The OSI Model: Overview on the Seven Layers of Computer Networks

Ms. Priti V. Jasud
Assistant Professor
Department of Information Technology
KDK college of Engineering, Nagpur, Maharashtra, India

Abstract

Due to the urgency in the need for standards for heterogeneous computer networks, International Standard Organization (ISO) created a new subcommittee for “Open System Interconnection” (ISO/TC97/SC16) in 1977. The first priority of subcommittee 16 was to develop architecture for Open System Interconnection which could serve as a frame work for the definition of standard protocols. As a result 18 months of studies and discussions, SC16 adopted a layered architecture comprising seven layers (Physical, Data Link, Network, Transport, Session, Presentation, and Application). In july 1979 the specifications of this architecture, established by SC16, were passed under the name of “OSI Reference Model” to Technical committee 97 “Data processing” along with recommendations to start officially, on this basis a set of protocols standardizations to start projects to cover the most urgent needs. These recommendations were adopted by TC97 at the end of 1979 as the basis for the following development of standards for Open System Interconnection within ISO. The OSI Reference Model was also recognized by International Telegraph and Telephone Consultative Committee (CCITT) Rapporteur’s Group on “Layered Model for Public Data Network Services”. This paper explains the OSI Reference Model, which comprises of seven different layers. Each layer having their own responsibilities.

Keywords: Open System Interconnection (OSI), International Standard Organization (ISO), Protocol

I. INTRODUCTION

The International Standard Organization (ISO) is a multinational body dedicated to worldwide agreement on international standard which was established in 1947. The ISO proposed a model named as OSI (Open System Interconnection) in 1983, which covers all aspects of network communication. The purpose of OSI model is for open communication between different systems without requiring changes to logic of the underlying hardware and software. The OSI model is not a protocol, it is a model for understanding and designing a network architecture that flexible, robust and interoperable. In 1977, the International organization for Standardization (ISO) recognized the special and urgent need for standards for heterogeneous informatics networks and decided to create a new subcommittee (SC16) for “Open Systems Interconnection” [1], [2]. The universal need for interconnecting systems from different manufacturers rapidly became apparent [3], leading ISO to decide for the creation of SC16 with the objective to come up with standards required for “Open Systems Interconnection”. The term “open” was chosen to emphasize the fact that by conforming to those systems obeying the same standards throughout the world.

The first meeting of SC16 was held in March 1978, and initial discussions revealed [4] that a consensus could be reached rapidly on a layered architecture which would satisfy most requirements of Open Systems Interconnection with the capacity of being expanded later to meet new requirements. SC16 decided to give the highest priority to the development of standard Model of Architecture which would constitute the framework for the development of standard protocols. After less than 18 months of discussions, this task was completed, and the ISO Model of Architecture called the Reference Model of Open Systems Interconnection [5] was transmitted by SC16 to its parent Technical Committee on “Data Processing” (TC97) along with recommendations to officially start a number of projects for developing on this basis an initial set of standard protocols for Open Systems Interconnection. These recommendations were adopted by TC97 at the end of 1979 as the basis for following development of standards for Open Systems Interconnection within ISO. The OSI Reference Model was also recognized by CCITT Rapporteur‟s Group on Public Data Network Services. CCITT (Consultative Committee for International Telephony and Telegraphy) is a part of the ITU (International Telegraph Unit) which defined many important standards for data communications and it coordinates standards for telecommunications. The CCITT, now known as the ITU-T (for Telecommunication Standardization Sector of the International Telecommunications Union), is the primary international body for fostering cooperative standards for telecommunications equipment and systems.

II. RELATED WORK

Considering the urgency of the need for standards which would allow constitution of heterogeneous computer networks, International Standard Organization (ISO) created a new subcommittee for “Open System Interconnection” in 1977. The main objective of subcommittee 16 was to develop architecture for Open System Interconnection (OSI) which could serve as a frame work for the definition of standard protocols. As a result 18 months of studies and discussions, SC16 adopted a layered architecture comprising seven layers shown in figure(1).

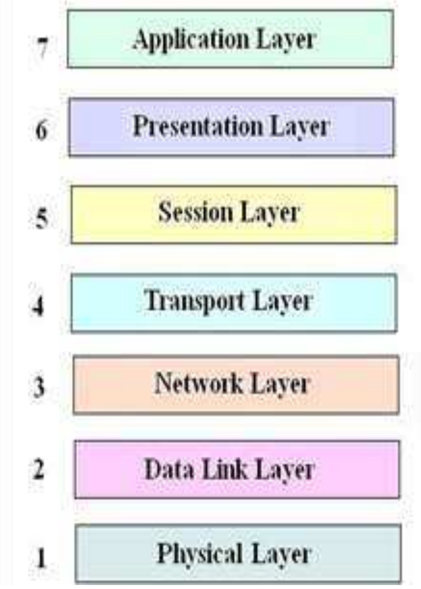

Fig. 1: OSI Layers

International Standard Organization (ISO) created a Subcommittee (SC16) whose basic objective is to standardize the rules of interaction between interconnected systems. Thus, only the external behaviour of Open Systems must conform to OSI Architecture, while the internal Organization and functioning of each individual Open System is out of scope of OSI standards since these are not visible from other systems with which it is interconnected [6]. It should be noted that the same principle of restricted visibility is used in any manufacturer's network architecture in order to permit interaction of systems with different structures within the same network. These considerations lead SC16 to prefer the term of "Open Systems Interconnection Architecture" (OSIA) to the term of "Open System Architecture" which had been used previously and was felt to be possibly misleading. However, for unclear reasons, SC16 finally selected the title "Reference Model of Open Systems Interconnection" to refer to this Interconnection Architecture. The next section presents a description of OSI layering and principles of ISO for the seven layers of OSI Architecture followed by a brief explanation of how the layers were chosen. There after seven layers of OSI architecture and conclusion are presented in the subsequent sections, followed by an acknowledgement section and a summary of references for this manuscript.

III. OSI LAYERING

Layering is a structuring technique which permits the network of Open Systems to be viewed as logically composed of a succession of layers, each wrapping the lower layers and isolating them from the higher layers.

The subcommittee (SC16) which is created by ISO has given an illustration of layering shown in figure (2) where successive layers are represented in a vertical sequence, with the physical media for Open Systems Interconnection at the bottom.

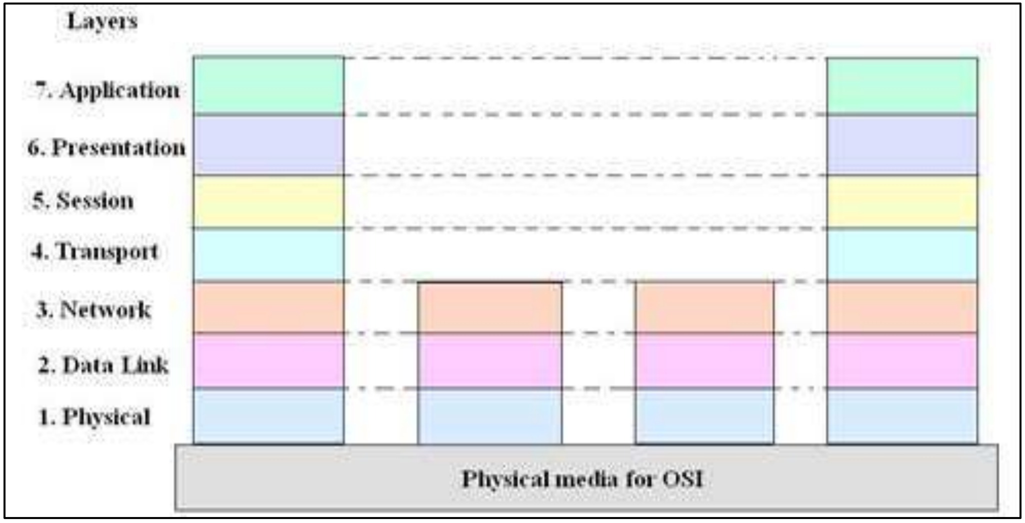

Fig. 2: illustrates an example of OSI representation of layering

Each individual system itself is viewed as being logically composed of a succession of subsystems, each corresponding to the intersection of the system with a layer. In other words, a layer is viewed as being logically composed of subsystems of the same rank of all interconnected systems. Each subsystem is, in turn, viewed as being made of one or several entities. In other words, each layer is made of entities, each of which belongs to one system. Entities in the same layer are termed peer entities.

In the OSI layering any layer is referred to as the (N) layer, while its next lower and next higher layers are referred to as the (N-1) layer and the (N+1) layer, respectively as illustrated in figure (3). The same notation is used to designate all concepts relating to layers, e.g., entities in the (N) entities.

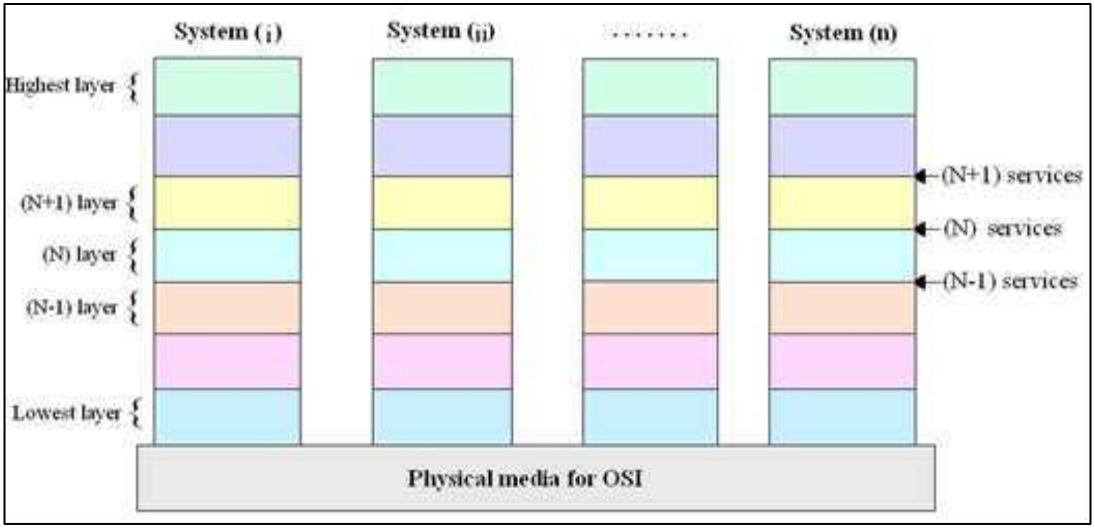

Fig. 3: illustrates of OSI layering

The basic idea of layering is that each layer adds value to services provided by the set of lower layers in such a way that the highest layer is offered the set of services needed to run distributed applications. Layering thus divides the total problem into smaller pieces. Another basic principle of layering is to ensure independence of each layer by defining services provided by layer to the next higher layer, independent of how these services are performed. This permits changes to be made in the way a layer or a set of layers operate, provided they still offer the same service to the next higher layer.

IV. Principles of ISO for the Seven Layers of the OSI Architecture

ISO determined a number of principles to be considered for defining the specific set of layers in the OSI architecture, and applied those principles to come up with the seven layers of the OSI Architecture.

Principles to be considered are as follows-

1) Do not create so many layers to make difficult the system engineering task describing and integrating these layers.
2) Create a boundary at a point where the services description can be small and the number of interactions across the boundary is minimized.
3) Create separate layers to handle functions which are manifestly different in the process performed or the technology involved.
4) Collect similar functions into the same layer.
5) Select boundaries at a point which past experience has demonstrated to be successful.
6) Create a layer of easily localized functions so that the layer could be totally redesigned and its protocols changed in a major way to take advantages of new advances in architectural, hardware, or software technology without changing the services and interfaces with the adjacent layers.
7) Create a boundary where it may be useful at some point in time to have the corresponding interface standardized.
8) Create a layer when there is a need for a different level of abstraction in the handling of data, e.g., morphology, syntax, semantics.
9) Enable changes of functions or protocols within a layer without affecting the other layers.
10) Create for each layer interfaces with its upper and lower layer only.
11) Create further subgrouping and organization of functions to form sublayers within a layer in cases where distinct communication services need it.
12) Create, where needed, two or more sublayers with a common, and therefore minimum, functionality to allow interface operation with adjacent layers.
13) Allow by passing of sublayers.

V. BRIEF EXPLANATION OF HOW THE LAYERS WERE CHOSEN

1) It is essential that the architecture permits usage of a realistic variety of physical media for interconnection with different control procedures. Application of principles 3, 5, and 8 leads to identification of a physical layer as the lower layer in the architecture.
2) Some physical communications media (e.g., telephone line) require specific techniques to be used in order to transmit data between systems despite a relatively high error rate (i.e., an error rate not acceptable for the great majority of applications). These specific techniques are used in data-link control procedures which have been studied and standardized for a number of years. It must also be recognized that new physical communications media (e.g., fiber optics) will require different data-link control procedures. Application of principles 3, 5, and 8 leads to identification of Data link Layer on top of Physical Layer in the architecture.
3) In the Open Systems Architecture, some systems will act as final destination of data. Some systems may act only as intermediate nodes (forwarding data to other systems). Application of principles 3, 5, and 7 leads to identification of a Network Layer on top of Data link Layer. Network-oriented protocols such as routing, for example, will be grouped in this layer. Thus, the Network layer will provide a connection path (network connection) between a pair of transport entities.
4) Control of data transportation from source end system to destination end system (which need not be performed in intermediate nodes) is the last function to be performed in order to provide the totality of the transport service. Thus, upper layer in the transport-service part of the architecture is the Transport Layer, sitting on top of the Network Layer. This Transport layer relives higher layer entities from any concern with the transportation of data between them.
5) Inorder to bind/unbind distributed activities into a logical relationship that controls the data exchange with respect to synchronization and structure, the need for a dedicated layer has been identified. So the application of principles 3 and 4 leads to the establishment of the Session Layer which is on top of Transport Layer.
6) The remaining set of general interest functions are those related to representation and manipulation of structured data from the benefit of application programs. Application of principles 3 and 4 leads to identification of a Presentation Layer on top of the Session Layer.
7) Finally, there are applications consisting of application processes which perform information processing. A portion of these application processes and the protocols by which they communicate comprise the Application Layer as the highest layer of the architecture.

The resulting architecture with seven layers is illustrated in figure (4).

VI. SEVEN LAYERS OF OSI ARCHITECTURE

The following figure illustrates OSI Reference Model, a seven layered OSI Architecture:

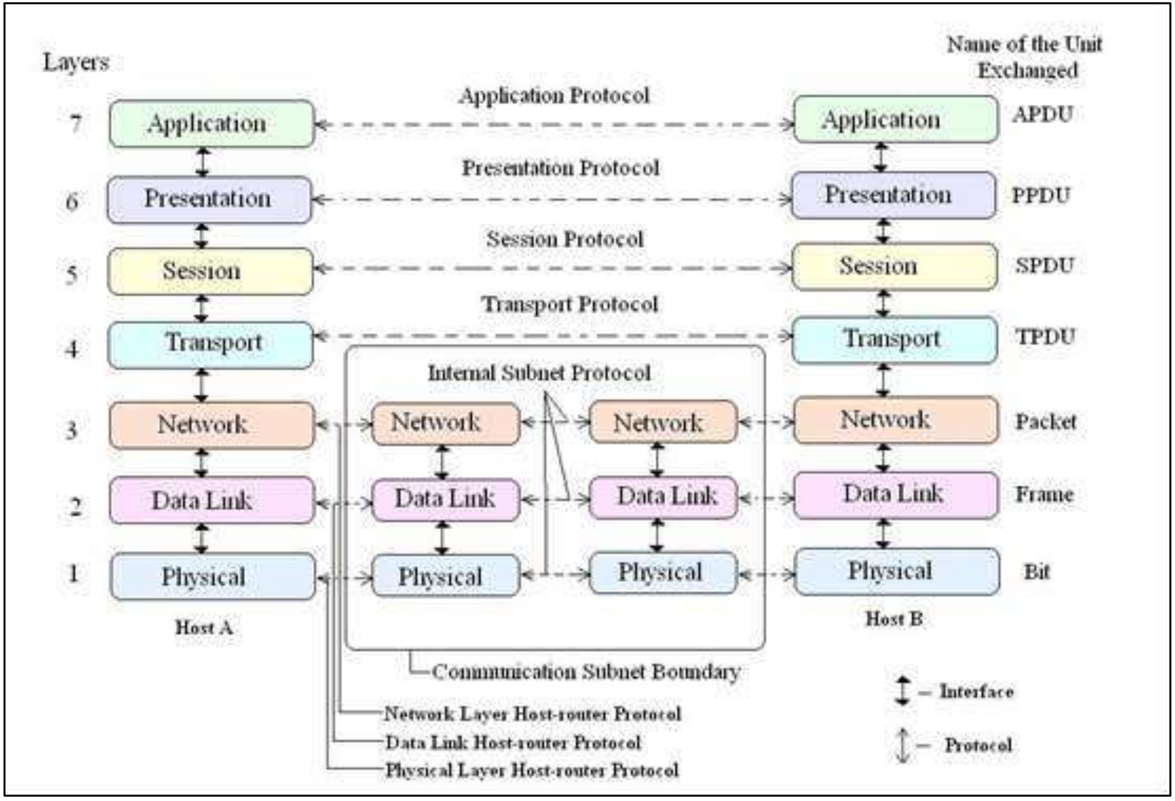

Fig. 4: illustrates OSI Reference Model - A Seven Layered OSI Architecture

Seven layers of the OSI Architecture:

A. *Physical Layer:*

The physical layer is responsible for individual bits from one node to another. It coordinates the rule for transmitting bits. It launches the raw bits in the channel or link.

This layer defines,

1) Physical network structures.
2) Mechanical and electrical specifications of the transmission medium.
3) Bit transmission encoding and timing rules.

The following network connectivity hardware are normally associated with physical layer are,

1) Hubs or Switches
2) Transmission media connectors
3) Modems

The major duties of physical layer is as follows,

1) *Physical Characteristics of Interface and Media:*

It defines the characteristics of the interface between the devices and transmission medium. It also defines the type of transmission medium used.

2) *Representation of Bits:*

The physical layer data consists of stream of bits(Sequence of 0‟s and 1‟s) without any interpretation. For transmission, bits are encoded into electrical or optical signals. The physical layer defines the type of representation (optical or electrical).

3) *Data Rate (or) Transmission Rate:*

The Sender and receiver must be synchronized at bit level.

4) *Synchronization of bits:*

The sender and receiver must be synchronized at bit level.

5) *Line Configuration:*

This layer is concerned with the connection of devices to the media. In a point-to-point configuration, two devices are connected through a dedicated link. In a multipoint configuration, a link is shared among several devices.

6) *Physical Topology:*

This defines how devices are connected to make a network. These devices can be connected by using a mesh, a star, a ring and bus topologies.

7) *Transmission Mode:*

This layer also defines the direction of transmission between two devices. Those two devices are simplex, half-duplex or full-duplex. Only one devices can be send in simplex mode, the other can only receive. It is a one way communication. Two devices can be send in half-duplex mode, but not at the same time. Two devices can send and receive at the same time in full-duplex mode.

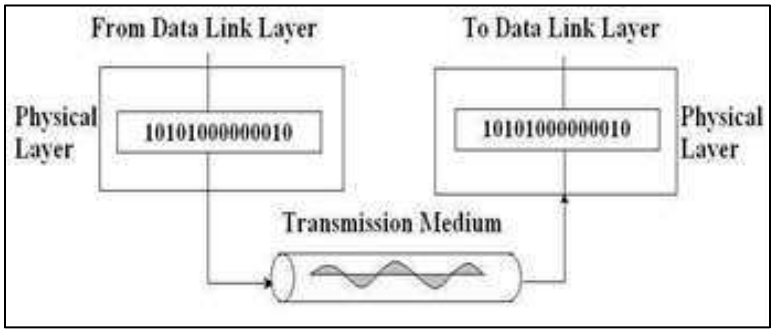

Fig. 5: illustrates data transmission in physical layer

B. *Data Link Layer:*

This layer is responsible for the two party communications by exchanging frames between the two nodes. It describes methods for moving information between multiple devices within the same logical network based on physical device addressing.

It makes the physical layer, a raw transmission facility, to a reliable link. It makes the physical layer error free to upper layer (network). The basic purposes for Data Link Layer protocol implementations are,

1) Organize Physical Layers bits into logical groups of information called frames.
2) Detect errors.
3) Control data flow.
4) Identify computers on the network.

Data link layer protocols,

1) HDLC (High-Level Data Link Control)
2) Frame relay

- The protocol packages the data into frames that contain source and destination addresses. These frames refer to the physical hardware address of each network card attached to the network cable. Ethernet, Token Ring, and ARCnet (Attached Resource Computer network) are examples of LAN(Local Area Network) data link protocols. If communication extends beyond the LAN onto the Internet, the network might use other data link protocols, such as Point-to-Point Protocol (PPP) or Serial Line Internet Protocol (SLIP). The data link layer sends block of data with necessary synchronization, bit error detection/correction error control, and flow control. This control of data flow controls approximately 70 percent of all error handling. Since the physical layer merely accepts and transmits a stream of bits without any regard to the meaning of the structure, it is up to the data link layer to create and recognize frame boundaries. This can be accomplished by attaching special bit patterns to the beginning and end of the frame as shown in figure (6).

Data link layer protocols,

1) HDLC (High-Level Data Link Control)
2) Frame relay

- The protocol packages the data into frames that contain source and destination addresses. These frames refer to the physical hardware address of each network card attached to the network cable. Ethernet, Token Ring, and ARCnet (Attached Resource Computer network) are examples of LAN(Local Area Network) data link protocols. If communication extends beyond the LAN onto the Internet, the network might use other data link protocols, such as Point-to-Point Protocol (PPP) or Serial Line Internet Protocol (SLIP). The data link layer sends block of data with necessary synchronization, bit error detection/correction error control, and flow control. This control of data flow controls approximately 70 percent of all error handling. Since the physical layer merely accepts and transmits a stream of bits without any regard to the meaning of the structure, it is up to the data link layer to create and recognize frame boundaries.

This can be accomplished by attaching special bit patterns to the beginning and end of the frame as shown in figure (6) T2 (trailer of layer 2) and H2 (header of layer 2) are attached at the beginning and ending of the frame.

Encryption can be used to protect the message as it flows between each network node. Each node then decrypts the message received and re-encrypts it for transmission to the next node.

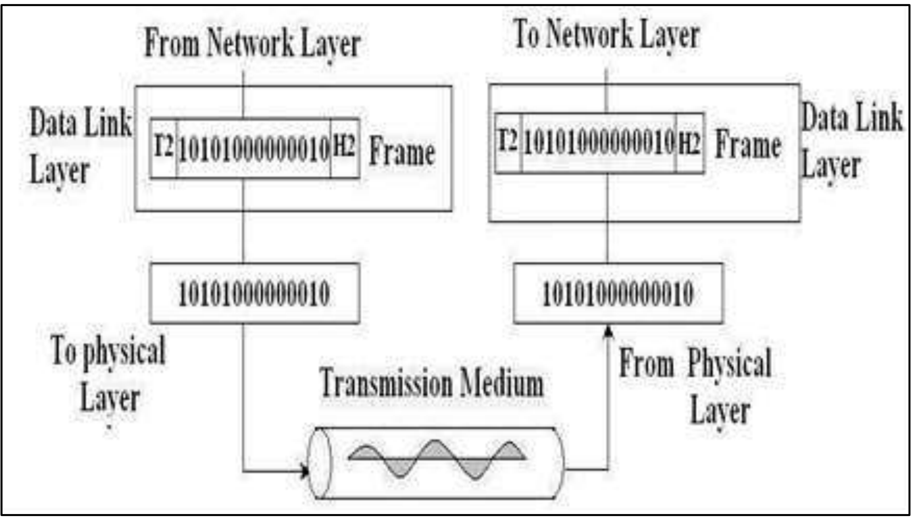

Fig. 6: illustrates Data Link Layer

This layer is subdivided into two sub-layers:

1) Logical Link Control (LLC):

This sub layer functions include – Managing frames to upper and lower layers, error control and flow control.

2) Media Access Control (MAC):

The MAC sublayer carries the physical address of each device on the network. This address is more commonly called a device‟s MAC address. MAC address is a 48 bit address which is burned into the NIC card (Network Interface Card) on the device by its manufacturer. NICs have a MAC address. A switch uses this address to filter and forward traffic, helping relieve congestion and collision on a network segment.

The major duties of the Data link layer are,

a) Framing:

The data link layer divides the stream of bits received from network layer into manageable data units called frames.

b) Physical Addressing:

The data link layer adds a header to the frame to define the sender and receiver of the frame. If the frame is for some system outside the network, then the receiver address is the address of the intermediate connecting device.

c) Flow Control:
The data link layer imposes flow control mechanism to prevent overwhelming the receiver, i.e., if the rate at which the data is absorbed by receiver is less than the rate of which it is produced at sender.
d) Error Control:
Error control is normally achieved by adding a trailer to the end of frame, which contains check sum. It is used to increase the reliability of the system.
e) Access Control:
When two or more devices are connected to the same link, then medium access mechanism should be evolved to determine which device has control over the link at any given time.

C. *Network Layer:*

The network layer is responsible for the delivery of packets from original source to final destination. Before receiving the packets to the respective destination, buffering of packets [7] [8] takes place in the network. This layer takes care of addressing and routing issues. It describes methods for moving information between multiple independent networks based on network layer addressing.

The data link layer oversees the delivery of the packet delivery between two systems on the same network, whereas the network layer that each packet gets to the final destination. If systems are connected on same link, there is no need for network layer. If systems are attached to different networks then it is inevitable. This layer defines,
1) Logical Network structures and addressing.
2) Route discovery and selection.
3) Network layer flow control and error control.
4) The network connectivity hardware associated with network layer.
5) Routers.
6) Network layer protocols.
7) IP (Internet Protocol) etc.

The major duties of network layer is,

1) Logical Addressing:
The physical address implemented at data link layer handles the addressing problem locally. If packet passes the network boundary, then we need logical address. The network layer adds a header to the packet coming from upper layer with source and destination logical address.

2) Routing:
In a large network, the connecting devices (router or switches) route or switch the packets to their final destination.

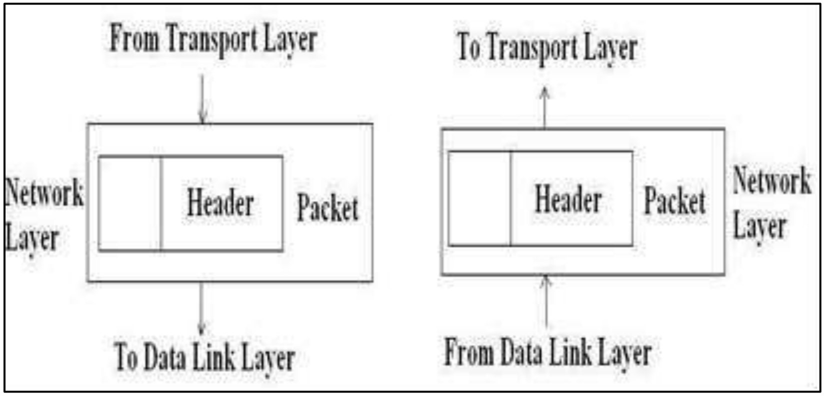

Fig. 7: illustrates Network layer

D. *Transport Layer:*

The transport layer is responsible for delivery of a message from one process to another. It is responsible for process to process delivery of the entire message.

The network layer ensures host to destination delivery of individual packets. It does not recognize the relationship between the packets. It does not recognize the relationship between the packets. So the transport layer ensures that the whole message arrives intact and in order, overseeing both error control and flow control at process-to-process level. It is responsible for end-to-end connection between the source and the destination. The name of the data unit in transport layer is TPDU (Transport Protocol Data Unit).

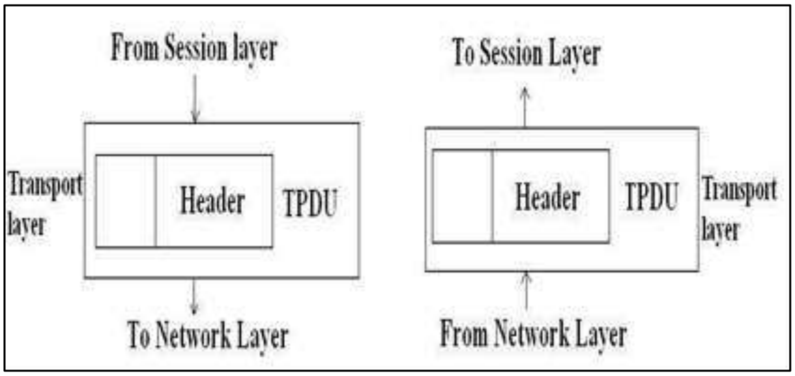

Fig. 8: illustrates Transport Layer

This layer defines,
1) Connection and transaction identifiers.
2) Segment development.
3) Connection services.
4) TCP
5) UDP etc.

The duties of transport layer are,

1) *Port Addressing:*

Computers often run several processes (running programs) at same time. So to ensure process-to-process delivery, the transport layer header (H4) includes port address which specifies the process.

2) *Segmentation and reassembly:*

If the message is big, it is divided into transmittable segments. Each segment contains a sequence number, which enables the transport layer to reassemble the segments upon arrival at the destination. It is also used to identify the packet lost during transmission.

3) *Connection Control:*

The transport layer can be either connection-oriented or connectionless. A connectionless transport layer treats each segment as an independent packet and delivers to the transport layer of the destination machine. Whereas, a connection-oriented transport layer establishes a connection with the transport layer of the destination machine first before delivering the packets. After data is transferred, the connection is terminated.

4) *Flow Control:*

The flow control is done end-to-end rather than across a signal link.

5) *Error Control:*

The error control is done end-to-end rather than across a single link. The error connection is achieved to retransmission.

E. Session Layer:

The session layer permits two parties to hold ongoing communications called a session across a network. The applications on either end of the session can exchange data or send packets to another for as long as the session lasts. The session layer handles session setup, data or message exchanges, and tear down when the session ends. It also monitors session identification so only designated parties can participate and security services to control access to session information. A session can be used to allow a user to log into a remote time-sharing system or transfer a file between two machines.

The session layer has the option of providing one-or-two-way communication called dialogue control. Sessions can allow traffic to go in both directions at the same time, or in only one direction at a time. Token management may be used to prevent both sides from attempting the same operation at the same time. To manage these activities, the session layer provides tokens that can be exchanged. Only the side holding the token is permitted to perform the critical operation.

Another session service is synchronization. Consider the problems that occur when transferring a file between two machines and the system crashes not being able to complete the transfer. This process must be restarted from the beginning. To avoid this problem, the session layer provides a way to insert checkpoints into the data stream, so that after a crash, only the data after the last checkpoint has to be repeated.

It accepts the data from presentation layer and provides services to it and accepts the services of the transport layer. The name of data unit in the session layer is SPDU (Session Protocol Data Unit) or sessions. Therefore session layer functionality includes:
1) Virtual connection between application entities
2) Synchronization of data flow
3) Creation of dialog units
4) Connection parameter negotiations
5) Partitioning of services into functional groups.

6) Acknowledgments of data received during a session
7) Retransmission of data if it is not received by a device

F. Presentation Layer:

The presentation layer is responsible for the format of the data transferred during network communications. This layer is concerned with the syntax and semantics of the information transmitted. For outgoing messages, it converts data into a generic format for the transmission. For the incoming messages, it converts the data from the generic form to a format understandable to the receiving application. Different computers have different codes for representing data. The presentation layer makes it possible for computers with different representation to communicate. The presentation layer provides common communication services such as encryption, text compression, and reformatting.

The presentation layer is also concerned with other aspects of information representation. Data compression can be used to reduce the number of bits that have to be transmitted. Cryptography is frequently required for privacy and authentication. The name of data unit in the presentation layer is PPDU (Presentation Protocol Data Unit).

G. Application Layer:

The application layer enables the user, whether human or software to access the network. It provides user interface and support for services such as e-mail, remote file access and transfer, shared database management etc.

1) It accepts the services from presentation layer and data unit in this layer is called APDU (Application Protocol Data Unit).
2) File Transfer, Access and Management: Provides handling services in the network. This includes the movement of files between different systems, reading, writing and deletion of remote files, and management of remote file storage.
3) Network virtual Terminal: Provides services to access applications in different remote computer systems through stimulating a real terminal.
4) Electronic Mail and Messaging Handling: Facilitates the electronic exchange of documents.
5) Directory Services (DS): Provides services with the ability to match names with addressing information.
6) Common management Information Protocol: Provides services for network management.

VII. CONCLUSION

The development of OSI Standards is a very big challenge, the result of which will impact all future computer communication developments. If standards come too late or are inadequate, interconnection of heterogeneous systems will not be possible or will be very costly. The work collectively achieved so far by SC16 members is very promising, and additional efforts should be expended to capitalize on these initial results and come up rapidly with the most urgently needed set of standards which will support initial usage of OSI (mainly terminals accessing services and file transfers). Common standards between ISO and CCITT (Consultative Committee for International Telephony and Telegraphy) are also essential to the success of standardization, since new services announced by PTT‟s and common carriers are very similar to data processing services offered as computer manufacturer products, and duplication of now compatible standards could simply cause the standardization effort to fail. In this regard, acceptance of the OSI Reference Model by CCITT Rapporteur‟s Group on Layered Architecture for Public Data Networks Services is most promising. It is essential that all partners in this standardization process expend their best effort so it will be successful, and the benefits can be shared by all users, manufactures of terminals and computers, and the PTT‟s/common carriers.

REFERENCES

[1] Hubert Zimmermann, “OSI Reference Model- the ISO Model of Architecture for Open System Interconnection” IEEE transaction on communications, vol.28, issue 4, April 1980.
[2] J.Day, Zimmermann H, “The OSI reference model” published in IEEE vol.71, issue12,pages: 1334- 1340, 1983.
[3] H. Zimmermann “High level protocols standardization: Technical and political issues", Proc. ICCC, pp.373 - 376, Aug.1976.
[4] ISO/TC97/SC16, “Provisional model of open systems architecture”, DOC.N34, Mar. 1978.
[5] ISO/TC97/SC16, “Reference model of open systems interconnection”, 1979.
[6] H. Zimmermann and N. Naffah “On open systems architecture", Proc. ICCC, pp.669 -674, Sep.1978.
[7] Depavath Harinath, “Buffers in 802.11-Based Networks” International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE) ,Vol.2, Issue12, Dec 2012.
[8] Depavath Harinath, “Dynamic Buffer Sizing Algorithms for 802.11-Based networks” International Journal of Advanced Research in Computer Science (IJARCS),vol.3, No. 7, Nov-Dec 2012.
[9] J. H. McFayden “Systems network architecture: An overview", IBM Syst. J., vol. 15,no. 1,pp.4 -23,1976.