

Chapter 1

Computer Network Foundation

1.1 Computer network development

Since 1946, the first computer came out until the late 1950s, the computer can only support single-user use, all the resources of the computer for a single user occupied by the user can only use the computer to a fixed place (such as computer room). The emergence and use of computers for the development of human information civilization has brought a revolutionary opportunity, but alone on the computer itself. Although it can be in a very short time to deal with a lot of information, a single computer, can only worked on their limited range, even if its storage capacity is also very large and very limited. As a result, the need to exchange data between computers is required. In addition, each computer is equipped with external devices such as printers, how to save expensive external equipment, and causes computers to connect to each other over a communication channel so that to share the consideration of peripherals. Therefore, the computer network is the crystallization of computer technology and communication technology development. There are two main aspects of the combination of computer and communication. On the one hand, the communication network provides the necessary means for the data transmission and exchange between the computers. On the other hand, the development of digital computing technology infiltrates into the communication technology and improves the performance of the communication network.

According to the network architecture the development of the network can be divided into three times:

(1) Online terminal system

The early computer network is a host-centered online terminal network system as shown in Figure 1 – 1. It is a single host as the center of the star network, the terminal through the communication line to share the expensive central host hardware and software resources. In addition to the host system has the ability to deal with the data, the system connected to all the terminals have no data processing functions. At this time the essence of the computer network system is the online multi-user system, which is the terminal-oriented computer communication. In the system, the host should not only carry out data processing, but also bear the communication with the terminal, the host computer load increases, the actual work efficiency decline; and scattered terminals have to occupy a separate communication lines, communication line utilization is low, costly. To this end, in front of the system host to add a front-end processor FEP (Front End Processor), used to be responsible for communication work, in order to achieve the data

processing and communication control division of labor, to better play the central computer processing data capabilities. The structure shown in Figure 1 - 2. In the early 1960s, American Airlines built a computer and a number of more than 2,000 terminals in the United States composed of air booking system SABRE-1 is a typical computer communication network.

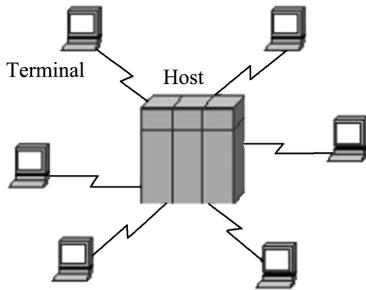


Figure 1 - 1 Focus on a single host

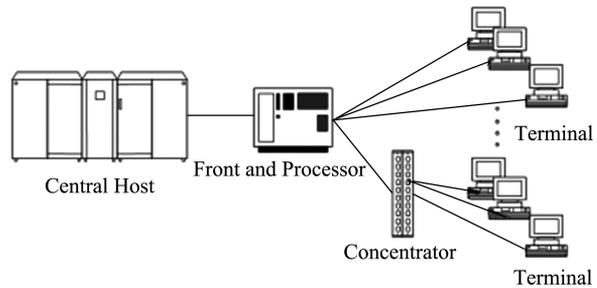


Figure 1 - 2 Online terminal system for FEP

(2) Host-host network phase

The early online terminal system has only one processor, and each terminal shares the hardware and software resources of the host through the communication line. Host-host communication computer network system, showing a number of host processing center features, the host through the communication line connection, mutual data, transfer software, to achieve the connection between the computer resource sharing. Figure 1 - 3 shows the host-host network structure.

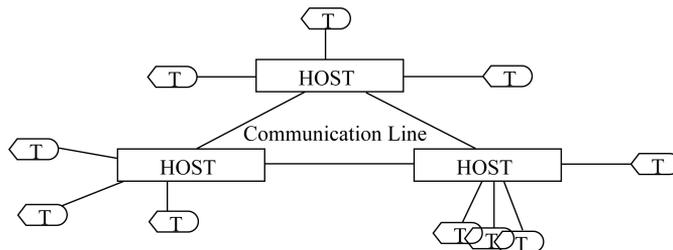


Figure1 - 3 Host-host network structure

In the early 1960s, the Advanced Research Project Agency, led by the US Department of Defense, proposed to develop a new network of survivability that could be adapted to modern warfare, with the aim of dealing with the Soviet Union The nuclear attack threat. We know that the traditional circuit switched circuit network has been in all directions, but during the war, once the communication circuit has a switch or a link was blown up, the entire communication circuit will be interrupted. If you immediately switch to other circuitous circuit communication, you must also re-dial to establish a connection. This will delay some time, but this can also cause irreversible heavy losses.

According to the then US military demand, this new type of network must meet some of the following basic requirements:

- (a) Unlike traditional telecom networks, this new type of network is not intended for telephony, but for data transfer between computers.
- (b) New networks can connect different types of computers, that is, not limited to a single type of

computer.

(c) All network nodes are equally important. Because the network must withstand the enemy's nuclear strike, so the network can not have some particularly important nodes, otherwise the enemy will first aim and destroy these important nodes. Will be all the nodes designed to be equally important, you can greatly improve the survival of the network.

(d) The computer in the communication, there must be redundant routing. When a node or link in the network is destroyed, the redundant route can make the ongoing communication automatically find the appropriate route, so that communication is maintained.

(e) The structure of the network should be as simple as possible, but can be very reliable transmission of data.

According to these requirements, the use of packet switching of the new computer network gave birth. In 1968, the world's first computer network-ARPANET was born. ARPANET main features are resource sharing, decentralized control, packet switching, the use of specialized communications control processor. Figure 1 -4 shows a typical ARPA network.

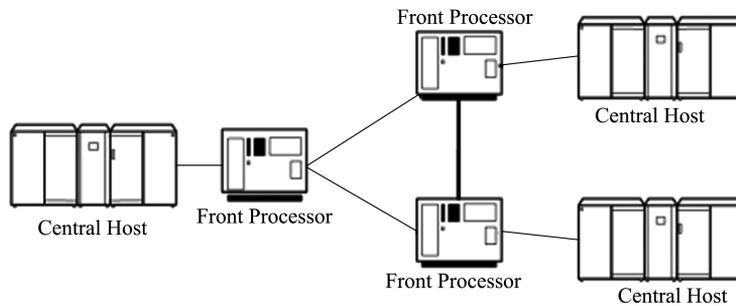


Figure 1 -4 A typical ARPA net

Here you can briefly introduce the concept of packet switching. Packet switching is the use of storage and forwarding technology. Figure 1 -5 shows the concept of grouping. The entire piece of data to be sent is called a message. Before sending a message, the longer message is divided into a smaller equal length segment. In front of each data segment, plus some of the necessary control information composed of the first part of the composition of a group. The head of the group becomes "Baotou". A packet is a unit of data that is transmitted in a computer network. The header of each packet contains control information such as address. The node switch in the packet switching network forwards the packet to the next node

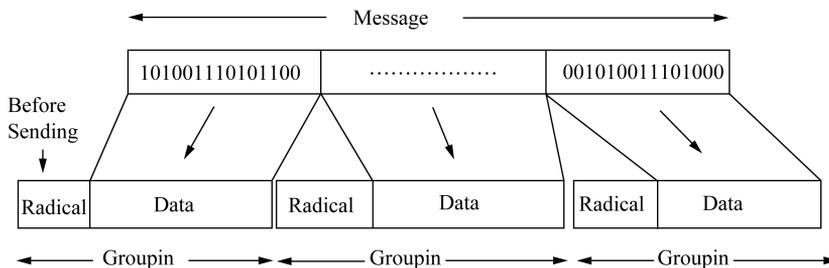


Figure 1 -5 The concept of packet

switch according to the address information in the header of the received packet. With such a store-and-forward method, packets can be delivered to the final destination. The receiver receives the packet and strives the header to restore the packet. Finally, at the receiving end, the received data is restored to the original message. (It is assumed that the packet does not have an error during the transmission and is not discarded at the time of forwarding.)

The diagram of the packet switching network is shown in Figure 1 – 6. Packet switching is the dynamic allocation of transmission bandwidth, the communication link is occupied by paragraph. To group the transport units and find the routes. And packet switching does not have to establish a connection to other hosts can send packets, so that the full use of the link bandwidth. Packet switching is highly reliable, has a perfect network protocol, adaptive routing protocol so that the network has a very good survivability. However, packet switching also brings new problems: packets need to be queued at each node store and forward, which can cause some latency. Moreover, the group must carry the head (which has essential control information) also caused a certain amount of overhead.

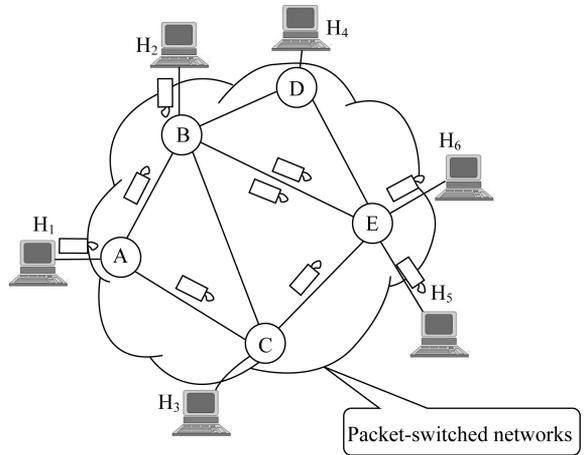


Figure 1 – 6 Packet switched networks and hosts

The success of ARPANET has led to a fundamental change in the concept of a computer network. The adoption of the new “store-and-packet switching” principle marks the beginning of a real computer network, resulting in a great development.

Later, the host-host network system system in the slow to the two forms of evolution, one form is the communication task from the host separation, adding communication control processor CCP dedicated to deal with the communication between the host task, the structure shown in Figure 1 – 7.

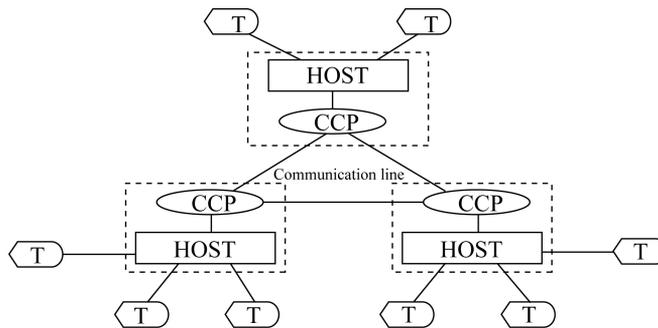


Figure 1 – 7 A host network for CPP

The transport network, which is composed of CCPs, is called a communication subnet and provides information transmission services for resource subnets. The set of hosts is called a resource subnet, providing a variety of network resources, based on the communication subnet (which can coexist more than

one system). Figure 1 – 8 is a network diagram centered on a communications subnet.

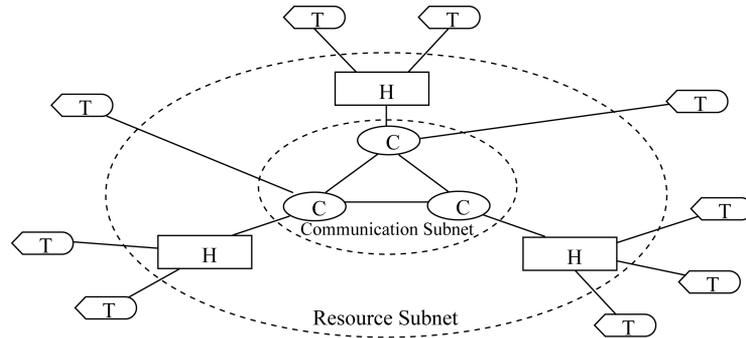


Figure 1 – 8 Network of communication subnets

The second evolution is that with the expansion of the communication subnet, the size of the communication subnet is also expanded by the private sector into the public, the original CPP has become a public data network. Its advantages are: to reduce the cost of building the user system, communication line utilization is high, good compatibility. The structure shown in Figure 1 – 9.

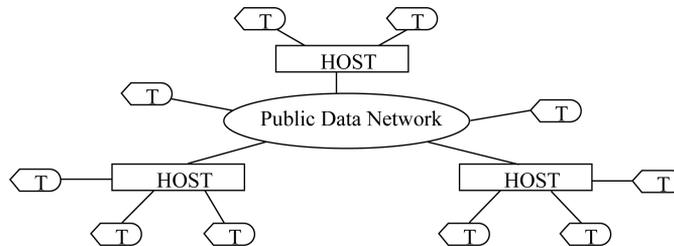


Figure 1 – 9 A network centered on public data communication networks

(3) Standardized network phase

The second generation of computer network is the lack of network popularity is low, the standard is not uniform, the network architecture research is not mature.

In the mid – 1970s, cheap personal computers (PC) came out, so that an enterprise or department can easily have one or more computers, there has been a local area network, and promote the development of computer networks. Compatibility and interoperability between different network devices is the driving force behind the standardization of network architectures. 70 years, entered the era of network protocol standardization, which is the beginning of the LAN era, which is representative of the network protocol is: OSI, TCP/IP.

Since the late 1980s, the most striking aspect of the network has originated from the rapid development of the Internet in the United States. Now the Internet has developed into the world's largest international computer Internet. As the Internet has affected all aspects of people's lives, which makes the 20th century, 90 years to become the Internet age, or simply referred to as the Internet age.

1.2 Computer network classification

1.2.1 Definition of computer network

More than 30 years of application and development, people on the computer network functions gradually have a deep impression. Many experts have tried to define the computer network itself, but because the scope of computer and computer networks is so extensive, this definition has never been unified.

The simplest definition of a computer network is the collection of some interconnected, autonomous computers. If this definition, the early terminal-oriented network can not be regarded as a computer network, but can only be called the online system (because then many terminals can not be regarded as autonomous computers). But with the decline in hardware prices, many terminals have a certain degree of intelligence, so “terminal” and “autonomous computer” gradually lost strict boundaries. Therefore, if the use of micro-computer as a terminal, according to the above definition, then the kind of early terminal-oriented network can also be called a computer network.

This book uses the following description of the computer network: computer network is shared resources (hardware, software and data, etc.) for the purpose of the connection, under the control of the agreement by one or more computer systems, a number of terminal equipment, data transmission Equipment and other components of the system set.

The computer network consists of network hardware, network software and network protocols. As shown in Figure 1 - 10. Network hardware typically includes workstations, servers, networking components, and communication media. Network software includes network operating system, management software, switching and routing software, network application software. The network protocol is attached to a certain network architecture, such as ISO-OSI reference model, TCP/IP and so on. The current dominant network architecture is TCP/IP.

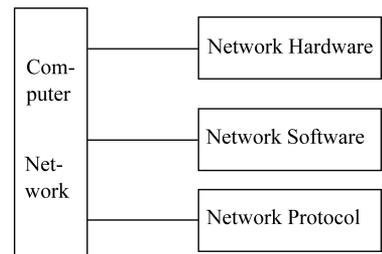


Figure 1 - 10 The composition of computer networks

1.2.2 Several different classification methods

(1) Sort by area of the network

The area of the network refers to the range that the network can cover. According to the coverage of the network, the network can be classified into wide area network, local area network, metropolitan area network and so on.

(a) Wide Area Network (WAN) also known as long haul network, WAN coverage is usually more than tens of kilometers, with large scale, large transmission delay characteristics. The task of the WAN is to send the data sent by the host over long distances. The links connecting the WAN switches are generally high-speed links and have large communication capacity.

(b) Local Area Network (LAN) also known as local area network, covering a few kilometers or



less, is now within the unit or within the building, often composed of a unit investment, with small, dedicated, transmission delay Small features. Local computers are typically connected to microcomputers or workstations via high-speed communication lines (typically 10 Mb/s).

(c) Metropolitan Area Network (MAN) metropolitan area network also known as the municipal area network, covering between the LAN and WAN. Such as covering a city that can span several blocks or even the whole city. Metropolitan area networks can be owned by one or several units, but they can also be a public facility used to interconnect multiple LANs. Metropolitan area network transmission rate higher than the local area network.

(d) Access Network (AN) access network, also known as local access network or residents access network, it is also in recent years due to the increase in demand for high-speed Internet users appear a network technology. The access network is the interface between the personal computer, the local area network and the metropolitan area network. Access network provides a variety of high-speed access technology, so that users access to the Internet bottleneck to a certain extent. Figure 1 – 11 shows the relationship between WAN, metro, access, and LAN.

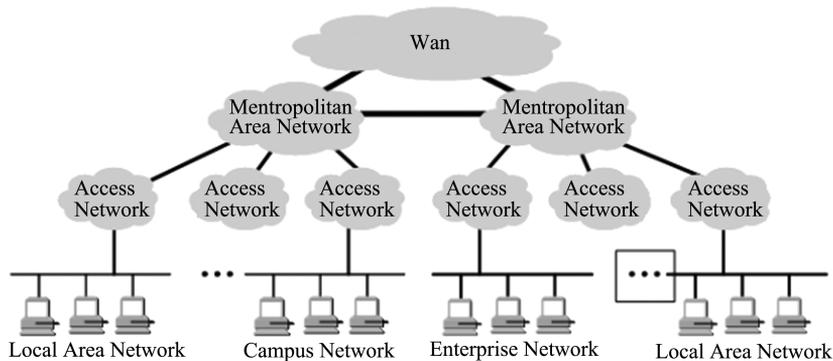


Figure 1 – 11 The relationship between wan, wans, access networks and lans

(2) According to the network topology

Network topology refers to the network equipment to connect the physical cable laying the form, common star, bus, ring and mesh shape.

(a) Star network; refers to the workstations connected to the star network. The schematic diagram is shown in Figure 1 – 12. That is, the network has a central node, other nodes (workstations, servers) are directly connected with the central node, this structure to the central node as the center, so also known as centralized network. It has a simple structure, easy to manage, simple control, easy to build network, network delay time is small, transmission error is low. But the shortcomings are also obvious, high cost, low reliability, poor ability to share resources.

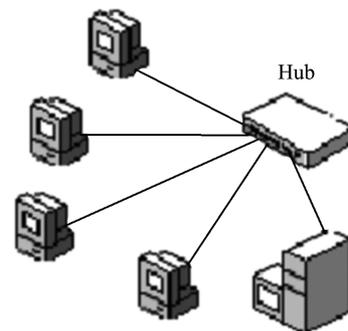


Figure 1 – 12 Star topology diagram

(b) Ring network: The structure diagram is shown in Figure 1 - 13. Some nodes in the network are connected to each other through a point-to-point link to form a closed loop. This structure makes the public transmission cable form an annular connection, and the data is transmitted in the loop along each direction in one direction. The node spreads to another node. The information flow in the network is moving in a fixed direction, the two nodes only one road, simplifying the control of the path selection; because the information source in the loop is serially through the various nodes, Point too much, it is bound to affect the information transmission rate, the network response time is extended, the loop is closed, not easy to expand, low reliability, a node failure, will lead to the whole network paralysis, maintenance difficult, Fault location is difficult.

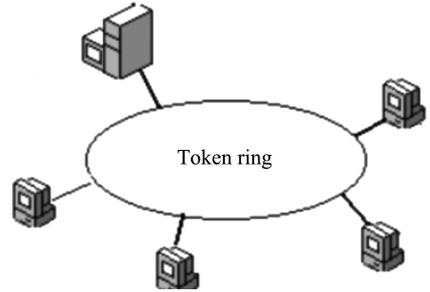


Figure 1 - 13 Circular topology diagram

(c) Bus network: the workstations and servers are linked to a bus, the status of the workstation equal, no center node control, public bus information on the baseband in the form of serial transmission, the transmission direction is always from the transmission of information Of the nodes began to spread to both ends, as the radio station to send the same information, the nodes in the acceptance of information when the address check to see if the address with their own workstation, consistent with the information received online. The structure diagram is shown in Figure 1 - 14.

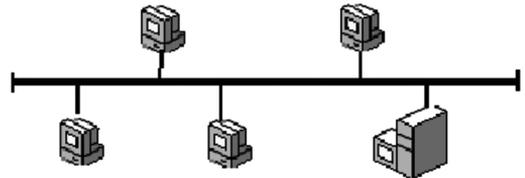


Figure 1 - 14 General linear topology diagram

Bus-based network features: simple structure, scalability is good. When you need to increase the node, only need to add a branch on the bus interface can be connected with the branch node, when the bus load is not allowed to expand the bus, the use of fewer cables, and easy to install, the use of equipment is relatively simple and reliable High, difficult to maintain, branch node fault finding difficult.

(d) Tree network: tree structure is a hierarchical centralized control network, compared with the star, its communication line length is short, low cost, easy to expand the node, find the path more convenient, but in addition to leaf node And its connected lines, any node or its associated line failure will make the system affected. The structure diagram is shown in Figure 1 - 15.

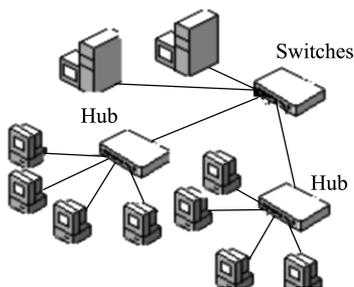


Figure 1 - 15 Tree topology diagram

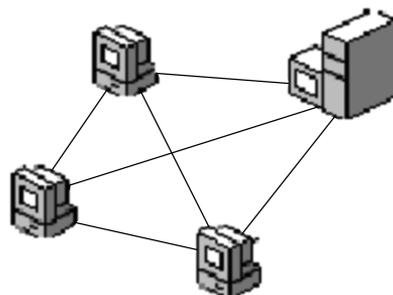


Figure 1 - 16 Network topology diagram

(e) Mesh network: In the mesh topology, there is a point-to-point link between each device in the network. This connection is not economical, and only every site has to send information frequently. It is also complicated to install, but the system is highly reliable and fault tolerant. Sometimes referred to as a distributed structure. The structure diagram is shown in Figure 1 – 16.

(3) From the network users to classify

(a) Public network, also known as the public network, refers to the country's telecommunications companies (state or private) funded the construction of large-scale network. The public means that all people willing to pay the cost of the telecommunications company can use.

(b) Private network A network in which a department is the need for special business work for the unit, which does not provide services to people outside the unit. For example, the military, railways, electricity and other systems have a dedicated network of the system.

(4) Sorting from the network switching function

The exchange is usually for network devices. Each network device typically has multiple input/output ports. The exchange means that the input signal of one port is forwarded (exchanged) to another port and transmitted to other devices through the line to which the port is attached.

(a) Circuit switching: the use of telephone work, with the establishment of links, data transmission and release links in three stages, the communication process, from beginning to end occupied the line, and does not allow other users to share its channel capacity.

(b) Message exchange: the switch with a "storage-forwarding" capacity of the computer, the user data can be temporarily saved in the switch, waiting for the line idle, and then a one-time transmission of user data.

(c) Packet switching: similar to the message exchange technology, but the provisions of the switch processing and transmission of data length (called the group), different user data packets can be interleaved in the network on the physical link transmission.

(d) Mixed exchange: refers to a data network in the use of circuit switching and packet switching.

(5) Classified by function

The ultimate goal of computer networks is application-oriented, so the computer network should also have the ability to provide information transmission and information processing. The information processing here is an action taken within the network in order to satisfy various applications, so that the transmission and processing can be divided into two parts from the logic.

(a) Communication subnet: the network for data transmission or data communication part of the collection of resources, mainly to support the transmission of user data, the subnet, including transmission lines, network equipment and network control centers and other hardware and software facilities.

(b) Resource subnet: network for data processing of the collection of resources, mainly to support the user's application. The subnet consists of the host resources of the user, including the user host accessing the network, and the application-oriented peripherals (such as terminals), software, and shareable data.

1.3 Three levels of computer network

1.3.1 The development of computer network system

In real life, hierarchical approach is widely used. In programming, usually a large procedure is divided into several layers of small modules to implement, such as the operating system. In the postal system, banking system, logistics systems and other large systems use a layered approach. A complex system is divided into a number of manageable subsystem, and then “divide and conquer”, a structural design method is a common means of engineering design. Hierarchical system is one of the best ways to decompose.

As show in Figure 1 - 17 are the hierarchy. n Layer is $n - 1$ Users, and $n + 1$ Service providers. $N + 1$ only use n layer provides services to, in fact it through n - layer also indirectly using the $n - 1$ layer and each layer is all of the following services.

Computer network is a very complex system. To illustrate this point, imagine a simple, connected two computers on the network to each other to transfer files.

Obviously, both computers must have a data transfer path between. But this is far from enough. There are at least some of the following work needs to be done:

- (1) Initiate communication computers data communication channel has to be activated (activate). The so-called “activation” is to send some signaling, ensure that computer data to be transmitted in this pathway on the right to send and receive.
- (2) To tell the network how to identify computers that receive data.
- (3) Computers that initiate communications must identify the other computer is ready to receive data.
- (4) Computers that initiate communication must be clear, in other files in the computer management program is to document receiving and storing file preparation.
- (5) If your computer’s file format is not compatible, at least one of these computers should complete the format conversion function.
- (6) For all kinds of errors and accidents, such as data transmission errors, duplicate, or missing, such as a node in a network switch is out of order, there should be reliable measures to ensure that the other party will eventually receive the correct file.

To designing complex computer networks, as early as in the first ARPANET Designed with a layered approach is proposed. “Tiered”-complex problem into smaller, single local problems be solved at different layers.

Network hierarchy of methods to resolve the problem:

Network should have which level? What is the function of each layer?

What is the relationship between the layers? How do they interact?

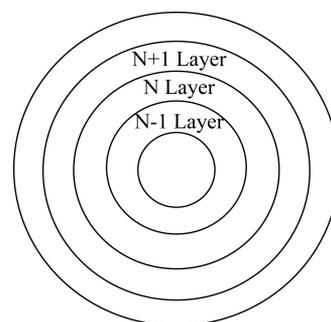


Figure 1 - 17 the hierarchy

Computer networking, layer, Protocol and interface are called computer network architecture. In other words, the architecture consists of three elements: the hierarchical structure and the function of each layer, indirect and services agreement.

Used level structure of system independence strong, that coupled degree low, upper just understand lower through layer between interface provides what service; adaptability strong, as long as service and interface not variable, each layer of achieved method can any change; easily achieved and maintenance, put complex of system decomposition into several a involved range small, and function simple of child unit, makes system of structure clear, achieved, and debugging and maintenance became simple and easy. So that designers can focus on the design and development of the module of interest.

1.3.2 The hierarchical principles of network architecture

1. Basic concepts

Entity is an abstract noun, any hardware or software process can send or receive information, in many cases, the entity is a specific software modules.

Peer layer, that is, two different systems of sibling hierarchy.

Peer entity, which are located in two entities in different systems of peer.

Interfaces, interaction between the two adjacent layers of interfaces, defined between two adjacent layers of operation and service of the lower to the upper.

Services, that is, one layer and the layer below a capacity, through the interface, to its adjacent upper.

The agreement, which both sides must abide by the rules in the communication. Network protocol has three elements:

- (1) Semantics (Semantics). Control information related to coordination and error handling.
- (2) Syntax (Syntax). Relating to data and control information formats, coding and signal level.
- (3) Timing (Timing). Related rate matching and sorting, and so on.

Under the control of the agreement, communications between two peer entity makes the layer up layer provides services to achieve this agreement, you also need to use the services provided by a layer below.

Protocol implementations provide services guaranteed to up one level. This service the user can only see service and unable to see the following agreement, following agreement on the above service is transparent to the user. Level of agreement is, in which the Protocol is to control the rules for communication between peers, but the service is vertical, that service is from the lower to the upper layer interface offers. In addition, all functionality is not completed within a layer are referred to as service, only those who can is higher see service functions can be called.

2. Network architecture and the Division of the principle

The benefit of this is that each layer of the hierarchy to implement a relatively independent function. Hierarchical structure also facilitates communication, understanding and standardization.

The so-called network architecture (Architecture) Is a collection of computer networks and agreements at

all levels. Hierarchical structures are generally expressed as vertical layering model (Figure 1 – 18).

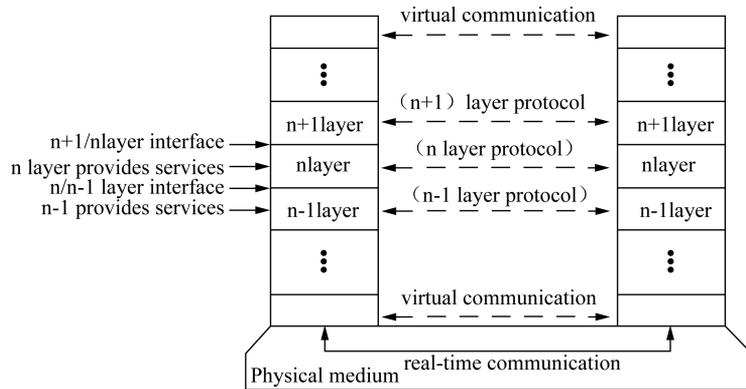


Figure 1 – 18 Hierarchical model of computer network

Elements of the hierarchy :

- (1) In addition to the physical media is on solid communication , the rest are virtual communication between peers.
- (2) Peer layer virtual communication must follow the Protocol.
- (3) n Virtual communication through n/n – 1 Layer at the interface between n – 1 Provide services as well as n – 1 Communication (and often virtual communications) to achieve.

Principles of hierarchies :

- (1) The function of each layer should be clear , and are independent of each other. When an update is a layer of concrete realization method , as long as you keep the upper and lower interface does not change , it will not have an impact on neighbors.
- (2) Indirect must be clear , across the interface of information should be as little as possible.
- (3) Layers should be moderate. If too few layers , resulting in each layer protocols are too complex if too many layers , the architecture is too complex , make it difficult to description and realization of functions of each layer.

Network architecture is characterized by :

- (1) To function as a basis for division level.
- (2) n Layer entities in the implementation of their own when the function is defined , you can use n – 1 Provide service.
- (3) n n + 1 When providing services , this service includes not only n Functions of the layer itself , also contain features that are provided by the underlying services.
- (4) Interfaces between adjacent layers only , and the specific implementation details of services provided on a layer of masking.

1. 3. 3 An open systems interconnection/reference purposes

1974 Years , United States IBM Development company announced its systems network architecture SNA(System Network Architecture) . This famous method of Web standards is the subject of the development. Yihou SNA and steadily improved , updated several versions. Now it is more widely used in the

world of a network architecture. Soon after, other companies have also launched a system structure of the company, and with a different name.

Due to the company's proposed network architecture independent of each other, although easy interconnection of the company's products, but was unable to, and other companies' products interoperate so that manufacturers and users are not satisfied.

As a user, hopes based on performance/Price than alternative products from different manufacturers and ensure that these new interoperability of devices and existing products to meet the application requirements.

As the manufacturer, and hope that their products have good versatility, has a larger market.

However the global economic development allowing different network architecture users urged to exchange information with each other. In order to interconnect different architecture of computer networks, the International Organization for Standardization ISO Yu 1977 year established specialized agencies to study the problem. Soon after, they made an attempt to make all kinds of computers in the worldwide standard framework for interconnection networks, known as the Open Systems Interconnection Basic reference model OSI/RM (Open System Interconnection Reference Model), referred to as OSI. "Open" means: as long as the OSI standard, a system column and place in the world, also follows this same standard to any other system of communication. It is like a telephone and postal systems around the world, these two systems are open systems. The "system" refers to the reality of each part of the system relating to the interconnection. Open Systems Interconnection reference model OSI/RM is an abstract concept. In 1983 years form the official documents of the Open Systems Interconnection Basic reference model, known as ISO 7498 to international standards, the so-called seven-layer protocol architecture.

As ISO itself, through establishment of standards, to achieve the following objectives:

- (1) To support the interconnection and communication between heterogeneous computer;
- (2) Can support a variety of communication media (offers a variety of communication and control procedures, support for different lines of business and communication requirements);
- (3) Can support a variety of business processes (interactive timesharing processing, remote batch processing);
- (4) To support advanced human-machine interfaces (graphics I/O, Text processing, voice recognition and synthesis);
- (5) With Extensible capabilities (support expanding application requirements and adapt to new technology, the development of new technologies).

Standards should be as inclusive as possible, all aspects should be covered as far as possible, and the realization of a comprehensive standard, not only requires more system resources and investments, but also for specific applications. In the 20 century 80 years, many large companies and even government agencies in some countries have expressed support for OSI. At that time seems to be in the near future the world will follow OSI develop standards to construct your own computer network. But 20 centuries 90 early, although the entire OSI international standards have been developed, but due to the Internet has a head start in the world covering a considerable range, but this time, hardly find what the manufacturers produce in line with OSI standards for commercial products.

Due to the above reasons, by the international standards organization ISO Network architecture developed by the international standards OSI/RM Have not been widely used, rather than the international standards TCP/IP Known as the de facto international standard.

1.3.4 OSI network architecture

Level divide is logically on the communication function of the Division. Levels should not be too few, so that each level is easy to implement and manage, not too much, otherwise the overhead of collection functions of each layer is too large. ISO in the stratified on the basis of the above, OSI/RM is defined as seven levels.

OSI Includes architecture, service definition and Protocol specification for three levels of abstraction. OSI of system structure defined has a seven layer model, to for process between of communications, and as a framework to coordination the layer standard of developed; OSI of service defined description has the layer by provides of service, and layer and layer Zhijian of abstract interface and interactive with of service original language; OSI the layer of agreement specification, Precisely defined which should be send control information and to explain the process of the control information.

Needs to be emphasized is that OSI Description of the reference model is not specific about the implementation, it just provides the conceptual framework for the standard machine. At OSI, various agreements were achieved, network device and OSI and unanimously when the agreement interconnects.

OSI Reference models, peer protocols exchanged between units of information called protocol data units (PDU, Protocol Data Unit). The transport layer and below the layer PDU also has its own specific name. Transport layer-the data segment (Segment), the network layer-packet (datagram) (Packet), the data link layer-the data frame (Frame), the physical layer-bit (Bit).

As Figure 1 – 19 As shown in, OSI Seven-layer model separately for the physical layer from bottom to top (Physical Layer, PH) Data link layer (Data Link Layer, DL), Network layer (Network Layer, N) And transport layer (Transport Layer, T), Session layer (Session Layer, S), Presentation layer (Presentation Layer, P) And application layer (Application Layer, A).

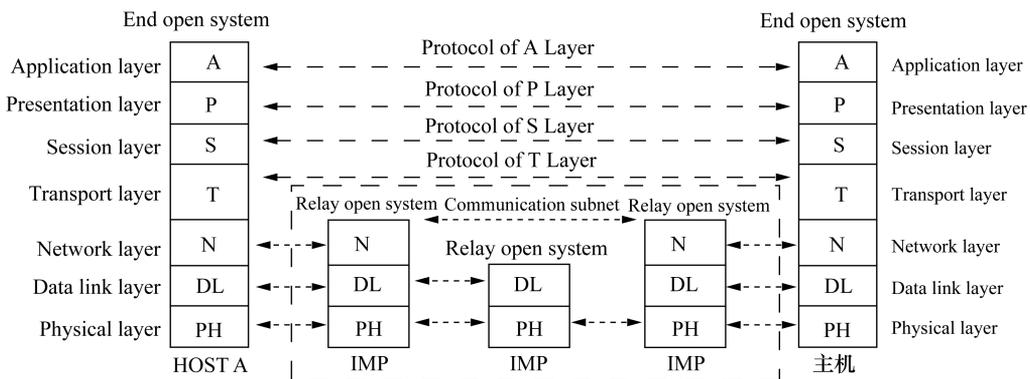


Figure 1 – 19 OSI The reference model

From Figure 1 – 19 Visible throughout the open system environment as the source and the host-side opening system and a number of open relay system through physical media connection. Open systems and



open relay systems, is an international standard OSI7498 in the terminology used. Informally, they are equivalent to the resources of the subnet host and communication nodes on the subnet (IMP). Only in the host may be required to include all seven levels of functions, but in communication subnet IMP usually require only the lowest three floors of even the lowest two layers function can.

OSI Reference model each layer relies on the next level of service. In order to provide services, lower the upper layer PDU as data encapsulation of this layer, then add the layer's head (and end). Head contains the control information needed to complete the data transfer.

In this way, data submitted from the top down is actually the process of continuous packaging process. Presented by the arrival from the bottom up process is constantly open process. Therefore, the data transmitted over the physical line, which is actually outside encapsulation multi layer "envelope".

However, a layer can only be recognized by the peer layer "envelopes", and is encapsulated in an "envelope" internal data only after opening will submit it to the top, this layer does not do any processing.

Hierarchy model of the actual transmission of data in Figure 1 – 20 As shown in. Sent to the receiving process and the process data in the chart is actually after the sender passed from the top to the bottom of each layer to the physical media through physical media after transmission to the receiver, then passed on each floor from bottom to top, and finally to the receiving process.

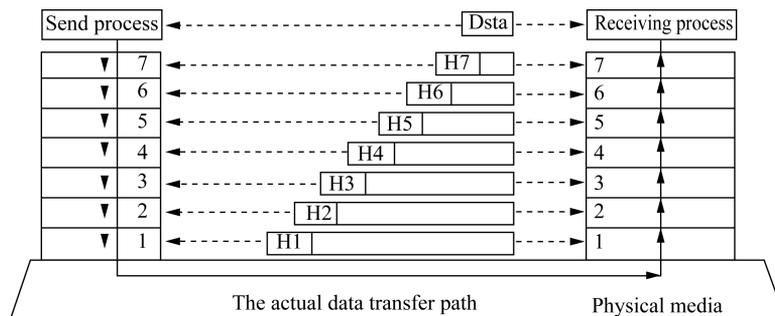


Figure 1 – 20 Actual data transfer process

On the sending side from top to bottom in the process of layer-by-layer transmission, each with appropriate control information diagram H7、H6、…、H1, Collectively known as the header. To the bottom by "0" or "1" component of the data bit stream, and then convert it into electrical signals transmitted over the physical media to the receiver. Passed up the receiver in the process is reversed when, to send a corresponding layer peeled layer by layer and control information.

Under the receiver does not receive a layer of control information for each layer, and high-level control information and is transparent to the data for it, so it only reads and removes the layer of control information and make the appropriate protocol operation. Peers of the sender and receiver see message is the same, just as if the information directly to each other through virtual communications.

Brief presentations by the function below:

(1) Physical layer—Defined as the establishment, maintenance and demolition machinery required for physical links, electrical, functional features and procedures, its effect is to make the origi-

nal data bits can be transmitted over the physical media. Connector-specific specifications, “0”, and “1” level, coordination of the sending and receiving of signals and so on.

(2) Data link layer—Bits are organized into data link protocol data unit (Often referred to as frames) And in its transmission, frame contains information such as the address, control, data and checksum. Data link layer’s main role is to pass the verification, validation and feedback means, no reliable physical link into the network layer for error-free data link. Data link layer also coordinate sending and receiving data rates, namely, flow control, to prevent the receiver from time to handle high-speed data sender and cause a buffer overflow and line congestion.

(3) The network layer—Data network protocol data unit (Group) Units for transmission. Concern is communication subnet of the network layer control, mainly deals with the how to make a data packet across subnets problem from the source to the destination, this requires routing in communication subnet. In addition, in order to avoid too much communication subnet network congestion caused by the grouping, you need to control the flow into the number of groups. When a packet spans multiple subnets to reach the destination, is to address Internet interconnection problems.

(4) Transport layer—Is the first end—The end, that is, host—Host level. Transport layer provides end-to-end transparency data transport service that makes top communication subnet of users do not exist, in a unified transportation primitive writing high-level software can run on any communication subnet. Also have to deal with end-to-end transport layer error control and flow control.

(5) Session layer—Is the process—Process level, its main function is to organize and synchronize communication between different processes on the host (Also known as dialogue). The session layer is responsible for dialogue between the two session entities in connection setup and teardown. In half-duplex case, session layer provides a data token to control when a party has the right to send data. Session layer also provides inserted into the data stream synchronization mechanism after the data transmission was interrupted due to a network failure, you don’t have to start from scratch and data re-transmission only recently after a synchronization point.

(6) The presentation layer—For upper-level users with common syntax transform data or information. In order for computers that use different encoding methods able to understand the contents of the data in a communication, abstract standard methods can be used to define your data structure, and using a standard encoded representation. Presentation layer management abstract data structures and internal representation of a computer representation of the criteria used in network communication. Data compression and encryption are also available in the presentation layer represents the transformation feature.

(7) The application layer is the highest level of the open systems interconnection environment. Different application layers for specific types of network applications provide access to OSI environment means. Different hosts in the network environment of file transfer access and management (FTAM), transmitting standard e-mail message handling system (MHS), the different types of terminals and hosts interactive access over the network virtual terminal (VT), the scope of application of the agreement.

From a communications perspective, each of the services can be divided into two major categories, that is, connection-oriented (connection-oriented) With no connection (connectionless).

**Table 1 –1 Of connection-oriented and connectionless services difference**

	Connection-oriented service	Connectionless service
Reference model	Telephone systems	Postal system
Features	Static allocation of resources, the transmission needs to establish a connection before	Dynamic resource allocation
Reliability	Providing reliable message flow services	Unable to prevent packet loss, corruption, duplication, and disorder
Destination address required	During the connection phase only needs to complete destination address	Need to provide each message with a complete destination address
Application	Sends a large number of messages to the same destination over time, real-time	A few sporadic messages
Classification and examples	1. And reliable message flow- File transfer 2. Reliable byte stream-Remote login 3. Unreliable connections-Digital voice	1. Data reported-Radio /Multicast 2. And reliable datagram-Registered mail 3. Request responses-Database query

1. Connection-oriented service

So-called connections, are two peers to communicate in a combination.

Connection-oriented service connection is established, data transfer and connection to release the three stages. Connection-oriented service, is before the data exchange, you must first establish a connection; when data exchange is over, you must terminate the connection. When transferring the data is transmitted in a sequence. Connection-oriented service is suitable in many messages to be sent to the same destination within a certain period. For sporadic sent a short message, connection-oriented service becomes too expensive.

2. Connect to the service, without

In the case of the connectionless service, communications between the two entities do not need to set up a connection, so the resources below, you don't need to book reservations in advance. These resources will be allocated dynamically when data transfer. Connectionless service another feature is that it does not require the communication of both entities are active simultaneously. When the sending entity is sent, it must be active. Receiving entity and does not have to be active. Only if the receiving entity is accepted, it must be active.

Advantages of connectionless service is convenient and relatively quickly. No connection service cannot prevent packet loss, duplication, or disorder.

Table 1 – 1 Illustrates a connection-oriented service and connectionless service difference in two ways.

1.3.5 OSI and TCP/IP architecture

In addition to OSI/RM Zhiwai, Also popular in practical application of some other network architecture, The most famous is the United States used the Internet TCP/ IP System.

Figure 1 – 21 Drew TCP OSI The comparison of two kinds of architecture. Is on the left in the picture has become history of the OSI architecture. Right now the Internet using TCP architecture (but also the following network interface layer is divided into two levels, namely, network interface layer and the physical layer. Thus become a five-layer architecture).

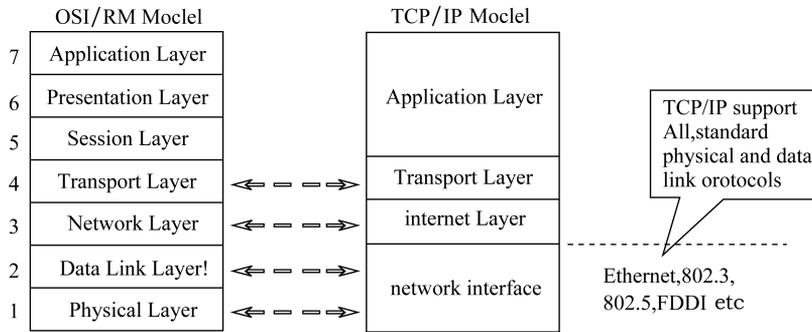


Figure 1 – 21 OSI TCP/IP Correspond to the reference model

TCP/IP Required by the application layer to provide users with a variety of services. Its main services are provided by remote login (Telnet), file transfer (FTP), e-mail, and, as shown in Figure 1 – 22 is shown.

TCP/ IP The transport layer provides for the application-End communication. Transport layer 2 protocols, Transmission Control Protocol (TCP, Transmission Control Protocol), User Datagram Protocol (UDP, User Datagram Protocol), as in Figure 1 – 23 is shown. TCP Protocol to establish a connection to a high reliability of message transfer purposes, it is responsible for a large number of user data according to a certain length packets are sent, and received packets in the decomposition order restructuring and recovering user data. In order to complete and reliable data transfer tasks, TCP Protocol with packet sequence control, error detection, test and then send the control functions. Table 1 – 2 lists the TCP segment format. UDP Protocol provides connectionless packet delivery service, it puts the user data into multiple packets and send to the receiver. However, theUDP Protocol to make the connection, packet sequence control, and traffic control, and other functions. Data transfer reliability guaranteed by the user program. UDP Protocol to perform code small, low overhead and high processing speed and so on.

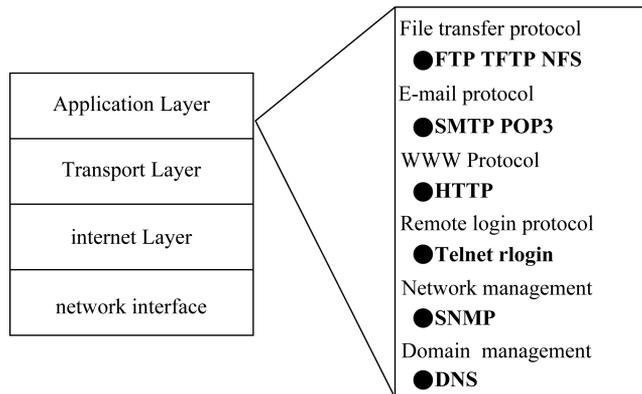


Figure 1 – 22 TCP/IP Application layer

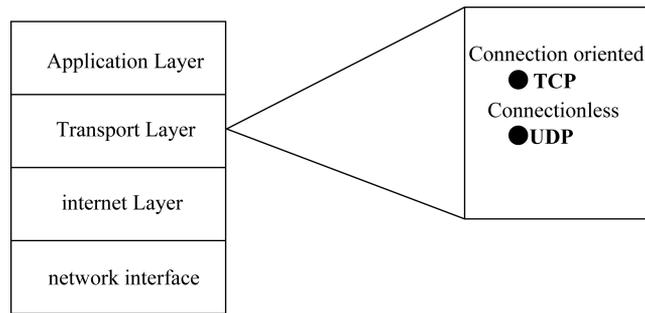


Figure 1 – 23 TCP/IP The two protocols of the transport layer

Interpretation of each field is as follows:

Source port (Source Port) Call port number;

Destination port (Destination Port) : Is called port number;

Sequence number (Sequence Number) : The sequence number of the first byte of the data;

Confirmation number (Acknowledgment Number) : Looking forward to the next sequence number;

Header length (HLEN) : 32 The length of the header bytes;

Reserved domain (Reserved) : Is set to 0;

Coding (Code Bits) : Used to control transmission (such as session establishment and suspension)

Includes: URG , ACK , PSH , RST , SYN , FIN Six-bit;

Window size (Window) : The receiver can continue to receive the number of bytes;

Checksum (Checksum) Include: TCP Header checksum and data;

Urgent pointer (Urgent Pointer) : The current offset of the sequence number to the emergency location data;

Options (Option) : For additional information, general application has been omitted.

Data (Data) : Upper layer protocol data.

Table 1 – 2 TCP Paragraph format

0 3 4 9 10 16 17 31

Source port			Destination port			
Sequence number						
Confirmation number						
Header length	Reservation	Bits of code		Window size		
Checksum			Urgent pointer			
Options (optional)						
Data						

Port is the application layer and the transport layer interface. Application to invoke the service, it is necessary to use specific ports, port number is 16 bits. TCP and UDP use port (socket), pass the information on to the upper. Port indicates the upper layer protocol being used. 1 – 24 lists some of the commonly used port number and the associated services.

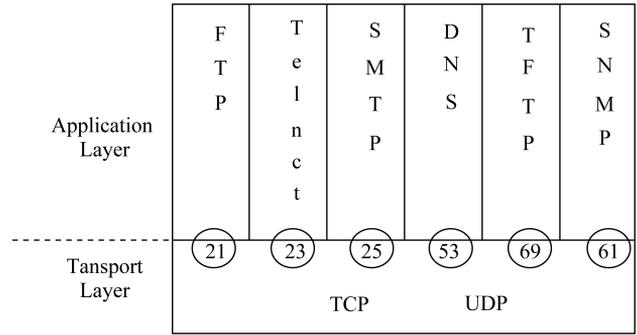


Figure 1 – 24 Commonly used port numbers and their corresponding service

As we all know, reserved port range 1 – 1023: < 255, As a public application; 255 – 1023, As the company; > 1023 Are not specified.

TCP/ IP Internet layer protocol is Internet Protocol IP(Internet Protocol). Assemble it to the transport layer to send messages into IP packets, and IP packets to the network layer. IP Protocol provides a unified IP packet format, to eliminate differences in network layer communication networks, thus for the sender and receiver of information with the alpha channel. Figure 1 – 25 shows the four main protocols of the Internet layer.

TCP/IP The network interface layer protocol and OSI Protocol physical layer, Part of the data link layer and network layer corresponds to. This layer to the communication protocols used in a subnet inherent in the Protocol, such as Ethernet, 8802.3 protocols, packet-switching network x.25 agreements. Network layer is transferred via the Internet processed messages.

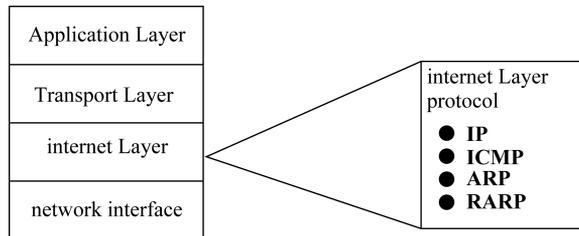


Figure 1 – 25 TCP/IP Internet layer four protocol

1.4 Theory of computer networks

1.4.1 Theoretical research of computer network

Computer network technology is 20th centurie one of the most rapidly developing technologies, is also very much related to the computer network teaching material, this textbook will be computer-related course is divided into three main levels: computer network (undergraduate), principles of computer networks and protocols (undergraduate and graduate), network theory (graduate students). Network theory course in computer studies, design and applications continue to produce new theories and new technologies, and graduate-level teaching contents and adapt, enhance the level of graduate courses.

Theory of computer networks is Research computer networks, in the planning, design, implementation, testing, operation, use, maintenance and management of typical problems and solutions in the process of theory. It is a synthetic science, in addition to direct use of communication theory and technology, computer science and technology, as well as logic, model theory, operational research, statistics, graph theory, information theory, control theory, simulation, neural networks, artificial intelligence, cognitive science, and other disciplines. In view of the huge computer network and wide contacts, involving multiple factors, usually divided into specific issues, highlighting the main factors, ignored or weakened secondary factors, and summarize, abstract, establishment of typical models to be examined.

Optimization design of computer networks is when setting up a computer network, first to resolve specific problems and theoretical issues. Purpose is in meet application needs and objective constraints conditions Xia, to at least of input (including human, and material, and financial, and time,), design, and built a security, and reliable, and effective, and run good, and adaptability strong, and easy management, and easy maintenance, and easy transformation, and easy expanded of computer network, and is expected to answered funds recycling term and may get of maximum of social benefits and economic, problem. Optimized design is divided into 3 phases: ① demand analysis and planning stage. Needs and circumstances should be investigated, collect, collate the necessary information and data, including application format, information, communications, response time, error rate and reliability requirements, selection criteria, and the distribution of existing devices, users, geographical features, natural condition, meteorology, outside influences, to clarify needs, identify the key links, planning the overall outline of the project. II stage of network design. In survey analysis of based Shang, should according to application needs, determine network of general framework and important of network parameter, must on some important of key problem made choices, as selection what topology structure, equipment of selection, and placed and connection method, communications media of select, and line layout and capacity distribution, communications procedures and routing, and flow and errors control technology, network business of type, and service quality and the senior agreement of select, . ③ design evaluation phase. According to the evaluation objectives, creation of mathematical models (such as predictive models, Optimization modeling, performance evaluation model, etc), so that network performance, cost, duration, time limits and benefit estimates, capital recovery period of analysis and evaluation are given technical and economic feasibility of conclusions. If those conclusions do not meet expected requirements, depending on the circumstances, some or all network optimization again.

Computer network architecture is a set of planning, design and set up computer networks required to follow the principles and basis, including hierarchies, function, protocol specifications, process descriptions, and so on. Network architecture for computer network development of the most influential international organization for Standardization (ISO) recommends that the open System Interconnection (OSI) reference model. It is through the system model and service definition and Protocol specification for 3 levels of abstraction step by step, gradually thinning being developed and described. Architecture model is the OSI highest level of abstraction, it functions and building an abstract concept level, hierarchical system models, depicting the opening performance of the whole system, structure factors, behav-

ioral characteristics, hierarchies, data formats, and so on. OSI architecture model by the application layer, presentation layer, session layer, and transport layer, network layer, data link layer and physical layer 7 layer. Service definition is OSI low level of abstraction, it defines in more detail each layer provides services provides, each layer features and abstract interfaces, but contained no details of whether and how to implement it. Protocol specification is OSI the lowest level of abstraction, it accurately defines a layer entities to work together and interaction semantics and syntax of the send control information required, as well as what kind of procedures to analyze, interpret, and process them. Architecture model for further development trends study, development of network applications architecture model in order to create a good environment for network users and the development environment. For example, some Internet experts in the OSI model on the basis of an open application architecture (OAA) model assumptions. OAA by the operating environment and development environment consists of two parts.

Routing, flow control and congestion control is also a popular research topic in the field of early computer, multiple, large amount of literature. Choice's primary objective is to choose the best path in the network, the source site sends the message fast and effective delivery to the destination site, focusing on improving network quality of service, reduced latency, and reduce transmission costs. Standards for measuring path selection algorithm include: ① the message information in the shortest time, shortest path, or minimal cost, delivery to the destination. ② algorithm is simple and easy to implement, flexible (able to network failures, and the impact of structural changes). ③ not overweight increase the overhead and nodes of the network (including the amount of processor time, storage capacity, and information transfer). It helps improve network performance, stable throughput rate and the average delay time, balanced network load, and so on. Spread the typical path selection algorithm for path selection, random choice, fixed-route selection, Adaptive Path, and so on.

Flow control and congestion control is designed to control network and the communication traffic on the line, keep the work of the network is in a steady state, in order to improve throughput and reduce the average delay time, focusing on improving network efficiency and resource utilization, congestion and prevent deadlocks. Flow control can be divided into between adjacent nodes flow control, between the source node and the target node flow control, host and node flow control, flow control between the host and the host of four types. Commonly used control methods have limited transfer rate and refuse to send window size limit, pause, send, receive, booking of buffer, and so on. Methods for congestion control is booked for buffers, limit the pipeline flow, network license, reverse inhibition.

In addition, error control is also an important research topic in the network design, which is based on application requirements, line quality, equipment performance, and environmental factors, choosing the appropriate control mechanisms and methods, identifying and correcting errors in information transmission, to be reduced to within the allowed level. In computer networks, usually by two basic strategies for handling errors in information transmission: ① using error-correcting codes. Information that you want to send message add enough redundant information, the receiver can not only identify, but also to correct errors in the information message. Because of too much redundant information, and control of complex, often used for one-way transmission, or as a secondary measure. ② use error detection



code. Information that you want to send message add a certain redundancy in information, enables receivers to detect errors in the information message (but I do not know what kind of error), and notify the sender to retransmit the information message. Communication protocols and network protocol often uses this approach.

Protocol engineering is one of the most active in the field of computer networking, aim of software engineering principles and methods for description, implementation, and verification of computer network protocols that work. Agreement the main contents of the project includes 3 aspects: ① Protocol formal specification and its formal description language. ② automatic generation of protocol software and development tools. ③ Protocol conformance testing technique and testing tools. Protocol engineering study helps deepen our understanding of computer networking protocols, helps to increase the productivity of software, helps to improve the maintenance and management of network protocol software level. However, compared to the Protocol engineering and software engineering, in both depth and breadth of research, development, application, said there is distance, there are broad prospects for development, development.

1.4.2 The new theory and technology of computer network

1.4.2.1 Grid technologies

Supercomputers are usually made up of hundreds of thousands of more processors, can calculate the common PC Machine and server computers that cannot complete large complex projects. If adult normal computer operation faster than walking speed, then the supercomputer reached a speed of rocket. With the development of supercomputer processing speed, it has also been an increasing number of applications in the industrial, scientific and academic fields. Figure 1 – 26 shows the Supercomputing Center. Although the supercomputer's processing power is very strong, but its cost is extremely high, cannot share the use of resources, limited to a few sectors at the national level, such as space, meteorological departments have such equipment, not user-friendly, the application may be unable to perform. Now it is higher requirements for computer processing power, but supercomputers cost apparently prevented it from entering areas of the work of ordinary people. So people started looking for a low cost and data processing capacity of superior computing model, in this needs grid technology was born.



Figure 1 – 26 Supercomputer Center

About 1910 Year of the grid concept, today When in power, regardless of what it is, where is. Figure 1 –27 shows a schematic of the power grid.

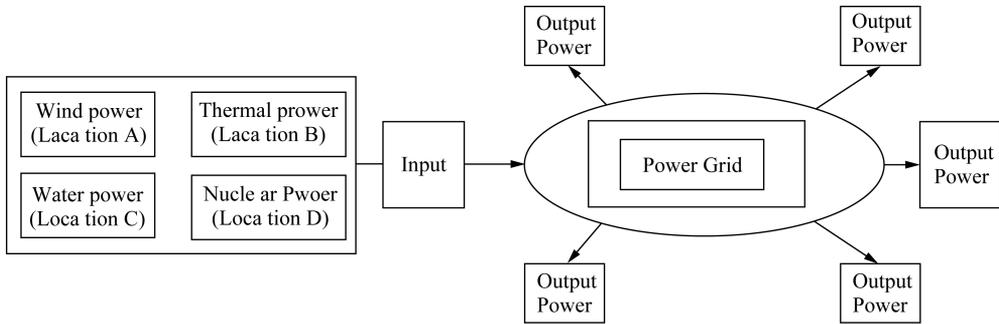


Figure 1 –27 schematic diagram of electric power network

Grid reference grid (ElectricPowerGrid) Came up with the concept, the end in the hope that users of the grid using the grid computing capabilities, as easy as now uses electric power. When electricity is in use, do not need to know that it is the location of power stations, do not need to know what kind of generator to produce power through, whether it is hydro-electric power, or by nuclear power generation, using a uniform form of “energy”. Grid diagram as in Figure 1 –28 shows. Grid and electricity consumers and providers has its own resources, for grid resource providers are power plants, for grid resource providers are computer; for grid resource consumer is the power consumption of devices for grid resource consumer users who are using grid computing to solve problems.

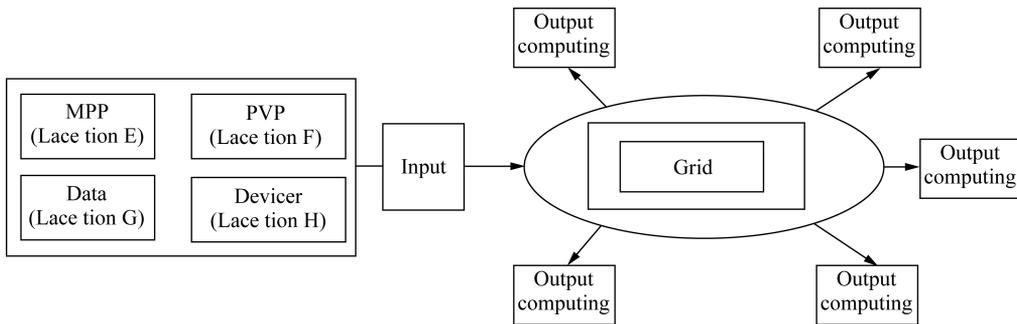


Figure 1 –28 sketch map of grid

Broadly speaking, the grid can be understood as an integrated computing environment and resources, or is it a pool of computing resources. Grids can absorb all kinds of computing resources, and translate them into an efficient, reliable, standard or economy of computing power at the same time. In addition to the various types of computers, computing resources include network communications, data, equipment, even people and other related resources. Problem solving is based on grid for grid computing. Grid computing schematic in Figure 1 –29 shows.

Narrowly speaking, the grid definition of grid resources mainly refers to the distribution of computing resources, grids are usually viewed as a grid, that is mainly used to solve computational problems in science and engineering grids; grid computing means to organize separate computers together to solve complex scientific and engineering problems.



Figure 1 – 29 schematic diagram of grid computing

Network grid break has traditional of shared or collaboration aspects of limit, yiqian on resources of shared often stay in data file transmission of level Shang, and network grid resources of shared allows directly control other resources, through network connection Internet Shang of various resources, including super computer, and mass storage system, and personal computer, and various equipment, , formed on user relative transparent of virtual of high performance calculation environment, can achieved distributed calculation, and high throughput calculation, and data intensive calculation, and collaborative engineering and data query, many function. Grid computing is defined as a wide range of “seamless integration and collaborative computing environments”, become an infrastructure and harmonization of various remote resources.

From the development of computer history, grid computing is accompanied by the rapid development of the Internet, specifically for complex scientific calculations or business of a new computing model. Internet and development on people’s way of thinking, working patterns and life philosophy has had a huge impact and shock. Figure 1 – 30 shows the development of the Internet. To E-mail as the main application of the first generation of Internet computers around the world over the TCP/IP Protocol to connect together, second generation Internet through Web Information browsing and e-business applications and other information service, achieving the world’s Web-connected third generation



Figure 1 – 30 Internet development diagram

Internet Will attempt to achieve full connectivity for all resources on the Internet, including computational resources, storage resources and communication resources, software resources, information resources, knowledge resources, and so on, this is the grid computing (GridComputing)。Grid computing will change the traditional Client/Server and Client/Cluster structures, forming new Pervasive/Grid architecture This architecture would enable users to see the entire network as a giant computer, and from

which to enjoy integrated, dynamic, flexible control, intelligent, cooperative information services, to realize “the network is the computer” concept.

We figure grid technologies fall into three categories 1 – 31 Shown here:

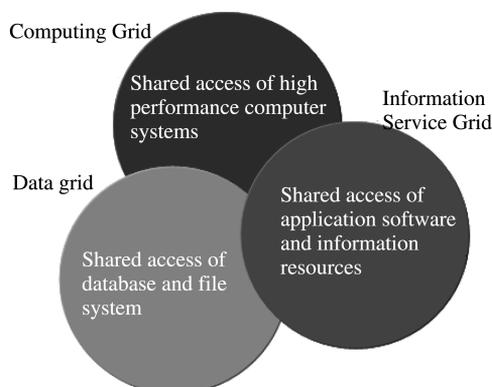


Figure 1 – 31 classification of Grid Technology

(1) Data (information) grid: mainly data is taken as needed.

(a) Data grid can make a set of distributed and heterogeneous data into a single view of the data.

(b) Can manage a large number of data grid, distributed in the different fields of data, storage and network resources, taking into account the local and overall security strategy, efficient resource use, high performance, and taking into account the speed and reliability.

(2) Grid: processing power is taken as needed.

(a) Depending on the application requirements and system load allocation of server resources in real time, and optimize the efficiency of the system.

(b) Provides a unified interface for heterogeneous systems, System Administration and configuration overhead.

(3) Service grids: applications and services is taken as needed.

(a) Web Services (WebServices) Standard, providing application access and integration mechanism for heterogeneous platforms.

(b) Shorten the time and cost of application development and integration, higher reusability, flexibility, scalability and availability.

(c) With the new Web Publication of service standards, and grid computing Web Services is moving towards unity.

Sharing and collaborative nature of grid problems. Sharing network Shanghai, autonomous, distributed and heterogeneous resources for effective organization, providing unified services for grid users transparent access mechanism. These resources include compute, storage, data, software, instruments, sensors, information, knowledge and experts, and so on. Collaboration refers to the resource can mutual interaction, understanding and collaboration, with a view to develop complex grid applications. Including human-human, human-resources, resources-resource cooperative.

The beginning of grid technology in 20 Century 90 Age, experience 10 Years of research and development, grid technologies began to take shape, and continues today to continue to develop. Can be sum-



marized in the development is divided into the following four distinct phases:

(1) Custom exploration programme; 20 Early medieval grid of groping, computation and related areas of work related to grid problems to develop custom solutions. Research work focuses on bring the system to run and explore those are possible. Directly based on the Internet Protocol. From the security, scalability and robustness point of view, there are some limited knowledge of typical features. And interoperability is not the main concern.

(2) Globus Working set: GlobusToolkit Kits from Globus Project Globus Project is the world's most influential and grid computing-related projects, are from all over the world focus on grid technology researchers and developers to work together. GlobusToolkit is an open source grid infrastructure, resources and software based on open architecture, open services, and support grid and grid applications, aimed at providing middleware services for building grid applications and libraries. Now, the latest version is 2005 years 1 month 31 days released GlobusToolkit4(GT4)。

(3) Open grid services architecture; 2002 Year of the open grid services architecture (OGSA) Is a variety of a real community standards. In particular, it includes the 2003 year publishing based on OGSA GT3.0. OGSA based GT2, and greatly expanded the GT2 concepts and techniques. Through the use of service-oriented architecture and Web services technologies, OGSA firmly to grid computing and a broad coalition of industry innovation. OGSA provides a framework, in this framework, you can define a wide range of lightweight services interoperable. OGSA provides a basis, on the basis of this form has more than one technology to provide a rich ecosystem of grid technology.

(4) Managed shared virtual systems; OGSA The initial definition of technical specifications is a big step forward, but in a complete blueprint for the grid there are lots of things to do prior to implementation. Based on OGSA-oriented services infrastructure, will be an extension of the interoperability of services and systems, and they can handle a large number of entities, can also handle small device description, it can increase the degree of Visual, provides a wealth of shared form, through various forms of active management to enhance the quality of services. These efforts produced results at the high end of the computer sciences, areas such as P2P systems, knowledge based systems and automated systems.

Currently most famous grid architecture has two; one is a Globus project hourglass's five-layer structure, and the other is GGF 's OGSA. Globus system architecture shown in Figure 1 – 32 is shown.

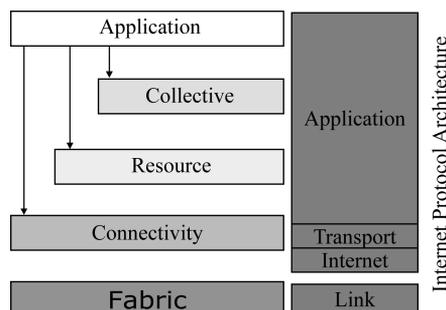


Figure 1 – 32 architecture of Globus system

1.4.2.3 Cloud computing

Cloud computing is 2007 Years the rise of the new term. Divergent views on its definition and meaning, and can be found on the Web at least 100 words, but there is no universally agreed upon definition. China cloud computing Expert Committee Liu Peng Professor give a definition of cloud computing that reflects the nature of cloud computing, covering typical of today's popular cloud computing solutions (including Google cloud computing, Amazon cloud computing, Salesforce Cloud computing, cloud security, etc), but to be able to distinguish other related concepts (such as grid computing, parallel computing, etc).

1. The definition of cloud computing

The definition of cloud computing: cloud computing (CloudComputing) Is a new business model. It distributes tasks across a large number of computers consisting of resource pools, application enables on-demand access to computing power, storage space and a variety of software services.

This pool is known as the "cloud". "The cloud" is some self-maintenance and management of virtual computing resources, usually for a large server clusters, including servers, storage, servers, broadband and so on. Put together, all of the computing resources of cloud computing and managed by the software automatically, without human intervention. This allows application providers do not need to worry about tedious details and can focus on their business, is conducive to innovation and lower costs.

It is called "cloud", because it is realistic in some ways characteristic of cloud: clouds are generally larger; the size of the cloud can be dynamically scalable, its borders are vague; clouds drifting in the air, you could not determine the exact location, but it is there somewhere. It is called "cloud", and also because cloud computing is one of the fathers of the Amazon will have something everyone referred to as grid computing, taking a new name "Elastic Compute Cloud"(EC2), and achieved commercial success.

Was a figure of speech: this is from the old single centralized power supply mode in generator mode to the power plant. It means computing power can also be used as a kind of commodity circulation, like gas, electricity, access to convenient, affordable price. The biggest difference is that it is transmitted through the Internet.

Cloud computing is a parallel computing (ParallelComputing), Distributed computing (Distributed-Computing) And grid computing (GridComputing) Development, or the commercial implementations of these concepts in computer science. Is virtualization cloud computing (Virtualization), utility computing (UtilityComputing), andIaaS (infrastructure as a service), and PaaS (platform as a service)、SaaS (software as a service) Concepts such as evolution and jump results mixed.

Total in cloud computing can be counted as a business evolution of grid computing. Early in 2002 years, Liu Peng on for traditional network grid calculation thought exists not practical problem, proposed calculation pool of concept: "it put dispersed in around of high performance computer with high-speed network connection up, with specifically design of middleware software organic to bonded in with, to Web interface accept around science workers proposed of calculation requests, and will of distribution to right of knot points Shang run. Compute pools can greatly improve service quality and resource utilization,

while avoiding the cross-node application brought about by inefficiency and complexity, able to meet practical requirements under current conditions. ” If the text of the “ high performance computing ” and substitute “ server clusters ”, “ Science ” and substitute “ commercial user ”, it is very close to the current cloud computing.

2. The characteristics of cloud computing

Cloud computing has the following characteristics :

(1) Super size. “ The cloud ” have a considerable scale, Google cloud computing already has 100 million servers, Amazon, and IBM, Microsoft, Yahoo “ In the cloud ” with hundreds of thousands of servers. Enterprise private clouds generally have hundreds or thousands of servers. The “ cloud ” to give users unprecedented computing power.

(2) Virtualization. Cloud computing to support users in any location, using a variety of terminal access to application services. The requested resource from the “ cloud ”, rather than a fixed physical entities. In the “ cloud ” somewhere in the running, but users do not need to know, have to worry about running the exact location. Only need a laptop or a cell phone, everything can be done via Web services people need, including supercomputing tasks.

(3) High reliability. “ The cloud ” using multiple copies of data for fault-tolerance, isomorphic interchangeable compute nodes and other measures to guarantee high service reliability, using cloud computing more reliable than using a local computer.

(4) Universal. Not targeted at specific applications of cloud computing in the “ cloud ” application can be constructed with the support of the ever-changing, with a “ cloud ” can support different applications running simultaneously.

(5) High scalability. The “ cloud ” can scale dynamically scalable, meet the needs of applications and user growth.

(6) On-demand services. The “ cloud ” is a huge pool of resources, you need to purchase; clouds can be like running water, electricity, gas billing.

(7) Extremely cheap. Due to “ cloud ” of special fault-tolerant measures can used extremely cheap of node to constitute cloud, “ cloud ” of automation concentrated type management makes large enterprise without burden increasingly high of data center management cost, “ cloud ” of versatility makes resources of utilization than traditional system sharply upgrade, so user can full enjoy “ cloud ” of low cost advantage, often as long as spent hundreds of dollars, and days time on can completed yiqian need tens of thousands of dollars, and number months time to completed of task.

3. Principles of cloud computing

Cloud computing is divided into IaaS 、 PaaS SaaS Three types of different manufacturers offer different solutions, there is no single technology architecture, constitute an obstacle to readers to understand the principle of cloud computing. Liu Peng, Professor of integrated programmes by different manufacturers, we construct a cloud computing architecture. This architecture as in Figure 1 – 33 shows, it summarizes the main characteristics of different solutions, each of these programmes may only implements some functionality, perhaps there are some relatively minor feature is not yet generally came in.

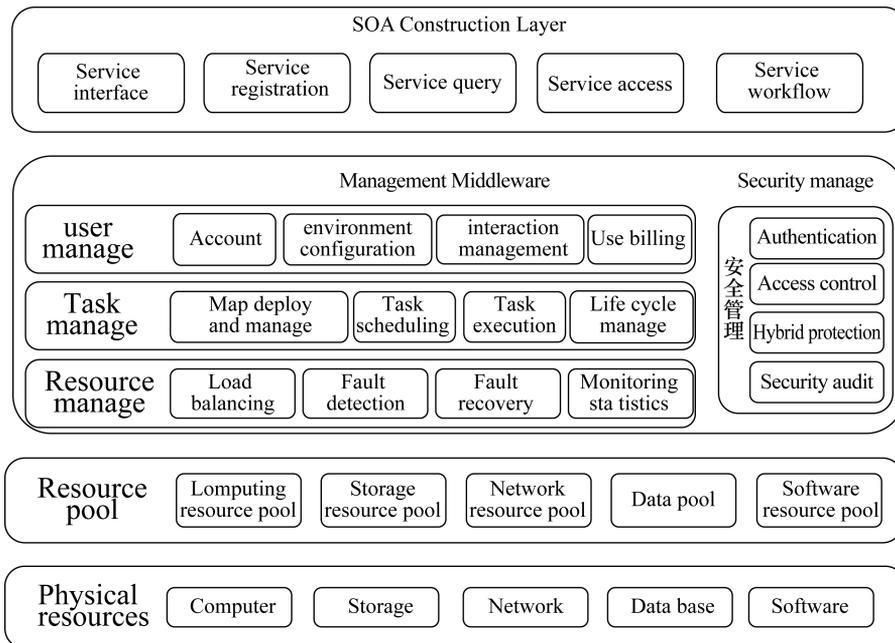


Figure 1 – 33 architecture of cloud computing technology

Cloud computing architecture is divided into 4 Layers; physical resources, resource pooling, management middleware layers and SOA Building layers, as shown in Figure 3 As shown in. Physical resources including computer, memory, network, facilities, databases, and software such as resource pool is a large amount of resources of the same type structure or near-homogenous resource pool, such as resource pooling, data pool, and so on. Building resources pool more is physical resources of integrated and management work, for example research in a standard container of space how loaded Xia 2000 a server, and solution thermal and fault node replaced of problem and reduced energy; management middleware is responsible for on cloud computing of resources for management, and on many application task for scheduling, makes resources can efficient, and security to for application provides service; SOA Build a layer of cloud computing ability packaged into standard WebServices Services, and is included in the SOA System management and use, including service register, locate, access and building services such as workflow. Management middleware, and resource pool is the most critical part of cloud computing, SOA building features more dependent on outside facilities.

The management of cloud computing middleware is responsible for resource management, task management, user management and security management, and so on. Resources management is responsible for balanced to using cloud resources node, detection node of fault and tries to recovery or shield of, and on resources of using situation for monitored statistics; task management is responsible for implementation user or application submitted of task, including completed user task reflected as (Image) of deployment and management, and task scheduling, and task implementation, and task life period management and so on; user management is achieved cloud computing commercial mode of a essential of link, Including user interface, management and recognition of users, create a user program execution environment, the use of user charges, such as safety management to ensure the overall security of cloud computing fa-

cilities, including comprehensive safety and security identity authentication, authorization, auditing, and so on.

Based on this architecture, IaaS Cloud computing, for example, briefly describe the realization mechanism of cloud computing, as in Figure 1 – 34 As shown in.

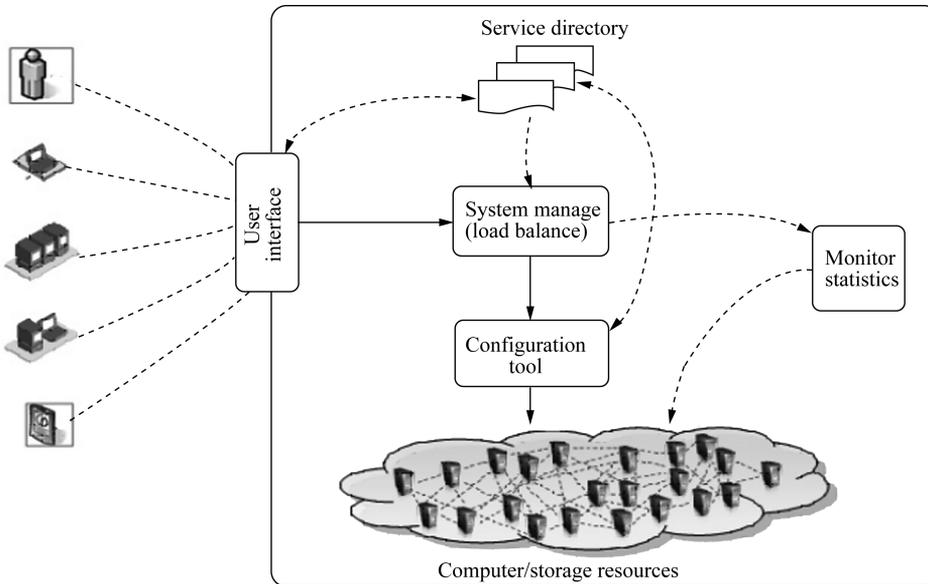


Figure 1 – 34 simplified IaaS implementation mechanism

User interface to the application to WebServices Provides access to the interface, user needs. Service directory is a list of services that users can access. System administration module is responsible for the management and distribution of all of the available resources, the core of which is load balanced. Configuration tool on the node that is responsible for assigning tasks to run environment. Monitoring statistics module is responsible for monitoring node operational status, and statistics on the situation of the users node. Implementation of the process is not complicated; the user interface allows the user to select from the list and call a service. The request is passed to the system management module, it will assign appropriate resources, and then calls the configuration tool to prepare the operating environment for the user.

Exercise one

1. Try brief packet switching points.
2. More than from the compare circuit switching, packet switching and packet switching are the main advantages and disadvantages.
3. Can say: “is equivalent to circuit-switched and connection-oriented, packet-switching and connectionless is the equivalent of”?
4. People often say “packet switching”. But often says “router forwards IP datagrams” or “frame forwarding on the router.” What “group” the word should be used in any situation?

5. What are the characteristics of network architecture?
6. Why adopt a layered structure, network architecture?
7. ISO Proposed OSI What is the purpose of the reference model?
8. OSI Which respectively address the following questions:
 - (a) Information transmission is divided into frames
 - (b) Deciding which path to use by subnet
 - (c) Provides synchronization
9. Discussion with the five points of the network layer protocol architecture, each layer contains the main function.
10. Test TCP/IP OSI Compare the architecture, discuss their similarities.
11. Protocols, and services what is the difference? What is the relationship?
12. Explain the following terms: stack, entity, peer, protocol data units, servers, clients, client—server mode.
13. What are the three elements of network protocol? What is the meaning of?
14. Talk about your understanding of your school's campus network?
15. And the brief reasons of grid technology?
16. Classification, narrative grid?