LESSON 13:
INTRODUCTION TO ESDS, STORED PROGRAM CONTROL AND CENTRALIZED AND DISTRIBUTED STORED PROGRAM CONTROL

Objective
The objective here is to learn electronic space division switching, SPC and types of SPC.

Introduction
We know that call processing speed were low in early cross bar switching systems because they used electromechanical components for common control systems. Several efforts are being done to improve the speed of control and signalling between exchanges. Therefore, we have used the application of electronics in the design of control and signalling subsystems. In late 1940s and early 1950s, control functions in the telecommunications switching systems were realised by using various electronic devices such as vacuum tubes, transistors, gas diodes, magnetic drums and cathode ray tubes (CRT). Circuits using gas tubes were developed and employed for timing, ring translation and selective ranging of party lines. Single frequency signalling operation was realised by using vacuum tubes. Transistors were used in line insulation text circuits.

Registers and translators of the common control systems could be replaced by a signal digital computer.

Stored Program Control (Spc)
Stored program concept is used by modern digital computers. Here, a program or a set of instructions to the computer is stored in its memory and the instructions are executed automatically one by one by the processor. Stored Program Control (SPC) performs control functions of a switching system through programs stored in the memory of a computer. A immediate consequence of SPC in the full-scale automation of exchange functions and the introduction of a variety of new services to users. We can obtain following features by the application of SPC to telephone switching.

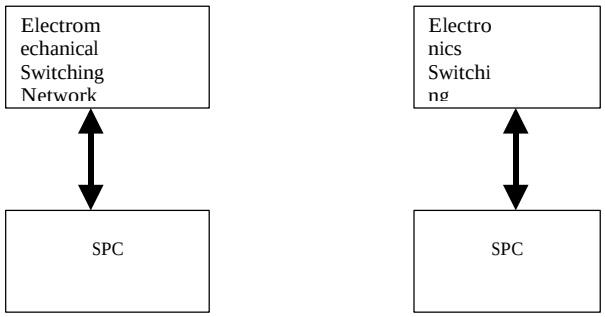
1. Common Channel Signalling (CCS)
2. Centralised maintenance and automatic fault diagnosis
3. Interactive human-machine interface.

Using a computer to carry out the control functions of a telephone exchange is not as simple as using a computer for scientific or commercial data processing. A telephone exchange must operate without interruption, 24 hours a day, 365 days a year and for say, 30-40 years. In other words, we can say that the computer controlling the exchange must be highly tolerant to faults. Early commercial computers had no fault tolerant features. Therefore switching engineers were faced with the test of developing fault tolerant hardware and software systems. In fact the field of telecommunication switching has given major contributions to fault tolerant computing.

Development of full-fledged electronic switching system takes place due to introduction of electronics and computers in the control subsystem of a telephone exchange. In a full fledge electronic switching system or telephone Exchange, the switching network is also electronic. The world's first electronic switching system (known as No.1 ESS) was commissioned by AT&T at Succasunna, New Jersey, in May 1965. Since, then, the history of electronic switching system and stored program control (SPC) has been one of rapid and continuous growth in versatility and range services. Today, all the electronic exchanges possess standard feature of 8PC. However, attempts to replace the space division electromechanical switching matrices by semiconductor cross point matrices have not been greatly successful particularly in large telephone exchanges. Therefore, switching engineers have been forced to return to electromechanical miniature crossbars and reed relays, but with a complete electronic environment.

As a result, many space division electronic switching systems use electromechanical switching networks with stored program control (SPC). Nonetheless, private automatic brand exchanges (PABX) and smaller exchanges do use electronic switching devices.

We have two types of space division electronic switching systems. One use electromechanical switching network and the other electronic switching network. These are shown in Fig. 13.1



(a) Electromechanical Switching Network (b) Electronic Switching Network

Fig.13.1 Electronic Space Division Switching Systems

Both types are called electronic switching systems. But only one of them is fully electronic. With the evolution of time division switching, which is done in electronic domain, modern telephone exchanges are fully electronic.

There are basically two methods to organising Stored Program Control (SPC):

1. Centralised Stored Program Control or Centralised SPC
2. Distributed Stored Program Control or Distributed SPC

Centralised SPC approach being used by early electronic switching systems developed during the period 1970-75. Although many present day telephone exchanges designs continue to use centralised SPC. With the advent of low cost

powerful microprocessors and VLSI chips such as programmable logic arrays (PLA) and programmable logic controllers (PLC), distributed SPC in gaining popularity.

Centralised Stored Programcontrol

It is denoted as centralised SPC. In centralised control, all the control equipments are replaced by a single processor. This processor must be quite powerful. This processor is capable of processing 10 to 100 calls second. The processing capability of the processor depends on the load on the system. It simultaneously performs many other ancillary tasks. Fig.13.2 shows block diagram of an Electronic Switching System (ESS) using centralised SPC.

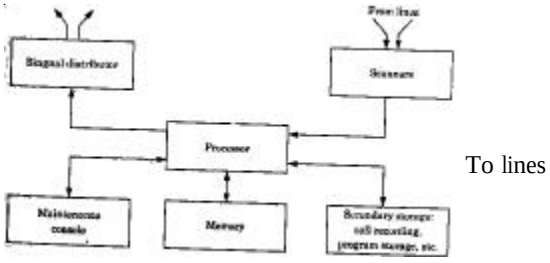


Fig 13.2 Block diagram of n ES system using centralised SPC.

A centralised SPC configuration may use more than one processor for redundancy purposes. Each processor has access to all the telephone exchange resources like scanners and distribution points. It is capable of executing all the control functions. A redundant centralised control structure is shown in Fig.13.3. Redundancy may also be provided at the level of exchange resources and function programs. In actual implementation the telephone exchange resources and memory modules contain the program for carrying out the various control functions. The exchange resources and the memory modules may be shared by processors or each processor may have its own dedicated access paths to exchange resources and its own copy of programs and data in dedicated memory modules.

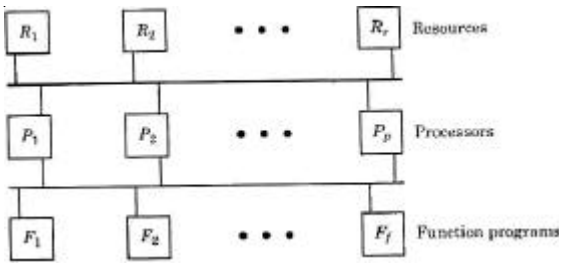


Fig. 13.3 A redundant centralised control structure.

At present, only a two-processor configuration is used in most of the present day electronic switching systems (ESS) using centralised control.

There are three modes and in one of them a dual processor architecture can be operated. These modes are:

1. Standby mode of operation
2. Synchronous Duplex mode of operation
3. Load sharing mode of operation.

Now we discuss each mode in detail.

Standby Mode of Operation

This mode of operation is the simplest of dual processor configuration operations. Normally, one processor is active and the other is on standby. Both hardware and software wise The standby processor is used on-line only when the active processor fails, this configuration should have an important requirement that the standby processor is capable of reconstituting the state of the exchange system when it takes over the control.

The State of the exchange means that the subscribers and trunks are busy or free or which of the paths are connected through the switching network etc. In small telephone exchanges, this may be possible by scanning all the status signals as soon as the standby processor is brought into operation. In such a situation only the calls which are being established at the time of failure of the active processor are distributed. But in large exchanges it is not possible to scan all the status signals within a reasonable time. Here, the active processor copies the status of the system periodically, say every 5 seconds into secondary storage. When switch over occurs, the on-line processor loads the most recent update of the system status from the secondary storage and continues the operation. In this case, only the calls which changes status between the last-update and the failure of the active processor are distributed. A standby dual processor configuration with a common backup storage is shown in Fig.13.4. The shared secondary storage need not be duplicated and simple unit level redundancy would suffice.

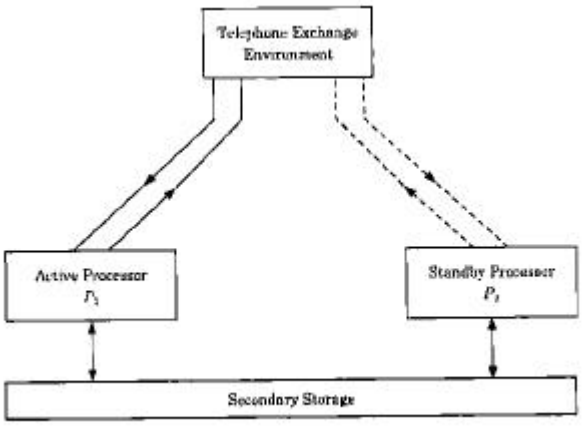


Fig. 13.4 Illustration of standby dual processor configuration.

Synchronous Duplex Mode of Operation

In this mode of operation, hardware coupling is provided between the two processors. These processors execute the same set of instructions and compare the results continuously. If a mismatch between the results obtained from two processors occurs, the faulty processor is identified. This identified processor is taken out of service within few milliseconds. If the two processors have the same data in their memories at all times, then the system will operate in normal manner. Hence, two processors simultaneously receive all information from the telephone exchange environment. One of the processors

actually controls the telephone exchange. But the other processor is synchronized with the first processor but does not participate in the exchange contro1.

Fig. 13.5 shows the configuration of synchronous duplex operation.

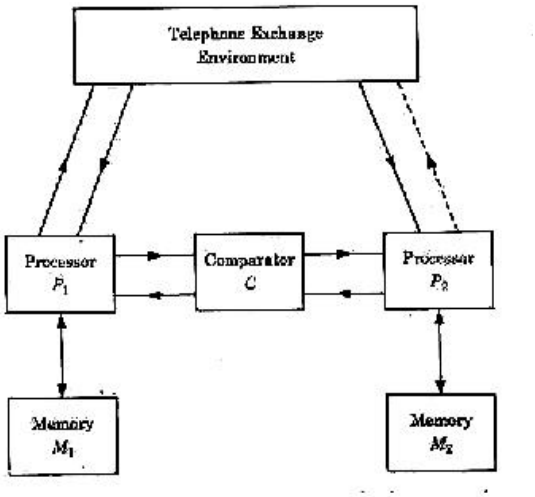


Fig.13.5 Configuration of synchronous duplex operation

If a fault is detected by the comparator C, the two processor P1 and P2 are decoupled and a checkout program is run independently on each of the machines to determine which one is faulty. The check-out program runs without disturbing the call processing which is suspended temporarily, The other processor operates independently in case if a processor is taken out of service on account of a failure or for maintenance, When a faulty processor is repaired and brought into service, the memory contents of the active processor are copied into its memory. Therefore, repaired faulty processor is brought into synchronous operation with the active processor and then the comparator is enabled.

It is also possible that a comparator fault occurs on account of a transient failure, which does not show up when the checkout program is run. In such cases, we have three possibilities to continue the operation,

1. Continue with both the processors.
2. Take out the active processor and continue with the other processor.
3. Take out the other processor and continue with the active processor.

Strategy (1) is based on the assumption that the fault is a transient one and may not reappear, Many times the transient faults are the forerunners of an impending permanent fault, These permanent faults can be detected by an exhaustive diagnostic test of the processor under marginal voltage, current and temperature conditions,

Strategy (2) and (3) are based on this hypothesis. The processor that is taken out of service is subjected to extensive testing to identify a marginal feature in these cases.

Load Sharing Mode of Operation

In this mode of operation an incoming call is assigned randomly or pre deterministically to one of the processors which then handles the call right through completion. Thus both the processors are active simultaneously and share the load and the resources dynamically. The configuration of load sharing operation is shown in fig. 13.6. Both the processors have access to the entire exchange environment. Therefore these processors are used to sense as we1l as control the exchange environment. Since the calls are handled independently by the processors. Both the processors have separate memories for storing temporary call data. Although programs and semi-permanent data can be shared, they tire kept in separate memories for redundancy purpose. There is an interprocessor link. Interprocessor link is used for sending processors exchange information. This information is needed for mutual coordination and verifying the state of health of the other.

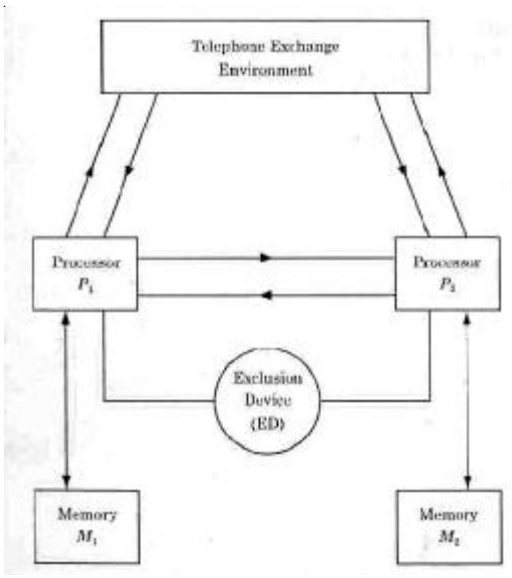


Fig. 13.6 Load sharing configuration that uses a hardware exclusion device (ED)

If the exchange of information between processor fails, one of the processors which detect the same takes over the entire load. This entire load includes the calls that arc already set up by the failing processor. However, the calls that were being established by the failing processor are usually lost. Sharing of resources needs for an exclusion mechanism so that both the processors do not seek the same resource at the same time. The exclusion mechanism may be implemented in software or hardware or both. Load sharing configuration that uses a hardware exclusion device is shown in Fig. 13.6. This exclusion device is set by one of the processors until it is reset by the first processor.

Under normal operation each processor handles 50% of the cells on statistical bases. The exchange operation can, however, send commands to split the traffic unevenly between the two processors. Such type of commands are required, for example to test a software modification on one processor at low traffic, while the other handles majority of the calls. Load sharing configuration gives much better performance in the presence of

traffic overloads as compared to other operating modes. Since the capacities of both the processors are available to handle over loads. Load sharing configuration increases the effective traffic capacity by about 30% when compared to Synchronous Duplex Operation. Load sharing is a step towards distributed control.

One of the main purposes of redundant configuration is to increase the overall availability of the system. A telephone exchange must show more or less a continuous availability over a period of 30 or 40 years.

Comparison of the availability figures of a single processor and Dual processor system:

The availability of a single processor system is given by

$$A = \frac{T_f}{T_f + T_r} \quad \ldots\ldots\ldots\ldots \text{Eq.1}$$

Where,

T_f = Mean time between failures (MTBF)

Tr = Mean time to repair (MTTR)

The unavailability of the single processor system is given by

(Unavailability)U = 1 - A (Availability

Or

$$U = 1 - A \quad \ldots\ldots\ldots\ldots \text{Eq.2}$$

Substituting Eqn.1 in Eqn. 2, we get

$$U = 1 - A$$

$$= 1 - \frac{T_f}{T_f + T_r}$$

$$= \frac{T_f + T_r - T_f}{T_f + T_r}$$

Or $$U = \frac{T_r}{T_f + T_r} \quad \ldots\ldots\ldots\ldots \text{Eq.3}$$

If $T_f >> T_r$, then, eq.3 can be written as

$$U = \frac{T_r}{T_f} \quad \ldots\ldots\ldots\ldots \text{Eq.4}$$

For a dual processor system, the mean time between failures, $MTBF_{D,}$ can be computed from the MTBF and MTTR values of the individual processors. A dual processor system is said to have failed only when both the processors fail find the system is totally unavailable.

The situation of total unavailability of system arises only when one of the processors has failed and the second processor also fails when the first one is being referred. In other words, we can say that this is related to the conditional probability that the second processor fails during the MTTR period of the first processor when the first processor has already failed. Here we arc not driving expression for result of dual processor system. MTBF of the dual processor system is given by,

$$T_f' = \frac{(T_f)^2}{2T_r} \quad \ldots\ldots \text{Eq.5}$$

Therefore the availability of the dual processor system is given by

$$A' = \frac{T_f'}{T_f' + T_r} \quad \ldots\ldots \text{Eq.6}$$

Substituting the value of T_f' from eqn. 5 in eqn. 6 , we get

$$A' = \frac{T_f'}{T_f' + T_r}$$

$$= \frac{(T_f)^2 / 2T_r}{\frac{(T_f)^2}{2T_r} + T_r}$$

or $$A' = \frac{(T_f)^2}{(T_f)^2 + 2(T_r)^2} \quad \ldots\ldots \text{Eq.7}$$

Therefore, the unavailability of the dual processor system is given by

$$U' = 1 - A'$$

$$= 1 - \frac{(T_f)^2}{(T_f)^2 + 2(T_r)^2} = \frac{2(T_r)^2}{(T_f)^2 + 2(T_r)^2}$$

or $$U' = \frac{2(T_r)^2}{(T_f)^2 + 2(T_r)^2} \quad \ldots\ldots \text{Eq.8}$$

If $T_f >> T_r$, then we have from eq.8,

$$U' = \frac{2(T_r)^2}{(T_f)^2} \quad \ldots\ldots\ldots\ldots \text{Eq.9}$$

Example 13.1: Determine the unavailability for single and dual processor systems. We have given that MTBF = T_f = 2000 hours and MTTR = T_r = 4 hours.

Solution:

Unavailability of the single processor is determined by,

$$U = \frac{Tr}{T_f} = \frac{4}{2000} = 2 * 10^{-3}$$

i.e.525 hours in 30 years.

Unavailability of the dual processor is determined by,

$$U' = \frac{2(Tr)^2}{(T_f)^2} = \frac{2(4)^2}{(2000)^2} = 8 * 10^{-6}$$

i.e.2.1 hours in 30 years.

As we know that control subsystem in an exchange performs four important functions.

Event monitoring

Call Processing

Charging

Operation and Maintenance (O&M)

Now we consider the real time response requirements. In such a situation, these functions may he grouped under three levels, These are shown in Fig. 13.7.

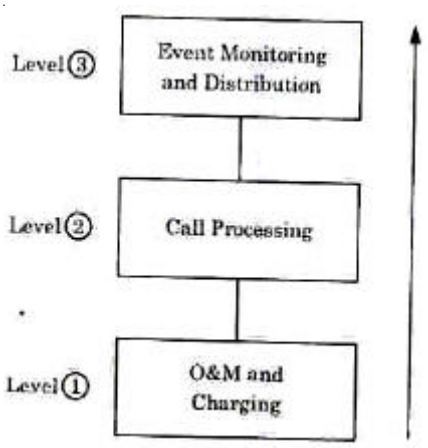


Fig.13.7 Illustration of 3 levels of control functions

Event monitoring has the highest real time constraint. But Operation and Maintenance (0&M) has the least real time constraint. The real time constraint necessitates a priority interrupt facility for processing in centralised control. If an event occurs when O&M function is being curried out by the control processor, the O&M processing has to be interrupted, the event processing taken up and completed and then the O&M functions processing resumed. Interrupts are necessary to suspend any low level function and take up the processing of higher-level functions. These illustrated in Fig. 13.8. When an interrupt occurs, program execution is shifted to an appropriate service routine address in the memory through a brand

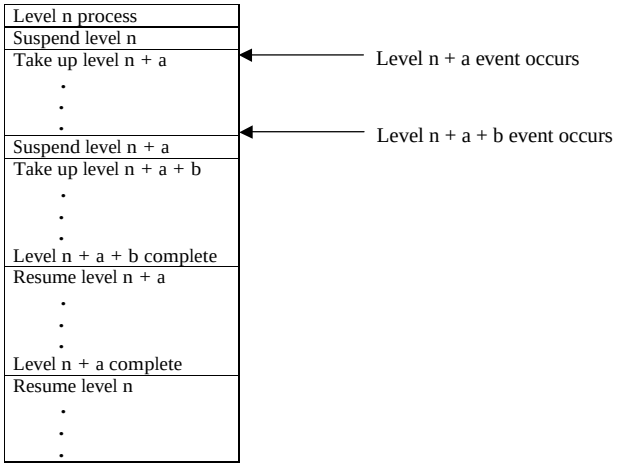


Fig.13.8 Illustration of interrupt processing.

There are two methods of accomplishing this,

1. Vectored Interrupt
2. Non-vectored Interrupt.

Vectored Interrupt

In this interrupt, the interrupting source supplies the brunch address information to the processor. The set of addresses supplied by different interrupting sources is called interrupt vector. Vectored interrupt is faster in response than the non-vectored interrupt.

Non-vectored Interrupt

In this interrupt, the brand address is fixed and a main interrupt service routine scans the interrupt signals and decides on the appropriate routine to service the specific interrupt.

Distributed Stored Program Control

It is denoted by distributed SPC. In this control, the control functions are shared by many processors within the telephone exchange itself. This type of structure uses the low cost microprocessors. This structure offers better availability and reliability than the centralized SPC.

Exchange control functions may be decomposed either horizontally or vertically for distributed processing.

Horizontal Decomposition of Exchange Control Functions

In this decomposition each processor performs only one or some of the exchange control functions. A chain of different processors may be used to perform various operations such as event monitoring, call processing and O&M functions.

The entire chain may be duplicated as shown in Fig. 13.9 for providing redundancy. Operating principles of both dual processor structure and dual chain configuration is the same.

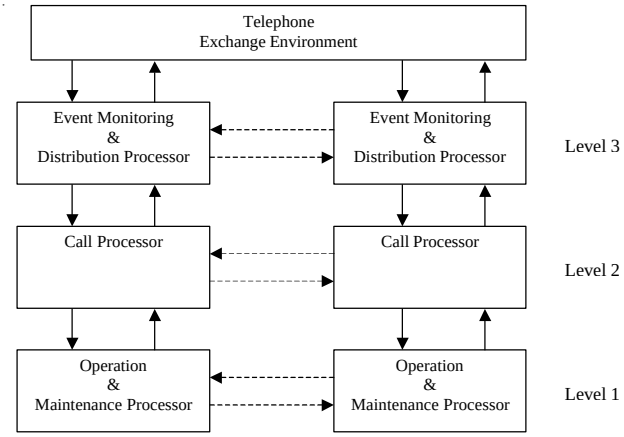


Fig. 13.9 Dual chain distributed Control.

Vertical Decomposition of Exchange Control Function

In this decomposition, the exchange environment is divided into several blocks and each block is assigned a processor. This processor performs all control functions related to that block of equipments. The total control system now comprises of several control units coupled together. The processor in each block may he duplicated for purposes. It is operated in one of the three dual processor-operating models.

Processing is done in three levels separately.

1. Level-3 processing
2. Level-2 processing
3. Level-l processing

Now, We shall discuss each processing in subsequent subsections.

Level-3 Processing

In distributed control, the specific functions arc performed by the processors. So these processors can be specially designed to curry out these functions efficiently. We can see from Fig. 13.9 that a level3 processor handles event monitoring and distribution functions. In other words, it handles scanning, distribution and marking functions. The processor and the associated devices are located physically close to the switching network junctions and signalling equipment processing operation~ involved are of simple, specialised and well-designed in nature. Generally, processing at this level performed in setting or sensing of one or more binary conditions in flip flops or registers.

It is necessary to sense and alter a set of binary conditions in a predefined sequence to accomplish a control function. We use either wired logic or micro programmed devices for performing: simple operation efficiently.

A control unit is designed as a collection of logic circuits using logic elements and therefore it is called electronic. Otherwise it is called u 'Hard Wired' control unit. A hard-wired control unit can be exactly tailored to the job in hand, both in terms of the function and the necessary processing capacity. But there is lack of flexibility in hardwired control unit. It cannot be easily adopted to new requirements. A micro programmed control unit is more universal. It can be put to many different uses by simply modifying the micro program and the associated data. The micro programmed control units are more expensive and slower than hardwired control unit for the same technology and an equivalent processing capacity.

When the processing is complex, the micro programming implementation is easier. Characteristics of micro programmed and Hard-wired control are summarized in Table 13.9.

Table 13.1 Characteristics of Electronic control schemes such as Hard-wired control and Micro programmed control.

Sr. No	Microprogrammed Control	Hard-wired control
1.	It is flexible	Not flexible
2	It is slower	Faster
3	It is more expensive for moderate processing function.	It is less expensive for moderate, simple and fixed processing.
4	It is easier to implement complex processing functions.	It is difficult to implement complex processing functions.
5	In the control we can easily introduce new services	In this control it is not easily possible to introduce new services
6	It is easier to maintain	It is not easier to maintain

With the advent of low cost microprocessors and VLSI PLAs and PLCs, microprogramming is the favoured choice for level-3 processing.

In microprogramming, the binary conditions required for control functions are changed through a control wont This control word contains a bit pattern that activates the appropriate control signals. By storing a set of control words in a memory and reading them out one after another, control signals may be activated in the required sequence.

There are two approaches to design of control word in micro-programmed system.

These approaches are given below:

1. The control word may he designed to contain one bit per every conceivable control signal in the system. This control scheme is called Horizontal control.
2. In other control scheme, all the control signals may he binary encoded and the control word may contain only the encoded pattern. This control scheme is called vertical control.
   a) Horizontal Control is flexible and fast in the sense that as many control signals as required may be activated simultaneously.
   b) But Horizontal control is expensive as the control word width may be too large to realise periodically.
   c) In vertical control only one signal at a time is activated. Therefore the time penalty to activate a set of signals may be unacceptably large.

Some of the recent designs use standard microprocessors for scanning and distribution functions instead of designing a micro-programmed unit. The Micro Processor based design is somewhat slower than the micro programmed unit. Micro-programmed unit itself is likely to dominate until low cost custom ICs for these functions become available.

Level-2 Processing

Level-2 processors are also caned switching processors. Earlier, general-purpose computers were not suitable for real time applications. They were 1arge in size and expensive. With the arrival of minicomputers and then microprocessors a number of real time applications outside the field of te1ecommunications have sprung up. Therefore it has led to the appearance of standard processors suitable for real time applications in the market.

Nonetheless the te1ephone exchange manufacturers have continued to prefer house-developed switching processors for some time in order to maintain full control over the products and to continue the costs. However new trend is to use commercially available standard microprocessors for the switching processor functions.

Fundamentally switching processors are same as general-purpose digital computers. However there are certain characteristics that are specific to switching processors.

Processor instructions for instance, are designed to allow data to be packed more tightly in memory without unduly increasing the to and fro time. Single-bit and 1·bit manipulation instructions are used extensively in switching applications.

Special instructions for task and event queue management are desirable. These special instructions would enable optimal run times for certain scheduler functions.

The architecture of switching processor is designed to ensure the following:

- 99.9% availability

- Fault Tolerance
- Security of operation.

In the input/output (I/O) area, the switching processor differ from general purpose computers mainly in telephone peripherals such as scanners, distributors, and markers along with the conventional data processing type peripherals like teleprinters, magnetic tapes etc.

The total I/O data transfer is not very high in switching processors. I /O data transfer is of the order of 100 k bytes/ second for large systems. I/O data transfer is done by two techniques. These are:

- Program controlled data transfer
- Direct memory access (DMA) techniques.

Sometimes the exchange peripherals are located far away from the switching processor. Consequently special communication links are required to connect them to the I / O controller.

The traffic handling capacity of the control equipment is usually limited by the capacity of the switching processor. The load on the switching processor is measured by its occupancy (O). Occupancy (O) is given by the simple formula.

$$O = h_f + t_a N \qquad \text{……. Eq.10}$$

Where,

h_f = Fixed overhead. it depends upon the exchange capacity and configuration.

t_a = Average time to process one call

N = Number of calls per unit time

The occupancy (O) is expressed as a fraction of the unit time for which the processor is occupied. The parameter h_f depends to a large extent on the scanning workload. The scanning workload depends usually on the number of subscriber lines, trunks and service circuits in the telephone exchange. The value of parameter h_f may be determined by knowing the total number of subscriber lines the number of instructions required to scan one line and the average execution time per instruction. The estimation of the value of the parameter to requires the definition of call mix, comprising incoming, outgoing, local and transit calls.

The number of instructions required to process each type of call varies considerably. For example, the number of instructions required to process an incoming call is much less than the number of instructions required to process a transmit call. This is why; there is no need to retransmit the address digits. The result of a call attempt such as call put through, called party busy and no answer also affects the number of instructions to be executed. The number of subscribers with DTMF and rotary dial telephones and the percentage of calls to grouped lines are also important factors. The grouped lines are also called PBX lines. Now we are taking above factors into account that a call mix may be worked out and the mean processing time per call attempt calculated. Mean processing time per call attempt is the weighted average of the processing times for various types of calls.

Usually, the switching processor is designed to handle a traffic load which is 40% higher than the nominal load. When this overload occurs, the processor may be overloaded only to 95% of its capacity so that traffic fluctuations can be absorbed.

Substituting $O = 95\% = 0.95$ and

$N = 140\%$ of N_N

$= 1.40\ N_N$ in eq.10 we get

$$O = h_f + t_a N$$

Or $$0.95 = h_f + t_a * 1.40\ N_N$$

Or $$0.95 = h_f + 1.4\ ta\ N_N \qquad \text{…….. Eq.11}$$

Where,

N_N = Nominal load in terms of number of calla per unit time

Eq.11 can be written as

$$N_N = \frac{0.95 - h_f}{1.4\ t_a} \qquad \text{……..Eq.12}$$

The average instruction execution time is dependent on the instruction mix as different instructions take different times.

Level-l Processing

This level processing handles operations and maintenance (O&M) functions. O&M functions involves the following steps:

1. They administer the telephone exchange hardware and software.
2. They add, modify or delete information in translation tables.
3. They change subscriber class of service.
4. They put a new subscriber line or trunk into operation.
5. They supervise operation of the telephone exchange.
6. They monitor traffic.
7. They delete arid locate faults and errors.
8. They run diagnostic and test programs
9. There is a man-machine interaction.

The complex nature of the functions demands a large configuration for the levels computer involving large disk or tape storage. Therefore, O&M processor in many cases is a standard general-purpose computer. This is generally a mainframe computer. The complexity and the volume of the software are also the highest as compared to level 2 and level-3 processing. The O&M functions are less subject to real time constraints and have less need for concurrent processing. Hence, it is common practice that a single O&M computer is shared among several telephone exchanges located remotely. It is shown in Fig. 13.10. In such an arrangement, the telephone exchanges contain only level-2 and level-3 processing modules. We require expert maintenance personnel for attending several exchanges from one central location for the purpose of remote diagnosis and maintenance.

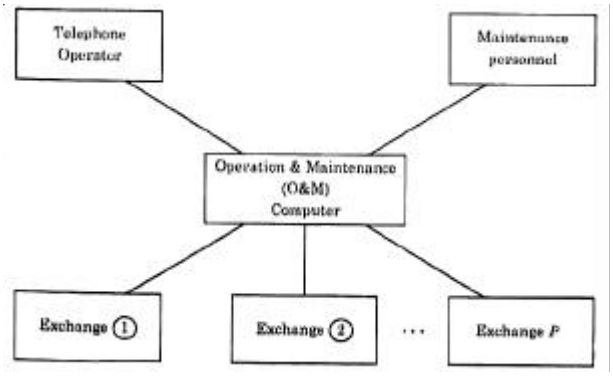


Fig.13.10 Remote Operation and Maintenance

The three-level distributed control though typical and popular is by no means the only method of distributing control function. In many telephone exchange designs we use a single computer located physically at the telephone exchange site, this computer is used to perform both the O&M and call processing functions. But in many exchange designs, we may use two different processors for O&M and call processing. Instead, they may use one dedicated Q&M processor for each telephone exchange.

Notes