Proposal a:

There are several telephone switching centers all around the city. If we install one system in each of the centers, inquiries to each system are of low volume and less than 14 operators are enough for good service. Each system (center) offcourse keeps complete directory of the whole city. To update the files of all systems a central supervisor system with access to all other systems, and no reverse communication is needed.

A major advantage in such a system is that telephone switching load reduces drastically. There would be in fact no switching at all if all switching centers are provided with their own directory assistance system.

Proposal b:

The hardware of our Laleh computer can easily handle upto 32 concurrently working terminals (ports). Since the directory assistance application is a simple retrieval system, using all of this terminal capacity will not sensibly degrade the computer response time. The problem, however is that the CDOS only handles 17 ports. To overcome this limitation we have proposed that upto 8 terminals be hardware multiplexed to key in a telephone subscriber's name is a few order of magnitude more than the response time of the system at a CDOS port. Therefore handling 8 inquires at one port will not degrade the total response time.

5: Conclusion

A detailed study of the problem including all the practical aspects and human engineering factors has resulted in a very attractive and cost and time efficient computerized telephone directory system, implemented using 100% native equipment and expertise. Some unique properties are: inquiries need not specify full name, data entry is done without the need for any high level skill and business coding is easily done by the telephone company without any prerequisites defined by the system. Use of a virtual data generator during test phase has removed any need for onsite test, and has reduced the total software development cost. Last but not least is the use of a novel data dependent indexing method which reduces the size of index table drastically, and results in a very fast retreival.

The system with small modifications can be adapted for other applications such as banking.

f: OTHER PROGRAMS

A few other programs such as directory book printer, security, and data entry programs are also run strictly by the supervisor.

3.5: Test and the Data Generator Program

Data entry is very costly. Data record formats are subject to change for better performance during the design, implementation and test of any storage and retrieval system.

Testing the system with small amount of data will not reveal all the possible programming and design bugs. Full data entry before test is very costly. This cost is multipled if the data formats change as a design or implementation requirement, such that new full data entries become unavoidable.

Our solution was to generate data in large practical volumes, by running a program that we named it as "Virtual Data Generator". The VDG is supported by a small data pool consisting of few Thousands of family names and first names, few hundreds of suffix and prefixes for family names and several hundreds of street names and other address components. The VDG program generates telephone directory data records by choosing record components from it's data pool through a random number generator.

We have generated up to 1 million virtual data records by our VDG program. Since the virtual data are structurally the same as real data, they serve as good as real data for all test purposes. In case of any change in data format the VDG is modified to meet the change and another file of virtual data is generated at almost no cost compared to the cost of full data entry.

3.6: Data Entry

Special data entry programs are provided to help the telephone company to edit their data. Data records could be in either formats of Fig 5. The two formats are equivalant to all the programs and the system, except for a flag which differs between business phones and house-phones.

The directory book printer uses this flag to distinguish the two types of phones for separate print outs. career and business codes can be defined by the telephone company and need not be introduced to the system. Only operators should be instructed to use them properly.

4: Limitations and Expandabilities

The 16–bit Laleh microcomputer with CDOS operating system currently is capable of handling 17 ports concurrently. This is more than enough for all major cities of Iran, except for Tehran. Telephone company is encouraged to install two identical systems per site. The second system can function as the backup (stand by) of the main one. Any time in future that during the pick hours the limited number of terminals in one system would not be enough, extra operators can operate with the second system and its terminals.

The Laleh microcomputer can operate with hard disk capacities upto 320 Mega bytes, therefore one system could hold more than 1,000,000 data records of the city of Tehran.

Some hardware modifications has been proposed and tested to enhance the addressing capabilities of the system and encorporate an extra Random Access Memory bank.

With these modifications we have been able to write retrieval programs that locate the right block of data and bring it to the primary memory with only one secondary storage reference, even when size of the file is over 1 million records.

The limitation of 14 terminals per system raises some difficulties for directory assistance of very populated cities. Fourteen operators are too few to handle inquiries of big cities. To overcome this problem two proposals are under consideration:

few characters of the search key. The retrieval program will bring into memory the block of data which includes all the records whose search keys start with the entered characters. One of the pages in the block is displayed. If the operator does not see the required record on the screen,.He/She will depress one of the numeric keys to bring another page on the screen. Our on the job experience has shown that after a few days of working with the system, operators bring the required record on the screen with their first choice of numeric keys. The total time for entering the first few characters of the search key, eye search of the screen, depressing a numeric key and eye search of the second displayed page is far less than the time needed for entering the whole search key.

Operators are first instructed to enter LITE characters of the search key equal to the length of index table entry. This will guarantee that the right block of data will be brought to primary memory. In practice, however, because the quantity LITE has been computed in a worst case circumstances, the experienced operator enters less than LITE characters and brings the right block of data into primary memory. Moreover our experience has shown that most operators, after hearing the search key from the caller, know that how many characters they should enter to hit the right page without the need for using the numeric keys.

3.4: Supervisor's Tools & Programs

The supervisor is provided by the following updating programs:

a: INSERT

This program asks for the new data items to be added. The right sorted place of the data is not immediately placed. It rather is temporarily stored in the new item file and a blinking astrisk with optional beep sound is enforced into the 5th field of the previous record. The new record's address (in new item file) is also stored in an extra unvisible field of the previous record.

b: DELETE

This program asks for the record to be deleted, finds it but does not delete it immediately, displays the record to the supervisor for confirmation, and then set a "Deleted" flag in the record. This flag will prevent the data to be visible for the operators and other programs.

d: MODIFY

To modify a record one can delete the old record and insert the modified one. The MODIFY program, however is a better facility which interactively modifies the required record.

c: UPDATE

This program scans the whole file and inserts and deletes permanently all the temporarily inserted and deleted records. If the permanent update was to be done instantly for a single record update, it would require thousands of I/O operations and holding the whole system for several minutes. That is why permanent updates are scheduled to take place weekly or monthly and overnight when the system load is at it's minimum.

e: INDEXGEN

After drastic changes in directory data in several years or accepting a very large number of new subscribers in a short period of time, this program recomputes LITE and reorganizes the index table storage and index table lookup routine accordingly.

۸۵۶۷۵	ناصرزاده - سلیمان	آیت‌ا... بهشتی
۷۵۶۵۴	ناصرزاده - عبدالحسین	قصرالدشت
.....	....	....
.....	....	....
.....	....	....
۷۵۶۷۷	ناصری - سلمان	احمدی
۶۷۸۵۴	ناصریان - کیومرث	آیت‌ا... طالقانی

Block i

۷۵۵۷۵	ناصری تهرانی - کیومرث	تختی
۸۷۶۵۵	ناظری زاده - مسعود	بلوار گلستان
.....	....	....
.....	....	....
.....	....	....
۶۶۵۸۸	ناطق - یدا...	قصرالدشت
۶۲۳۴۵	ناطقی - حبیبا...	آیت‌ا... بهشتی

Block i+1

Fig 7: Sample sorted data blocks

tory assistance center, and build up the index table according to that value of LITE, it will be valid for a quite long time. In spite of this characteristic of telephone directory data we have provided our system with a special program called INDEXGEN, which can be run after each permanent update of the data file. This program will recompute the value of LITE and reorganize the index table storage and index table look-up routine accordingly.

3.3: Retrieval

To find a telephone number, the operator enters the whole search key. Only the first LITE characters of the enterd search key is compared against the entries in the primary memory resident index table. One block of data is brought to primary memory which includes a record with the given (requested) search key if existed in the directory at all. One screen length page of directory data is then displayed including the required telephone number's record. The number of records per screen can be set by the operators between 1 to 24.

Keying in the whole 40 characters of the search key is very time consuming, and will increase the response time too much. In our system the number of records that are brought to primary memory in one I/O operation is equal to the number of records that can be displayed in eight 24 record pages or 12 sixteen record pages (other page lengths are arbitrarily possible under operator control). Ten numeric keys on the keyboard can be used by the operator to instantly bring on the screen one of the 5 previous (or 5 next) pages respective to the current displayed page. With this facility operators may only enter the first

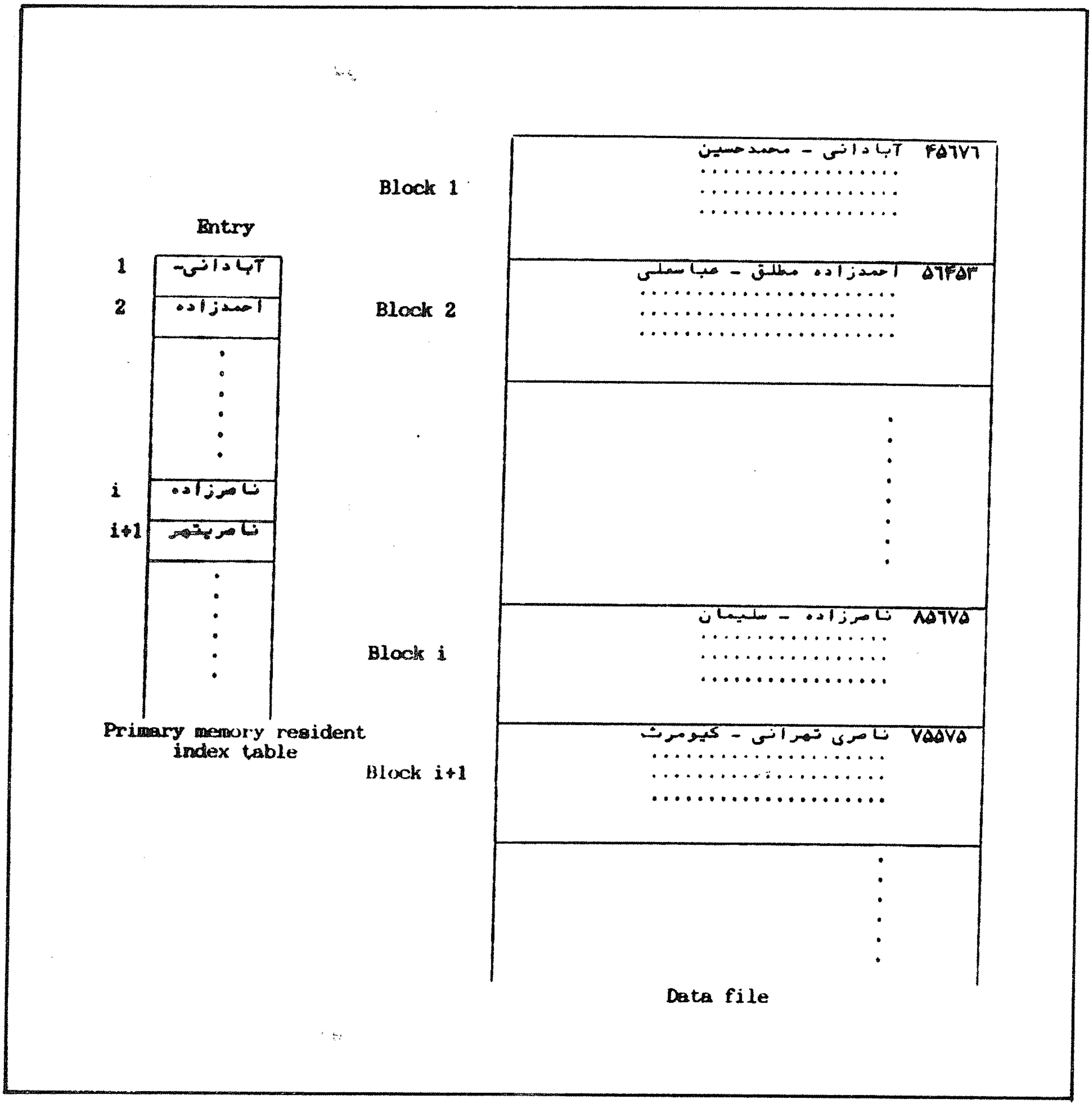

Fig 6: Schematic of the index table and data file.

reduction in index storage. This saving is also obtained in index table look-up time, because when searching the table for a given search key, at most 8 characters are compared instead of 40 characters.

For directory assistance centers with larger volumes of data, the LITE would be larger. Our investigations on LITE for the directory data of thecity of Tehran with more than half a million telephone numbers is about 12 to 14. This will result in more than 60% reduction in the index table storage and index table look-up time.

LITE is data dependent, but the subscriber names and business names do not vary that much. This is more so, when we consider only the beginning characters of subscriber or business names. Therefore if we compute LITE after the initial data entry for a direcappended, so the actual records are longer. There are 3 character string fields in each record, which are used in two formats as shown in (Fig 5). Data is kept sorted according to the second left field, which we call, the search key.

8 Bytes Tel_No	40 Bytes Name	32 Bytes Address	4 Bytes Link

8 Bytes Tel_No	40 Bytes Business Code & Address	32 Bytes Name	4 Bytes Link

Fig 5: Data record format.

3.2: Indexing

We consider the data file as an ordered set of equal size data blocks. Each block consists of data records of either formats of fig 5. The number of records in each block, that is size of each block, is equal to the size of a physical record of the system, where a physical record is transmitted to and from primary memory in one I/O operation. A schematic of index table for the data file is shown in Fig 6. Each entry in the index table is related to one and only one block, and holds the search key of the first record of that block.

To search for a record, its search key is looked up in the index table. Let BN denote the index of the found entry. This BN is equal to the number of the block holding the required record. The address of that block is computed as follows:

Block address = (BN − 1) Block-Size

To reduce the index table storage and index table look-up time, we keep only the first few characters of the search key as the index table entry. The number of sufficient characters for a successful search is computed as is followingly described.

3.2.1: Computing the length of index table Entry (LITE)

Fig 7 depicts two consecutive blocks of data. The first four characters of the search key are the same for all records of block i. The 4 character are "   ".

If in a block i, o or more consecutive beginning characters of the search key are the same for all records of that block, then let LITEi denote the count of those characters. This quantity for the blocks of Fig 7 is :

LITEi = 4   LITEi + 1 = 2

We compute LITEi for all (1   i   = Number of blocks), and find the maximum value, and call it LITE. We use this value as the length of the index table entry.

In a regular indexing method, one would take the length of the entry equal to maximum length of the search key. In our system 40 bytes are reserved for the search key. For real data of city of mashhad (80,000 Telephone numbers) LITE was computed to be only 8. Therefore the length of the entry of the index table is 8 instead of 40, which results in 80%

(Fig 4).

3: Internal File Structure and Programs

Our goal in designing the file structure has been to retrieve any permanently stored telephone-number with at most one I/0 refrence to secondary memory. This has been achieved by implimenting a small index table in primary memory. Details will come in section 3.2.

3.1: Data Record Format

Each record of directory data is at most 80 characters long equal to the length of one CRT line. A few extra bytes for some invisible system defined data are

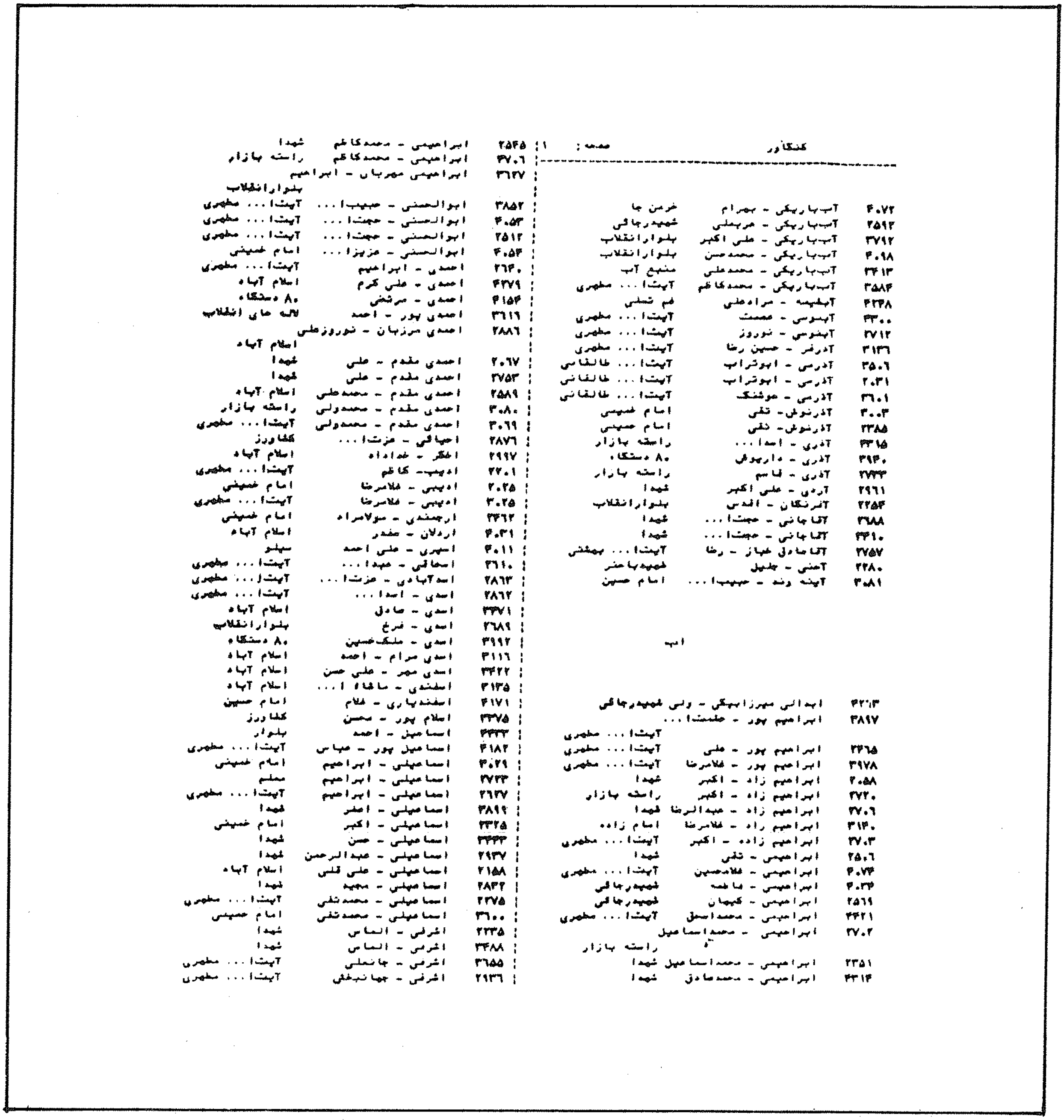

Fig 4: Sample directory book print.

۴۵۶۵۶	دک‌ـقدس: ابراهیمی - علی اکبر	زنان
۴۳۵۴۵	دک‌ـقدس: ابراهیمی - محسن	قلب‌و عروق
۴۶۳۲۴	دک‌ـقدس: ابراهیمی - محمد	عمومی
۵۴۵۶۷	دک‌ـقدس: ابراهیمیان - ناصر	زنان و زایمان
۵۶۵۶۷	دک‌ـقدس: ابراهیمیان کاشانی - عباس	متخصص‌بیهوشی
۳۷۸۵۴	دک‌ـقدس: ابراهیمی قمی - حسین	قلب
۴۵۷۶۶	دک‌ـقدس: ابراهیمی کیلان - احمد	بینایی
۴۸۷۷۵	دک‌ـقدس: ابوطالبی - حسن	اطفال
۵۶۶۵۴	دک‌ـقدس: ابوطالبی - کاظم	گوش‌وحلق و بینی
۷۸۹۶۷	دک‌ـقدس: ابومقداد - محمدحسین	قلب‌و ریه
۷۷۷۶۴	دک‌ـقدس: ابهری - محمدکاظم	داخلی
۵۶۷۵۶	دک‌ـقدس: ابهری - نورالدین	پوست
۸۵۷۵۴	دک‌ـقدس: ایزدپناه - حسینعلی	گوش‌وحلق و بینی
۵۷۸۶۵	دک‌ـقدس: ایزدپناه - محمدعلی	داخلی
۶۴۳۵۵	دک‌ـقدس: ایزدی - صادق	قلب‌و عروق
۴۵۶۷۷	دک‌ـقدس: ایزدی تهرانی - عباسعلی	بینایی

نام تقاضا شده :دک‌ـقدس:

Fig 3: Some of the physicians whose offices are located in Ghods Ave.

where data has been temporarily entered to the system by the supervisor. Later, He/She will permanently place the data in their correct sorted position by running a permanent update program. The update run usually takes place concurrently with system's regular service, but overnight when the system load is not heavy.

2.4: Security

When an Operator logs on the system, introduces himself by a password. A special security file is opened by the system to take note of attempting any unauthorized action. At the same time on such attempt the supervisor's process will be interrupted and some relevant message will be displayed on his/her terminal. This serves as an adequate security system to keep the data from unauthorized disturbances. The security files are also checked by the supervisor at the end of each work session or any other time that He/She wishes. some working statistics such as number of inquiries are also saved for each terminal per work session.

2.5: Directory Books

Telephone directory book prints in different sizes and page formats can be achieved by supervisor's commands. Business data can be printed separately. These prints can be used for mass production of telephone directory books. An example printout is shown in resembles exactly the search method of the microfiche system which operators are familiar with. If the caller does not specify the family-name and for example asks for the phone-number of a physician whose office is located in Ghods Avenue, the operator enters the first few (usually 2 suffices) characters of the business type followed by the first few characters of the main address component. Depressing the CR key will bring the consecutive list of the physicians whose offices are located in Ghods Avenue (Fig 3). From then on, the operator helps the caller by reading the other existing specifications of the data on the screen. This facility bears the cost of storing the business Data redundantly in two formats, one with the business owner's name as the search key, and the other with the business type and the address as the search key. Practically this neglectable extra cost is very well paid off.

2.3: New Items

Sometimes operators when looking up the screen for some data, see a blinking asterisk and/or hear a beep sound indicating that the required data has only been recently added to the directory and has not gone through a permanent update operation. The operator then depresses a special 'NEW ITEM' key which will change the screen to the one displaying the required data coming from a temporary update file (Fig 2),

۶۴۵۶۶ ناصری - محمدرضا   زند

۹۸۴۳۴ ناطقی - حسین   وصال

نام تقاضا شده : ناط

Fig 2: Recently entered records from the "NEWITEMS" file.

out of main memory, which is turn increases the search time. To avoid this increase, with the limited available data memory, we would be forced to reduce the number of operators which in turn would degrade the response time of the total directory assistance system. More on software will come in section 3.

2.2: Operators Interaction

To answer an inquiry, operators enter the first characters of the family name. One screen full of the data records appears on the terminal (Fig 1). Number of lines per screen can be changed by the operators and upto 24 lines.

Each record contains a family name, firstname(s), Address and telephone-number. The operator looks up for the complete name among the records on the screen. He or She might not find the name even if the name existed in the directory, because only the few first characters of the family-name were entered. In this case the operator pushes one of the ten numeric keys on the keyboard which are marked –5 to –1 and 1 to 5, to bring one of the 5 previous screen length pages or one of the next 5 pages on the screen. Usually the experienced operators find the right page with the first guess. This semi-computerized search method

۷۷۴۶۳	ناصری نژاد - ایرج	بلوار امیرکبیر
۸۷۹۴۵	ناصی - زهرابیگم	شمس
۵۴۶۶۷	ناطق الاسلام - حاج سیداسماعیل	زند
۳۴۵۶۷	ناطق الاسلام - سیدجعفر	قلعه خاتون
۵۴۶۶۷	ناطق فصیحی - سیدجمشید	بلوار آیتا...ربانی
۷۶۶۵۴	ناطق فصیحی - سیدمنصور	زند
۶۷۷۸۶	ناطق فصیحی - سیدنورالدین	زند
۸۵۶۷۵	ناطق فصیحی - سیدنورالدین	آیتا... بهشتی
۷۵۶۵۴	ناطق فصیحی - نورالدین	قصرالدشت
۴۶۷۶۷	ناطقی - علیرضا	جاده کفترک
۵۶۷۶۴	ناطقی - محمدباقر	زند
۶۵۷۵۶	ناطمیان - محمدباقر	زند
۸۶۷۵۴	ناظر - علی	زند
۵۶۷۶۵	ناظر - نیهتون	وصال
۶۷۵۶۵	ناظر - نیهتون	بلوار مدرس
۷۵۶۷۷	ناظرایلخانی - کیهان	قصرالدشت

نام تقاضا شده: ناط

Fig 1: Only 3 characters of the required name has been entered. Asterisk denote the existence of new items.

number of operators will not help much.

Use of high storage capacity and fast retreival capabilities of a computer system solves the stated problems. Most important is the possibility of immediate updating in a computerized directory without bearing any extra cost.

A commputerized telephone direcory system should respond faster, it should be capable of immediate updating, should provide proper tools for the operators to be able to answer to inquiries which may lack complete specifications and should be cost-effective.

A telephone directory data base is structurally and operationally a simple one. Generally speaking, it is a single entity data base with small and single key data records. There is only one query which, provides a name and asks for the telephone number.

The number of queries per time unit, however is very large. There are usually many operators interacting with the data base, trying to help callers with their requested telephone numbers.

A computerized telephone directory assistance with the stated characteristics should be implemented arround a powerfull multiprogramming operating system with a high degree of multi-programming. More than a hundred operator's concurrently interact with directory assistance system of big cities of the world. These systems have usually been implemented on a mainframe or at least a powerfull minicomputer system.

The Iran Telecommunication Research Center (I.T.R.C) has implemented a computerized directory assistance on her own made 16 bit Laleh microcomputer and bilingual CRT terminals. The system is currently operational and in service in several major cities of Iran with maximum capacity of 400,000 telephone numbers. A unique microsystem has been designed for city of Tehran, which will shortly be able to store more than a million telephone numbers. This paper describes the problem in general and the specifics of the I.T.R.C's design and implementation of such a system.

2- *General Description*

This section briefly describes the hardware, software and operations of our computerized telephone directory system.

2.1 :Hardware & Software

A 16 bit Lalehe microcomputer with up to 1MB primary memory, up to 160 MB hard disk, up to 14 operator bilingual CRT terminals with specially mapped keyboard, a supervisor terminal and two printers make up the hardware configuration of the system.

The maximum configuration has shown to be capable of providing good service for up to 400,000 telephone numbers.

Software consists of the main operating program, which is constantly running and interacting with the operators. Five other programs which are occasionally run by the supervisor, are used for temporary and permanent updates, and printing of telephone directory. A user friendly data entry program is used to input initial data. A security program will be invoked in case of unauthorized action through operator's password. All programs are written in 8086 assembly language and run on concurrent DOS operating system.

The reason why we chose to write assembly code instead of programming in a high level language or a 4th generation language such as dBase, Data Flex and so on, is that we are using a microsystem to implement a multiuser database. The micro version of 4th generation languages are usually good for few concurrent users. Their performance degrades heavily as the number of users and/or volume of data increase. The problem with micro version of 3rd generation programming languages is that different data areas should be allocated for each user process. Therefore, the amount of allocated main memory for each process is too small to hold the whole index table. This consequences in buffering portions of index table in and

A Microcomputer Design for Telephone Directory Assistance Centers

Ghasem Jaberipur, Ph. D

Shaheed Beheshti University

Ali Khojasteh Layen, M.Sc.

Iran Telecomunication Research Center

Abstract:

A computerized telephone directory assistance has been implemented on an Iranian 16–Bit Laleh16 microcomputer, with full Farsi language operations. Design and implementation specifics, including very short response time, responsivness to incomplete name entries, data dependent low volume indexing, career and address only specified inquiries, and use of a virtual data generator for test purposes are described.

1- Introduction:

Telephone directory data used to be kept on microfiches even for the city of Tehran with more than half a million telephone numbers. The major drawback of such a system is the difficulty and very high cost of keeping the directory upto-date. This had restricted the telephone company to update the data at most 3 times a year. The 4 month updating period, however is not tolerable, because most inquiries about a changed telephone number, or a new one, take place during the first month after assignment of the number.

While in many directory assistance centers, they are satisfied with the response time of the microfiche system, operator's duties are not simple, they become tired very soon, phisically and mentally. In these systems, the main, and most of the time the only operation is to search for the phone number of a house or a business. The caller should provide the name of the telephone subscriber or the business name exactly as is recorded in directory's microfiches. Name conflicts are usually resolved by address differences.

The mechanical restrictions and static nature of a microfiche system cannot cope with the rapid growth and frequent changes of directory data. Response time will start to degrade after the volume of telephone numbers passes a certain limit, and even a larger