5.1) to correct the condition. During such a mismatch condition, both call processing and file administration continue to function successfully. However, they will print messages calling attention to the anomaly whenever it is encountered. In addition, if neither the original abort message nor any subsequent mismatch messages gain attention to the anomaly, the nightly routine validity tests will also find and record it.

Consistent with the philosophy of redundant updating, an input message normally will be rejected if a file complex is unavailable to call processing at the time of request. There are, however, some special conditions which can be effected by input messages in order to continue processing intercept record changes in the face of file trouble.

One of these conditions directs the file administration executive to select the available file complex, use it, and not repeat the sequence on the unavailable file. This enables data management to proceed in the event of long-term outage of a file.

The other condition enables the use of an out-of-service file. A benefit of this feature is that a data base can be completely restructured on an out-of-service file. Operating in this manner protects call processing from incurring data errors and protects the restructuring from disturbance by maintenance activity.

V. FUNCTIONS

5.1 Interactive accesses

The most common file administration actions consist of deleting and inserting file entries. Both functions involve use of lookup I/O routines to locate the data block affected, then block transfer to obtain an image in call store, program manipulation of the image, and block transfer to write the revised block at its proper machine address in the disc memory. For a deletion, the revision consists of replacing the entry with blanks—one in place of the called number word, and one in place of an immediately following new number word if present. An insertion must be placed in numerical order. If one or two blanks are not present just before the next greater called number, they must be found elsewhere in the file. The preferred source is within the block already imaged in call store, but if necessary the data are rippled through successive sectors by transfer after transfer until blanks are found. The action is completed in both files before another request is accepted.

In case the action requested is inconsistent with data in file, it is rejected. Entries to be deleted must really be there, and numbers to be inserted must not already be in file. Other checks are made on existing file structure and hardware integrity with every action.

The call-count abstracting function uses a lookup to locate the start of a thousands group. Then the function obtains call store images of block after block of data which it scans for counts below the specified threshold. Other printing functions are also provided for administrative and maintenance purposes, some using lookups on data, some starting from given machine addresses. These print as little as one file entry or as much as all the data in one track of the disc memory. An ancillary function is available to print the machine address of a particular file entry.

Data maintenance functions use only block transfers to operate on one or both file complexes. Match and transfer actions, respectively, verify and produce the correspondence of blocks of data in the duplicated disc memories. Machine-addressed deletion and data writing functions provide for writing blanks or other specified bit patterns into any desired locations.

5.2 *Intercept data error checks*

Auditing functions check the intercept data for inconsistencies or violations of the file structure and coding scheme. Depending on the mode of initiation, either a single thousands group of numbers, a single track, or one-eighth of the intercept data base may be covered, in one file complex or both. The method of testing is to proceed word by word through the specified area in one file. If both files are to be checked, the corresponding area in the other file is tested by means of block comparisons with valid data in the initial file.

Each data word may be distinguished as to type by its two-bit tag and tested accordingly. The validity check function tests called number words for sequential order and verifies that the two-digit status code is one of an allowed set of values. This status code than serves as an indication of whether the called number should have an associated new number. It also indicates the format and type of information that should be contained in the new number NPA digits which are used by call processing programs to form the locality and NPA segments of the new number announcement.

In addition, tests are made that all telephone numbers consist of valid BCD digits, from 0 to 9 or from 2 to 9 as appropriate. Certain special codes are permitted, such as those used to represent groups of 100 or 1000 called numbers. Where group entries occur, checks are made that no individual entries occur within the range covered by the group entry.

Header and locator words for thousands groups should be in one-to-one correspondence and contain identical information. In addition, the machine address contained in each should point correctly to the location of the header word. Locators are tested for ascending sequential order, although the order of thousands groups on the data tracks need not be sequential.

Blank words are permitted to occur anywhere except between a called number and its associated new number. In addition, checks are made that areas of varying size between thousands groups contain only blank words.

An external initiation of validity checks may request that the tests cover a particular thousands group of numbers, or one-eighth of the data base. In the latter case, each subsequent request causes the next eighth to be tested. In both cases, either a particular file or both files may be tested. The automatic initiation of validity checks at one a.m. each day covers one-eighth of the intercept data in both file complexes so that all the data are audited in eight days. Timed and externally initiated eighths are incremented independently.

Internal indications of possible data anomalies, such as call processing detection of a header not-in-file, a mismatch between files, or a new number missing result in requests for file diagnostics. If all the hardware tests pass in these cases, validity checks are initiated in the area of the possible data error. These checks cover a single track on which the error should be located, and test that track in both file complexes if both are in service.

For all types of validity checks, records are printed of each individual error detected. At the end a summary record is printed including a total count of errors and an indication of the range of data tested. The information in the individual records may be used later to locate the errors and correct them by deletion of incorrect data and reinsertion of correct data. No automatic corrections are attempted on the intercept data because it is difficult or impossible to determine by program what the correct data is.

5.3 Recovery and restructuring

Recovery from a loss of intercept data involving one or both disc memories varies depending on the amount of data lost and whether the same data were lost from both file complexes. When the correct data are still present on one file they can be readily transferred internally via the No. 2 ESS processor to the other file. A maximum of about four hours is required if all the data in a four-disc unit must be transferred. In the unlikely event that both online copies are lost, even in part, then data must be entered externally, either individually using routine update messages or by using appropriate sections of the latest backup copy of the complete data base. If the backup copy is maintained on a spare disc memory unit, the backup unit can be temporarily placed on line allowing internal transfer of the desired data. If backup is in the form of paper tape, then recovery proceeds at about 10,000 entries per hour. Any recovery operation utilizing a backup copy of the data base must be supplemented by relevant update tapes generated since the last complete copy was made.

Should neither file be considered suitable for processing calls, all calls requesting file lookups are diverted to a simple "blank number" announcement. Because such a condition could conceivably persist over the potentially long time intervals involved with data base recovery operations, provisions are made to return a partially recovered file to service. When this is done the system is made temporarily insensitive to the expected flurry of not-in-file indications since their normal handling would cause flooding of the centralized intercept bureau operator positions.[1]

It may be found in normal operation of some AICs that occasionally a region of the file becomes so congested that insertion of new entries requires abnormally long times to complete all the data rippling required to maintain ordering within thousands groups. To relieve such congestion it is possible to do some restructuring of the data base. Since thousands groups are relocatable, restructuring can be done by extracting copies of one or more thousands groups from the congested region, removing their old image, and reinserting them into a less-congested region of the file.*

5.3.1 *Intercept data compression techniques*

Lengthy data transmissions from external sources are usually required for restructuring, recovering, or expanding the intercept data base. To reduce the time required for such extensive information exchanges, the intercept data are compressed so that fewer bits are required to record and subsequently reconstruct them than are stored on disc. Several characteristics of the intercept data format contribute to the applicability of a variety of data compression techniques.

One such characteristic is the existence of a significant amount of redundancy. In the section describing file organization it was noted that all entries belonging to the same thousands group are stored

* A recently completed automatic blank redistribution now accomplishes the desired restructuring. The FAS provides for a magnetic tape backup copy of the data base and an intermediate-speed recovery operation.

sequentially. In addition to such ordering, each entry includes its thousands group designation as part of the stored information. Since the thousands group designation can be inferred from the position of an entry in the data stream, it is not necessary to record it for each entry.

Another basic characteristic of the data base is that intercept number information is largely stored as BCD digits. The use of a BCD format, while convenient for many of the basic AIS functions, is inherently inefficient as a coding scheme. An alternative coding scheme having higher information content is used for compressing the disc data.

Throughout the data base, some blank disc words are usually interspersed for updating convenience. The exact placement of spare words is not critical, making it possible to omit them during data compression. Blank words may be inserted automatically during subsequent reconstruction if desired.

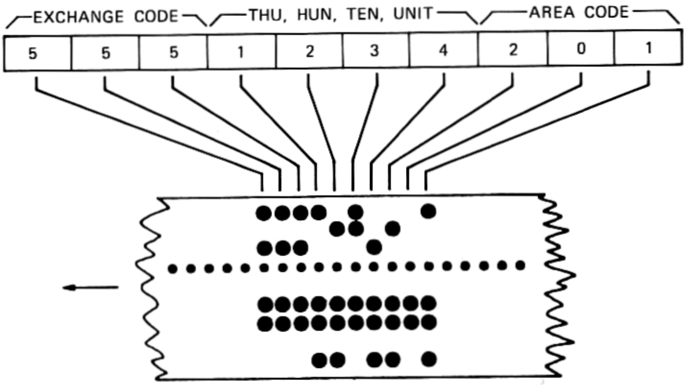
Still another source of data compression results from certain fields of each entry assuming predetermined values when being reconstructed from compressed code. These fields need not be recorded since they can be automatically initialized by a reconstruction program.

The implementation of encoding and redundancy omission for intercept data compression is described in greater detail in the next two sections.

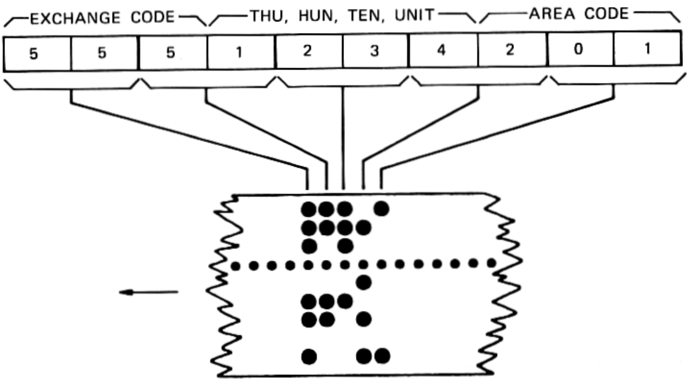
5.3.2 Encoding techniques

The conventional use of ASCII encodes *each* decimal digit in seven information bits and one optional parity bit. These seven or eight bits are then treated as one character. The coding scheme employed by the AIS data compression program converts each *pair* of BCD digits into a single seven-bit code which need not be equivalent to the decimal value of the BCD pair. This seven-bit code, plus an optional parity bit, is then treated as one character for transmission purposes. Compressed coding thus achieves a "two-for-one" reduction of character transmission compared to conventional use of ASCII coding. Figure 5 illustrates the difference in the codes as they would be used to punch a new number with its area code in paper tape.

A table lookup technique is employed both for converting pairs of BCD digits into single seven-bit characters and for expanding single seven-bit characters back into BCD pairs. The principal advantage of a table technique compared to conventional conversion algorithms is the ease of assigning arbitrary correspondences between BCD pairs



ASCII CODING OF 555-1234, AREA CODE 201



COMPRESSED CODING OF 555-1234, AREA CODE 201

Fig. 5—Compressed code vs ASCII coding.

and seven-bit codes. This feature is particularly useful in avoiding characters which are likely to result in undesired actions by the data processing facilities involved. An occurrence of an EOT character ("end of transmission") in the middle of a compressed code data stream, for example, could result in a premature disconnection since the receipt of an EOT character causes several types of commonly used data terminals to disconnect automatically. This and other troublesome characters indicated by the asterisks in the conversion table in Fig. 6 are avoided by the assignment scheme used.

Directly indexable tables for the expansion and compression functions tend to be relatively large. In AIS some memory saving techniques are employed to incorporate both tables within a single 128-word block of memory. To convert BCD pairs in the range 00–99 a

BCD PAIR	SYMBOL	DECIMAL EQUIVALENT
• 00	z	122
01	SOH	001
02	STX	002
03	ETX	003
• 04	v	118
• 05	u	117
06	ACK	006
• 07	t	116
08	BS	008
09	HT	009
10	LF	010
11	VT	011
12	FF	012
13	CR	013
14	SO	014
• 15	o	111
• 16	.	119
• 17	{	123
• 18	l	121
• 19	s	115
• 20	:	114
21	MAK	021
22	SYN	022
23	ETB	023
24	CAN	024
25	EM	025
26	SUB	026
27	ESC	027
28	FS	028
29	GS	029
30	RS	030
31	US	031
32	SP	032
• 33	q	113
34	"	034
35	#	035
36	$	036
37	%	037
• 38	p	112
39	'	039
40	(	040
41	)	041
42	*	042
43	+	043
44	'	044
45	–	045
46	.	046
• 47	x	120
48	0	048
49	1	049
50	2	050
51	3	051
52	4	052
53	5	053
54	6	054
55	7	055
56	8	056
57	9	057
58	:	058
59	;	059
60	<	060
61	=	061
62	>	062
63	?	063
64	.	064
65	A	065
66	B	066
67	C	067
68	D	068
69	E	069
70	F	070
71	G	071
72	H	072
73	I	073
74	J	074
75	K	075
76	L	076
77	M	077
78	N	078
79	O	079
80	P	080
81	Q	081
82	R	082
83	S	083
84	T	084
85	U	085
86	V	086
87	W	087
88	X	088
89	Y	089
90	Z	090
91	[	091
92	\	092
93	]	093
94	^	094
95	-	095
96		096
97	a	097
98	b	098
99	c	099
A0 (DELIMITER SERIES)	d	100
A1 (DELIMITER SERIES)	e	101
A2 (DELIMITER SERIES)	f	102
A3 (DELIMITER SERIES)	g	103
A4 (DELIMITER SERIES)	h	104
A5 (DELIMITER SERIES)	i	105
A6 (DELIMITER SERIES)	j	106
A7 (DELIMITER SERIES)	k	107
A8 (DELIMITER SERIES)	l	108
A9 (DELIMITER SERIES)	m	109

• INDICATES DEVIATIONS IN CONVERSION OF BCD PAIRS TO DECIMAL EQUIVALENTS.

NOTE: A = DECIMAL 10

Fig. 6—Compressed code conversion table.

directly indexable table of 154 words (nine blocks of 16 words each plus one block of 10 words) would be required. In AIS a delimiter digit equal to decimal ten is used which increases a directly indexable table size by one 16-word block, thus requiring 170 entries. The table size for expanding any seven-bit binary character to some BCD pair need only be 2^7 or 128 entries at most. In AIS the memory for storing tables consists of 22-bit words. This permits both the expansion and compression tables to share the same words by being assigned different bytes in each word. The combined table has the expansion table occupying the low-order eight bits and the compression table occupying the next higher seven bits. Just using shared memory alone would allow this combined table to fit into a single block of 170 words of memory. A further significant reduction in memory space for such a table is possible by using an "overlay" technique. The specific "overlay" structure for AIS is described below.

Whenever a BCD pair to be compressed is equal to or greater than BCD "80" (corresponding to the 128th entry), BCD "30" is subtracted and the compression table is indexed with the remainder. If the seven-bit binary number obtained from the table is added to the binary equivalent of decimal 30 the result will be a seven-bit binary equivalent for the original BCD pair. This kind of "overlay" is only possible if that portion of the table to be "overlaid" contains binary equivalents for the BCD ranges involved that differ only by a fixed constant. This happens to be the case for the conversion table in Fig. 6 since the portion of the table shared is BCD "50" to BCD "79" and all exceptions to straightforward decimal conversion lie below BCD "50." The fixed constant difference as noted earlier is the binary equivalent of 30 decimal. With the "overlay" technique the "two-for-one" conversion table can now be accommodated by a block of 128 words of memory.

5.3.3 Omitting redundancy

In addition to using the "two-for-one" coding scheme, the AIS achieves significant data reduction by omitting from compressed data certain types of the redundancy provided in the data base for ease of access. As noted previously, of the ten possible BCD digit fields for each disc word only seven digits are used to store the intercepted directory number. Of these seven BCD digits, four digits are used to designate the thousands group and the other three give the hundreds, tens, and units digits. The major source of redundancy is the repetition of the four digits designating the thousands group for each called number entry to be intercepted. Since all entries for a particular thousands group are placed sequentially on the file, the thousands group information need only be recorded once for each group.

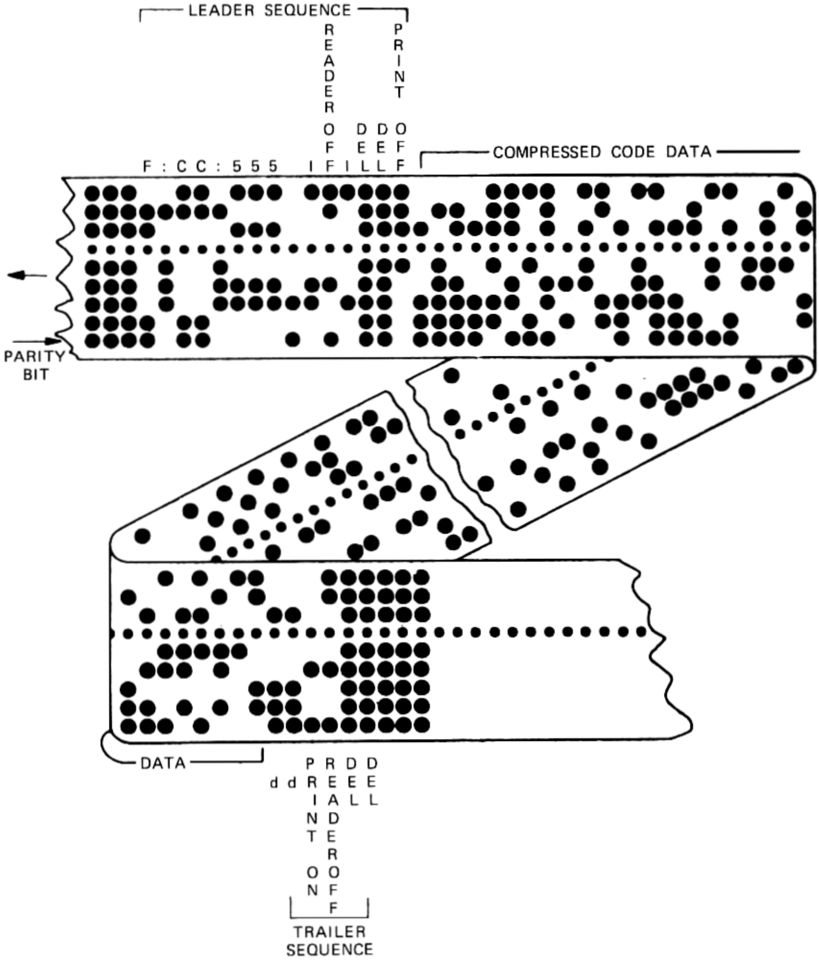


Fig. 7—Typical compressed code tape.

The call count and tag bits which account for six of the remaining useful bits need not be explicitly transmitted since the tag bits can be implied from entry type and the call count bits are initialized to a fixed value at time of reconstruction. It is thus found that, at most, only five BCD digits are required to record or reconstruct each called number entry. These are the hundreds, tens, and units, and two status digits. Sometimes an intercepted number has an associated new number reference. In this case the new number entry will always be found in the next consecutive disc word. The new number entry is treated as ten BCD digits consisting of a seven-digit directory number plus possibly either an area code or some special announcement code. Currently all ten digits for a new number entry are transmitted and recorded since a high degree of randomness is found for this type entry. The tag bits, however, need not be transmitted since they are implied by the entry type.

When full advantage is taken of all of the potential savings it is possible to reduce transmission times by almost an order of magnitude compared to methods using routine update facilities. In addition to reducing transmission time, the amount of storage media required to record all or part of the data base externally is also reduced. These savings benefit not only AIS, but also telephone company computer centers whenever such centers become involved in data base generation, backup storage, or routine transmissions to and from associated AICs.

At present for the AIS the storage medium for compressed data is paper tape. A typical compressed code paper tape consisting of a leader, data, and trailer is illustrated in Fig. 7. Teletypewriter page copy normally used to monitor paper tape transmissions would not be available, since to a teletypewriter compressed code appears as a random stream of both printing characters and nonprinting control characters. Except for the leader at the beginning of each thousands group and a summary message giving called and new number totals at the end of each thousands group, a teletypewriter page printer would normally be suppressed to avoid erratic printer operation. For AIS the loss of page copy is considered an acceptable trade-off in exchange for the overall time and storage media savings, particularly since an end result check of the reconstructed data can be made with audit programs and selective printouts of individual or group entries.

VI. SUMMARY

The AIS file administration programs provide both the external interactive human interface and the internal machine interface for managing the intercept information stored on duplicate disc files. Routine file administration activities have minimal effect on call processing as long as one good copy of the intercept data is available.

To aid in maintaining valid matched data on duplicate files, both automatic and externally initiated auditing functions are provided. Portions of the data base may be displayed in a variety of formats under the control of human requests.

Some special-purpose functions have been included for infrequent recovery and restructuring actions. The speed of both the special-purpose functions and the routine update functions is expected to improve when a data link interface to remote data processing equipment becomes available in the near future.

VII. ACKNOWLEDGMENTS

The authors wish to acknowledge contributions from many others associated with the AIS project, especially Mrs. T. C. Means for file administration teletypewriter programs and the late R. A. Sussman for initial development of the compressed code conversion tables.

REFERENCES

1. C. J. Byrne, W. A. Winckelmann, and R. M. Wolfe, "Automatic Intercept System: Organization and Objectives," B.S.T.J., this issue, pp. 1–18.
2. J. W. Hopkins, P. D. Hunter, R. E. Machol, J. J. DiSalvo, and R. J. Piereth, "Automatic Intercept System: File Subsystem," B.S.T.J., this issue, pp. 107–132.
3. T. E. Browne, T. M. Quinn, W. N. Toy, and J. E. Yates, "[No. 2 ESS] Control Unit System," B.S.T.J., *48*, No. 8 (October 1969), pp. 2619–2668.
4. H. Cohen, D. E. Confalone, B. D. Wagner, and W. W. Wood, "Automatic Intercept System: Operational Programs," B.S.T.J., this issue, pp. 19–69.