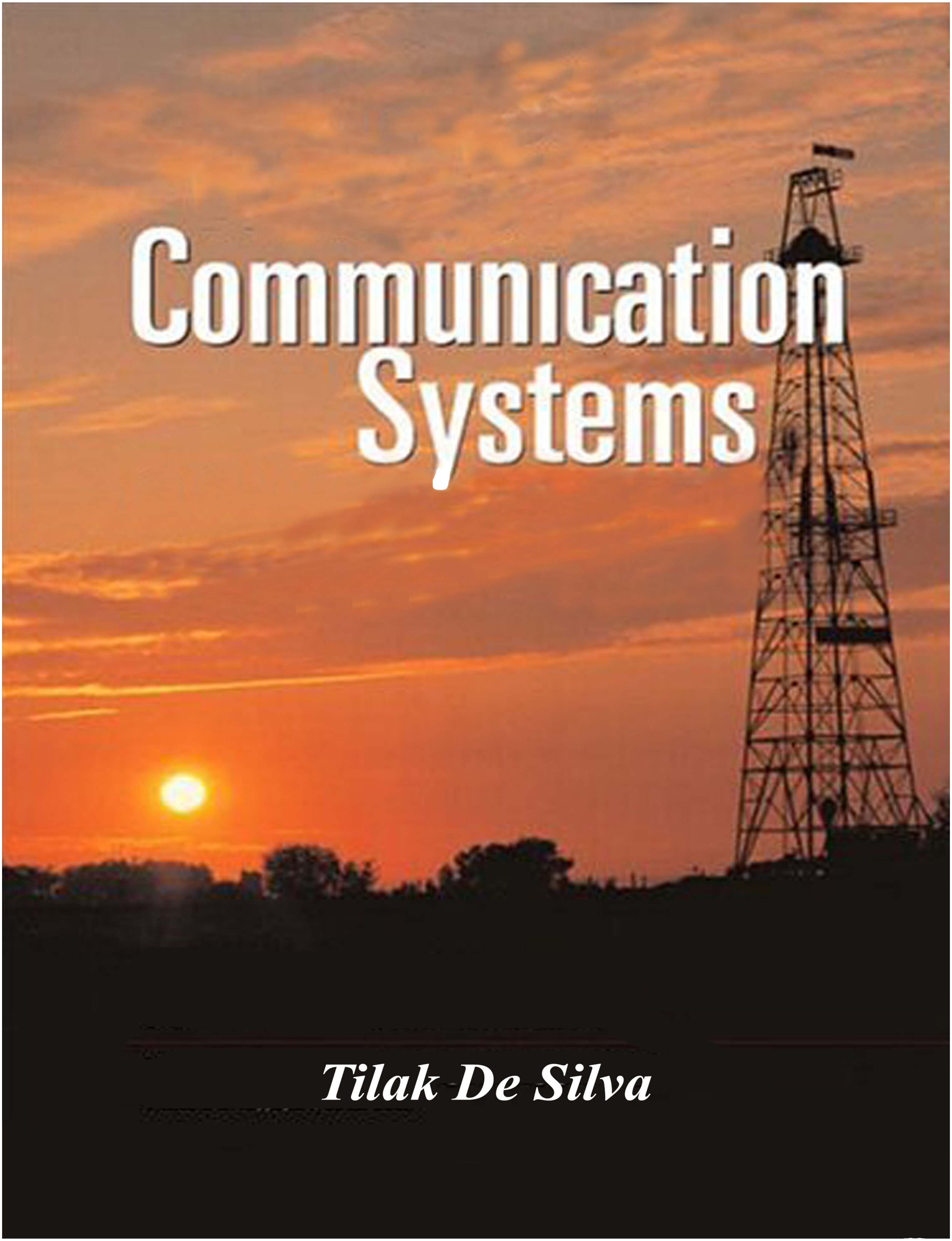


Communication Systems

Tilak De Silva

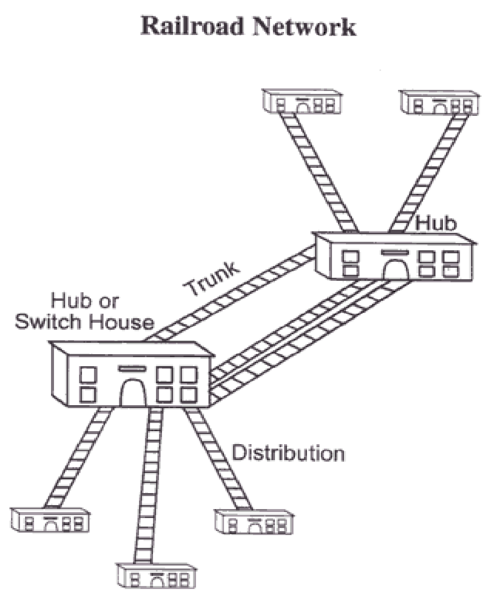
Content

Chapter -1 (Domestic Networks)

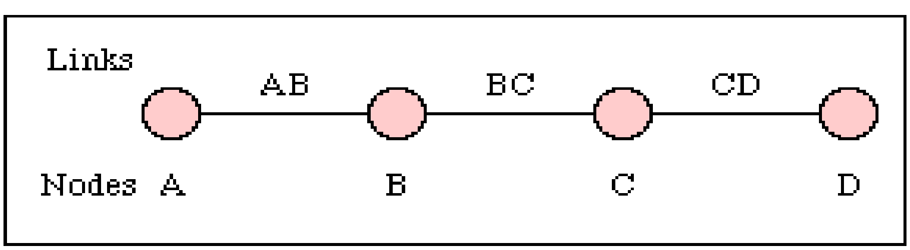
Different Type of Networks

- Road Network
- Railroad Network
- Network for Electricity Distribution
- Network for Water Supply
- Telecommunication Network

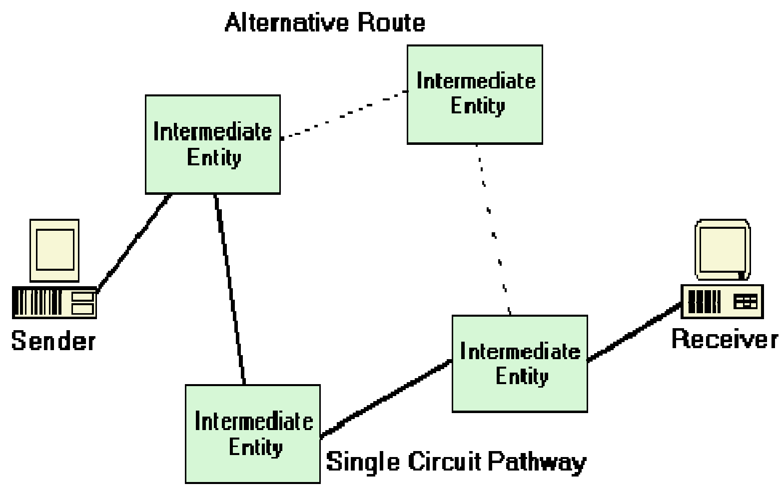
Network



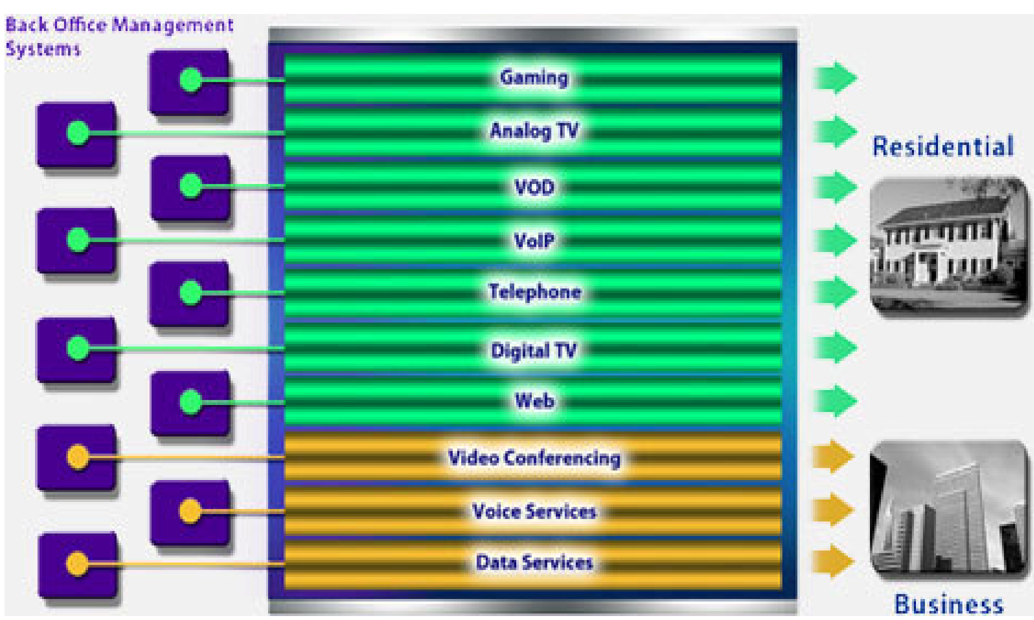
Communication



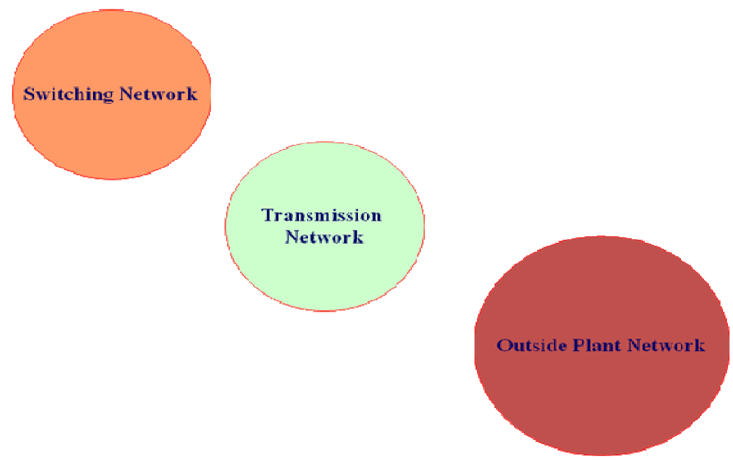
Communication Network



Why Communication Network is?



Key Elements of Telecommunication Network

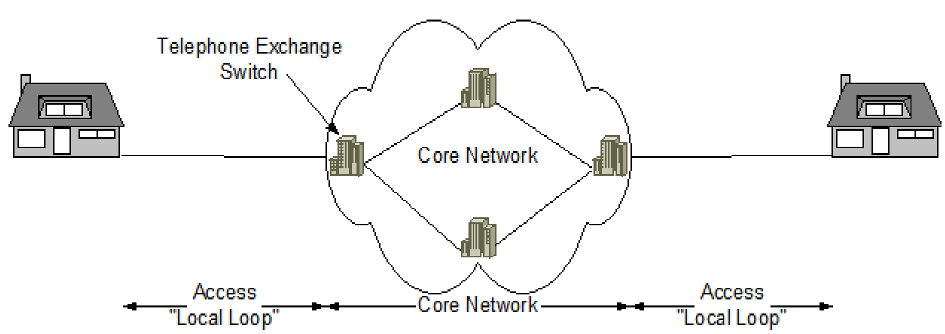


Major Elements

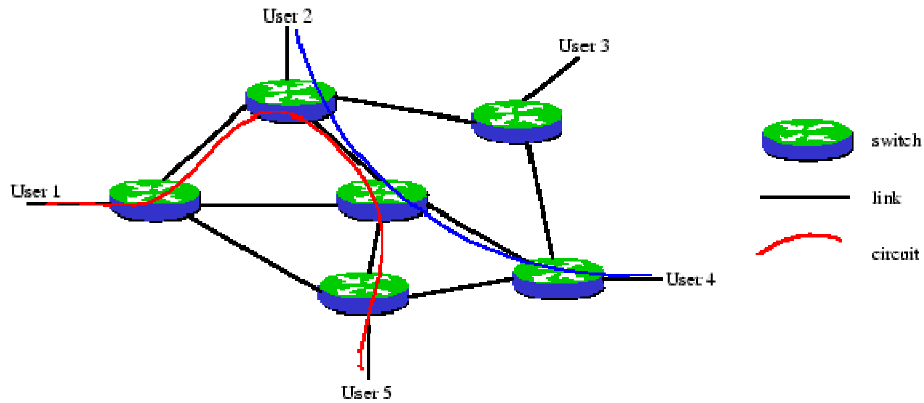
- Internal Plant
  - Switching
  - Transmission

- External Plant (wired line)
  - Subscriber cable network

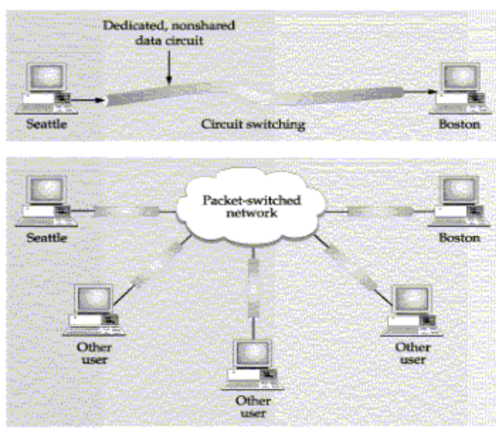
Communication Network



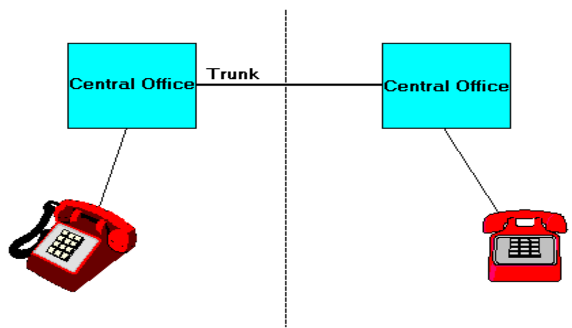
Link and Circuit



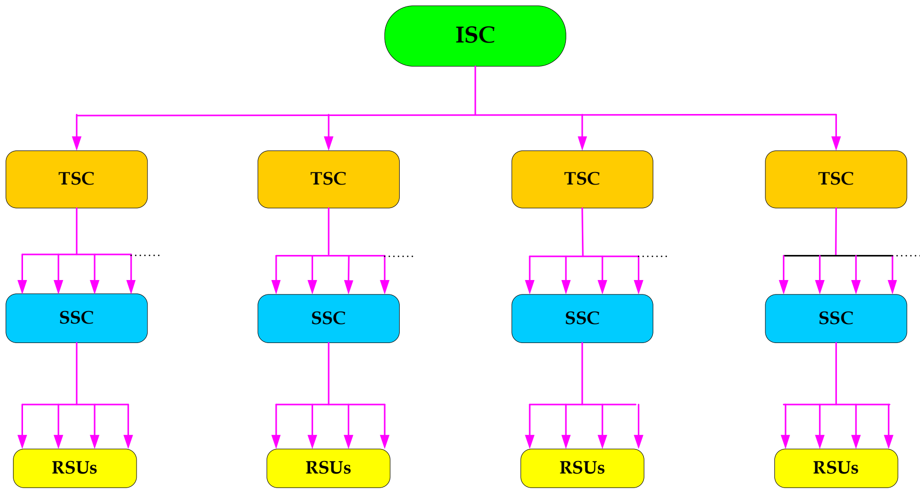
Switching



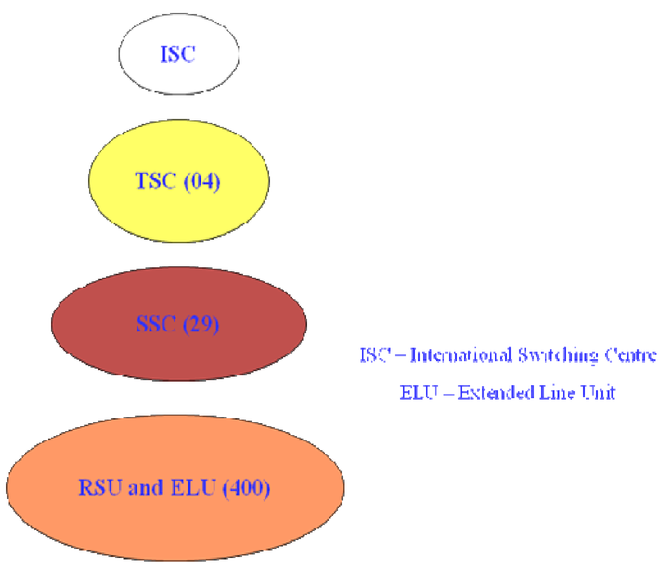
What is Trunk?



Switching Network Hierarchy



Over View of Switch Network



Switch towards next generation

- From
  - Conventional circuit switches
- To
  - IP based switches

Transmission

- Microwave Transmission
- Optical Fiber Transmission

Radio Transmission Tower

Copper Vs Optical

- Limited Band width
- Inflexibility
- Limited Reliability
- Long Installation time
- Maintenance
- Passive Network
- Losses
- Applicability
- Interference
- Security

Optical Transmission

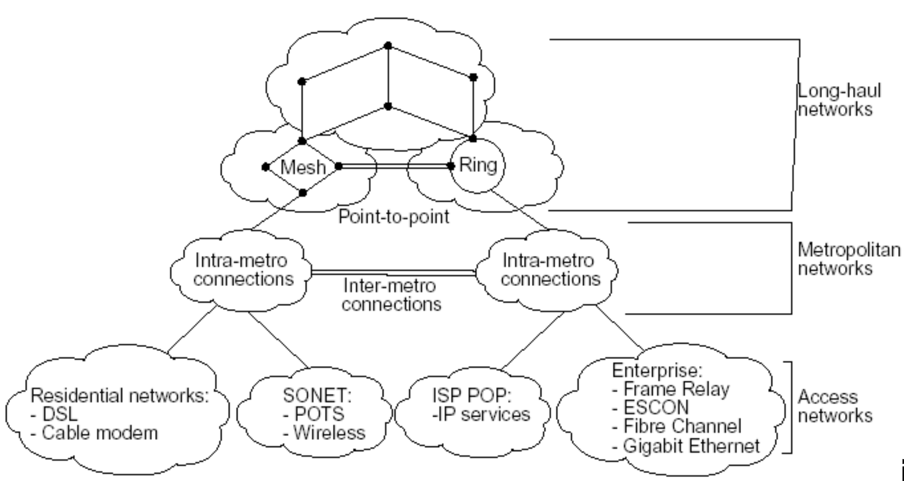
- who is using ?
- why is used ?
- when is used ?

Introducing DWDM

The following discussion provides some background on why dense wavelength division multiplexing (DWDM) is an important innovation in optical networks and what benefits it can provide. We begin with a high-level view of the segments of the global network and the economic forces driving the revolution in fiber optic networks. We then examine the differences between traditional time-division multiplexing (TDM) and wavelength division multiplexing (WDM). Finally, we explore the advantages of this new technology.

Global Network Hierarchy

It is the nature of modern communications networks to be in a state of ongoing evolution. Factors such as new applications, changing patterns of usage, and redistribution of content make the definition of networks a work in progress. Nevertheless, we can broadly define the larger entities that make up the global network based on variables such as transport technology, distance, applications, and so on. One way of describing the metropolitan area network (MAN) would be to say that it is neither the long-haul nor the access parts of the network, but the area that lies between those two.



Long-Haul Networks

Long-haul networks are at the core of the global network. Dominated by a small group of large transnational and global carriers, long-haul networks connect the MANs. Their application is transport, so their primary concern is capacity. In many cases these networks, which have traditionally been based on Synchronous Optical Network (SONET) or Synchronous Digital Hierarchy (SDH) technology, are experiencing fiber exhaust as a result of high bandwidth demand.

Access Networks

At the other end of the spectrum are the access networks. These networks are the closest to the end users, at the edge of the MAN. They are characterized by diverse protocols and infrastructures, and they span a broad spectrum of rates. Customers range from residential Internet users to large corporations and institutions. The predominance of IP traffic, with its inherently bursty, asymmetric, and unpredictable nature, presents many challenges, especially with new real-time applications. At the same time, these networks are required to continue to support legacy traffic and protocols, such as IBM's Enterprise System Connection (ESCON).

Metropolitan Area Networks

Between these two large and different networking domains lie the MANs. These networks channel traffic within the metropolitan domain (among businesses, offices, and metropolitan areas) and between large long-haul points of presence (POPs). The MANs have many of the same characteristics as the access networks, such as diverse networking protocols and channel speeds. Like access networks, MANs have been traditionally SONET/SDH based, using point-to-point or ring topologies with add/drop multiplexers (ADMs). The MAN lies at a critical juncture. On the one hand, it must meet the needs created by the dynamics of the ever-increasing bandwidth available in long-haul transport networks. On the other hand, it must address the growing connectivity requirements and access technologies that are resulting in demand for high-speed, customized data services.

Metropolitan and Long-Haul Networks Compared

There is a natural tendency to regard the MAN as simply a scaled-down version of the long-haul network. It is true that networks serving the metropolitan area encompass shorter distances than in the long-haul transport networks. Upon closer examination, however, these differences are superficial compared to other factors. Network shape is more stable in long-haul, while topologies change frequently in the MAN. Many more types of services and traffic types must be supported in MANs, from traditional voice and leased line services to new applications, including data storage, distributed applications, and video. The long-haul, by contrast, is about big pipes. Another important way in which metropolitan networks today differ from trunkoriented long haul networks is that they encompass a collection of low bit-rate asynchronous and synchronous transmission equipment, short loops, small cross-sections, and a variety of users with varying bandwidth demands. These fundamental differences between the two types of networks have powerful implications for the requirements in the metropolitan domain. Protocol and speed transparency, scalability, and dynamic provisioning are at least as important as capacity, which rules in the long-haul market.

An Alternative View

The preceding breakdown of the global network represents a somewhat simplified view. In reality, the lines between the domains are not always so clear-cut. Long-haul and metropolitan networks are sometimes not clearly delineated; the same holds true for the access and metropolitan domains. Furthermore, other views of the global network exist. One, for example, defines the access network as part of, rather than separate from, the MAN, while also including enterprise connectivity in the metropolitan domain. In this view, the metropolitan market breaks down as follows:

- Core—These are essentially scaled-down long-haul systems. They are considered the core of the MAN, because they interconnect carrier POPs and do not directly interface with end users.

- Metropolitan access—This is the segment between carrier POPs and access facilities, which could be equipment at customer premises or at an aggregation point.

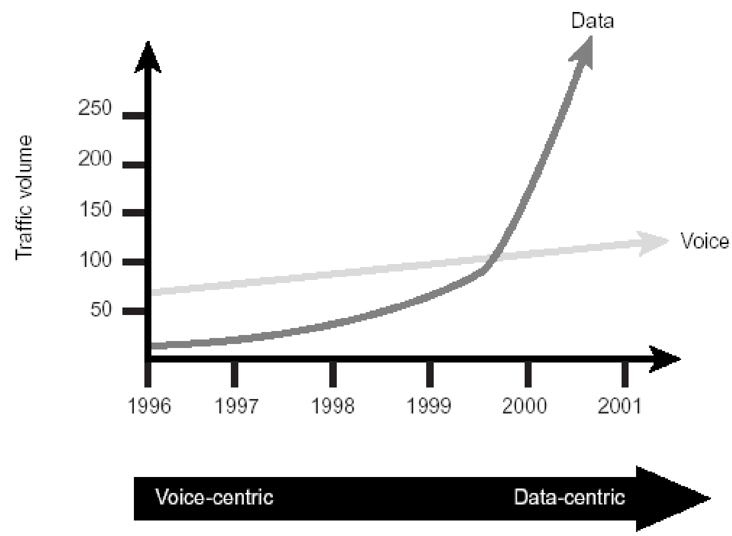
- Enterprise—This is the part of the network dedicated to serving the needs of enterprises. Using owned or leased fiber (or leased fiber capacity), connectivity is provided between geographically disparate enterprise sites and for new applications, such as storage area networks (SANs).

Economic Forces

As we enter the twenty-first century, it goes without saying that information services have permeated society and commerce. Information, while still a tool, has become a commodity in itself. Yet the universal acceptance and ubiquitous adoption of information technology systems has strained the backbones on which they were built. High demand—coupled with high usage rates, a deregulated telecommunications environment, and high availability requirements—is rapidly depleting the capacities of fibers that, when installed 10 years ago, were expected to suffice for the foreseeable future.

Bandwidth Demand

The explosion in demand for network bandwidth is largely due to the growth in data traffic, specifically Internet Protocol (IP). Leading service providers report bandwidths doubling on their backbones about every six to nine months. This is largely in response to the 300 percent growth per year in Internet traffic, while traditional voice traffic grows at a compound annual rate of only about 13 percent



At the same time that network traffic volume is increasing, the nature of the traffic itself is becoming more complex. Traffic carried on a backbone can originate as circuit based (TDM voice and fax), packet based (IP), or cell based (ATM and Frame Relay). In addition, there is an increasing proportion of delay sensitive data, such as voice over IP and streaming video. In response to this explosive growth in bandwidth demand, along with the emergence of IP as the common foundation for all services, long-haul service providers are moving away from TDM based systems, which were optimized for voice but now prove to be costly and inefficient. Meanwhile, metropolitan networks are also experiencing the impact of growing congestion, as well as rapidly changing requirements that call for simpler and faster provisioning than is possible with older equipment and technologies. Of key importance in the metropolitan area is the growth in storage area networks (SANs), discussed in the "Storage Area Networks"

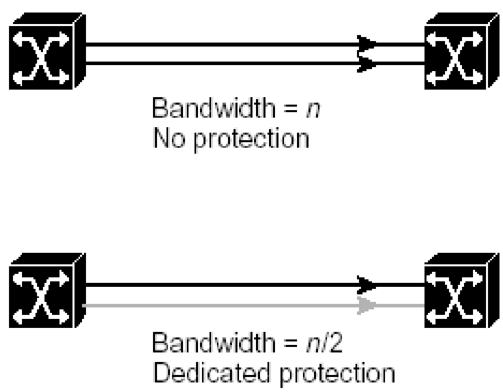
Competition and Reliability

While the demand for bandwidth is driven largely by new data applications, Internet usage, and the growth in wireless communications, two additional factors come into play: competition and network availability. The telecommunication sector, long a beneficiary of government regulation, is now a highly competitive industry. Competition was first introduced into the U.S. long-distance market in 1984, and the 1996 Telecommunications Reform Act is now resulting in an increasingly broad array of new operators. These new carriers are striving to meet the new demand for additional services and capacity.

There are two main effects on the industry from competition:

- Enhanced services are created by newcomers trying to compete with incumbents. In the metropolitan market, for example, there are broadband wireless and DSL services to homes and small and medium-sized business, high-speed private line and VPN services to corporations, and transparent LAN services to enterprise network customers.

- New carriers coming onto the scene create new infrastructure so that they do not have to lease from existing operators. Using this strategy, they have more control over provisioning and reliability. As telecommunications and data services have become more critical to business operations, service providers have been required to ensure that their networks are fault tolerant. To meet these requirements, providers have had to build backup routes, often using simple 1:1 redundancy in ring or point-to-point configurations. Achieving the required level of reliability, however, means reserving dedicated capacity for failover. This can double the need for bandwidth on an already strained infrastructure.



Options for Increasing Carrier Bandwidth

Faced with the challenge of dramatically increasing capacity while constraining costs, carriers have two options: Install new fiber or increase the effective bandwidth of existing fiber. Laying new fiber is the traditional means used by carriers to expand their networks. Deploying new fiber, however, is a costly proposition. It is estimated at about $70,000 per mile, most of which is the cost of permits and construction rather than the fiber itself. Laying new fiber may make sense only when it is desirable to expand the embedded base.
Increasing the effective capacity of existing fiber can be accomplished in two ways:

- Increase the bit rate of existing systems.
- Increase the number of wavelengths on a fiber.

Increase the Bit Rate

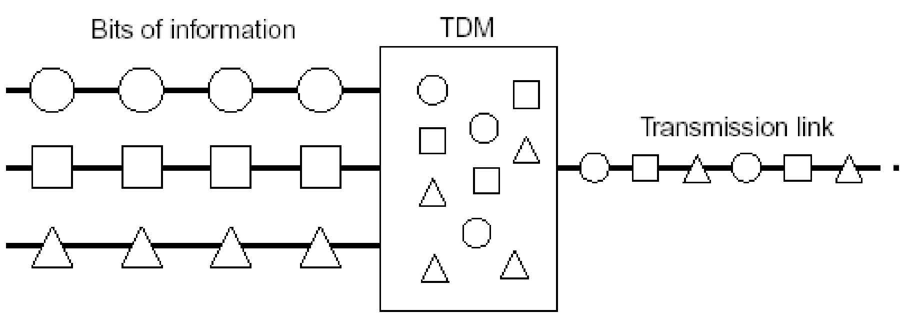
Using TDM, data is now routinely transmitted at 2.5 Gbps (OC-48) and, increasingly, at 10 Gbps (OC-192); recent advances have resulted in speeds of 40 Gbps (OC-768). The electronic circuitry that makes this possible, however, is complex and costly, both to purchase and to maintain. In addition, there are significant technical issues that may restrict the applicability of this approach. Transmission at OC-192 over single-mode (SM) fiber, for example, is 16 times more affected by chromatic dispersion than the next lower aggregate speed, OC-48. The greater transmission power required by the higher bit rates also introduces nonlinear effects that can affect waveform quality. Finally, polarization mode dispersion, another effect that limits the distance a light pulse can travel without degradation, is also an issue.

Increase the Number of Wavelengths

In this approach, many wavelengths are combined onto a single fiber. Using wavelength division multiplexing (WDM) technology several wavelengths, or light colors, can simultaneously multiplex signals of 2.5 to 40 Gbps each over a strand of fiber. Without having to lay new fiber, the effective capacity of existing fiber plant can routinely be increased by a factor of 16 or 32. Systems with 128 and 160 wavelengths are in operation today, with higher density on the horizon. The specific limits of this technology are not yet known.

Time-Division Multiplexing

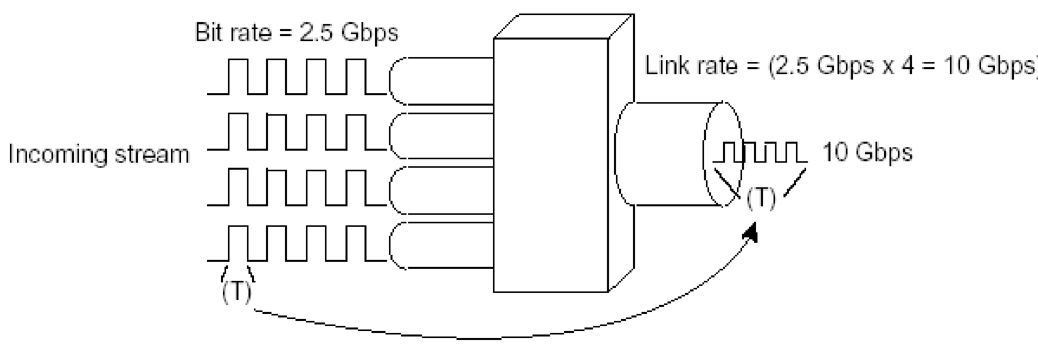
Time-division multiplexing (TDM) was invented as a way of maximizing the amount of voice traffic that could be carried over a medium. In the telephone network before multiplexing was invented, each telephone call required its own physical link. This proved to be an expensive and unscalable solution. Using multiplexing, more than one telephone call could be put on a single link. TDM can be explained by an analogy to highway traffic. To transport all the traffic from four tributaries to another city, you can send all the traffic on one lane, providing the feeding tributaries are fairly serviced and the traffic is synchronized. So, if each of the four feeds puts a car onto the trunk highway every four seconds, then the trunk highway would get a car at the rate of one each second. As long as the speed of all the cars is synchronized, there would be no collision. At the destination the cars can be taken off the highway and fed to the local tributaries by the same synchronous mechanism, in reverse. This is the principle used in synchronous TDM when sending bits over a link. TDM increases the capacity of the transmission link by slicing time into smaller intervals so that the bits from multiple input sources can be carried on the link, effectively increasing the number of bits transmitted per second.



With TDM, input sources are serviced in round-robin fashion. Though fair, this method results in inefficiency, because each time slot is reserved even when there is no data to send. This problem is mitigated by the statistical multiplexing used in Asynchronous Transfer Mode (ATM).Although ATM offers better bandwidth utilization, there are practical limits to the speed that can be achieved due to the electronics required for segmentation and reassembly (SAR) of ATM cells that carry packet data.

SONET and TDM

The telecommunications industry adopted the Synchronous Optical Network (SONET) or Synchronous Digital Hierarchy (SDH) standard for optical transport of TDM data. SONET, used in North America, and SDH, used elsewhere, are two closely related standards that specify interface parameters, rates, framing formats, multiplexing methods, and management for synchronous TDM over fiber. SONET/SDH takes *n* bit streams, multiplexes them, and optically modulates the signal, sending it out using a light emitting device over fiber with a bit rate equal to (incoming bit rate) x *n*. Thus traffic arriving at the SONET multiplexer from four places at 2.5 Gbps will go out as a single stream at 4 x 2.5 Gbps, or 10 Gbps. This principle is illustrated in Figure 1-5, which shows an increase in the bit rate by a factor of four in time slot *T*.



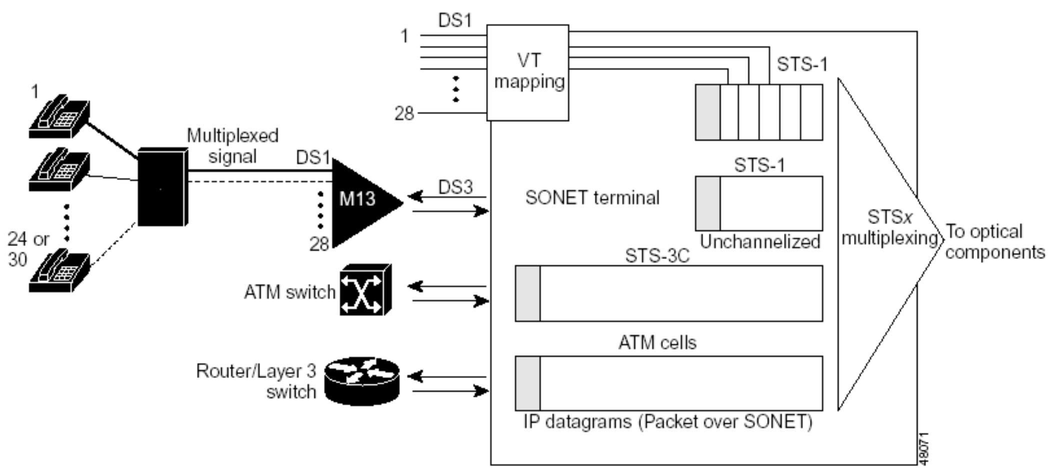
The original unit used in multiplexing telephone calls is 64 kbps, which represents one phone call. Twenty-four (in North America) or thirty-two (outside North America) of these units are multiplexed using TDM into a higher bit-rate signal with an aggregate speed of 1.544 Mbps or 2.048 Mbps for transmission over T1 or E1 lines, respectively. The hierarchy for multiplexing telephone calls is shown in Table.

Signal	Bit Rate	Voice Slots
DS0	64 kbps	1 DS0
DS1	1.544 Mbps	24 DS0s
DS2	6.312 Mbps	96 DS0s
DS3	44.736 Mbps	28 DS1s

These are the basic building blocks used by SONET/SDH to multiplex into a standard hierarchy of speeds, from STS-1 at 51.85 Mbps to STS-192/STM-64 at 10 Gbps. Table 1-2 shows the relationship between the telco signal rates and the most commonly used levels of the SONET/SDH hierarchy (OC-768 is not yet common).

Optical Carrier	SONET/SDH Signal	Bit Rate	Capacity
OC-1	STS-1	51.84 Mbps	28 DS1s or 1 DS3
OC-3	STS-3/STM-1	155.52 Mbps	84 DS1s or 3 DS3s
OC-12	STS-12/STM-4	622.08 Mbps	336 DS1s or 12 DS3s
OC-48	STS-48/STM-16	2488.32 Mbps	1344 DS1s or 48 DS3s
OC-192	STS-192/STM-64	9953.28 Mbps	5379 DS1s or 192 DS3s

depicts this multiplexing and aggregation hierarchy. Using a standard called virtual tributaries for mapping lower-speed channels into the STS-1 payload, the 28 DS1 signals can be mapped into the STS-1 payload, or they can be multiplexed to DS3 with an M13 multiplexer and fit directly into the STS-1. Note also that ATM and Layer 3 traffic, using packet over SONET (POS), can feed into the SONET terminal from switches equipped with SONET interfaces.



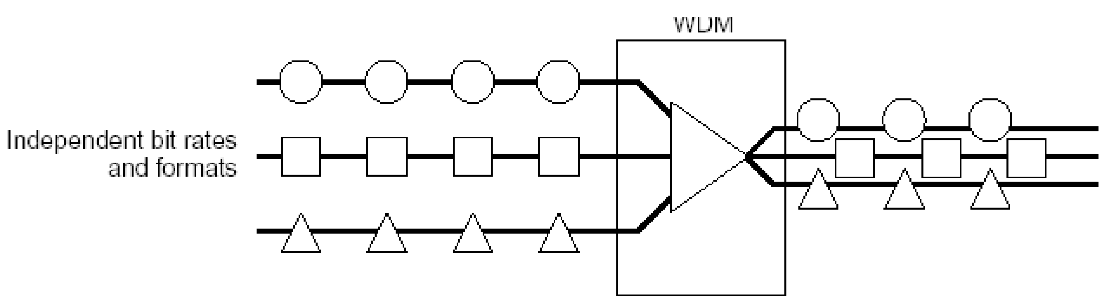
SONET/SDH does have some drawbacks. As with any TDM, the notions of priority or congestion do not exist in SONET or SDH. Also, the multiplexing hierarchy is a rigid one. When more capacity is needed, a leap to the next multiple must be made, likely resulting in an outlay for more capacity than is initially needed. For example, the next incremental step from 10 Gbps (STS-192) TDM is 40 Gbps (STS-768). Also, since the hierarchy is optimized for voice traffic, there are inherent inefficiencies when carrying data traffic with SONET frames. Some of these inefficiencies are shown in Table 1-3. DWDM, by contrast, can transport any protocol, including SONET, without special encapsulation.

Ethernet	SONET/SDH Signal	Bit Rate	Wasted Bandwidth
10BASE-T (10 Mbps)	STS-1	51.8540 Mbps	80.709%
100BASE-T (100 Mbps)	STS-3/STM-1	155.520 Mbps	35.699%
1000BASE-T (1000 Mbps)	STS-48/STM-16	2488.32 Mbps	59.812%

To summarize, the demand placed on the transport infrastructure by bandwidth-hungry applications and the explosive growth of the Internet has exceeded the limits of traditional TDM. Fiber, which once promised seemingly unlimited bandwidth, is being exhausted, and the expense, complexity, and scalability limitations of the SONET infrastructure are becoming increasingly problematic.

Wavelength Division Multiplexing

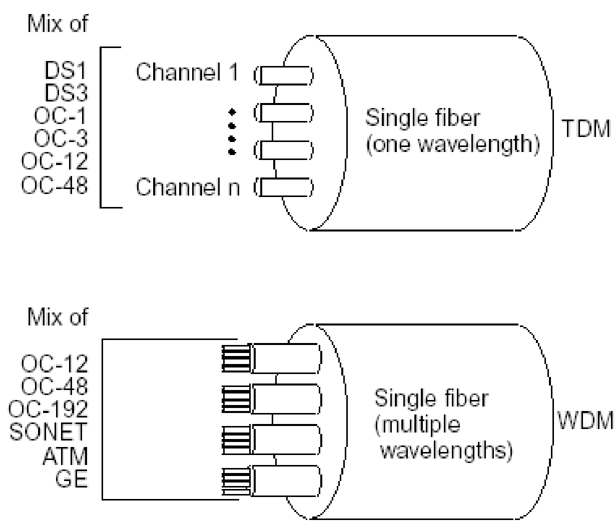
WDM increases the carrying capacity of the physical medium (fiber) using a completely different method from TDM.WDMassigns incoming optical signals to specific frequencies of light (wavelengths,or lambdas) within a certain frequency band. This multiplexing closely resembles the way radio stations broadcast on different wavelengths without interfering with each other (see Figure 1-7). Because each channel is transmitted at a different frequency, we can select from them using a tuner. Another way to think about WDM is that each channel is a different color of light; several channels then make up a "rainbow."



In a WDM system, each of the wavelengths is launched into the fiber, and the signals are demultiplexed at the receiving end. Like TDM, the resulting capacity is an aggregate of the input signals, but WDM carries each input signal independently of the others. This means that each channel has its own dedicated bandwidth; all signals arrive at the same time, rather than being broken up and carried in time slots. The difference between WDM and dense wavelength division multiplexing (DWDM) is fundamentally one of only degree. DWDM spaces the wavelengths more closely than does WDM, and therefore has a greater overall capacity.The limits of this spacing are not precisely known, and have probably not been reached, though systems are available in mid-year 2000 with a capacity of 128 lambdas on one fiber. DWDM has a number of other notable features, which are discussed in greater detail in the following chapters. These include the ability to amplify all the wavelengths at once without first converting them to electrical signals, and the ability to carry signals of different speeds and types simultaneously and transparently over the fiber (protocol and bit rate independence).

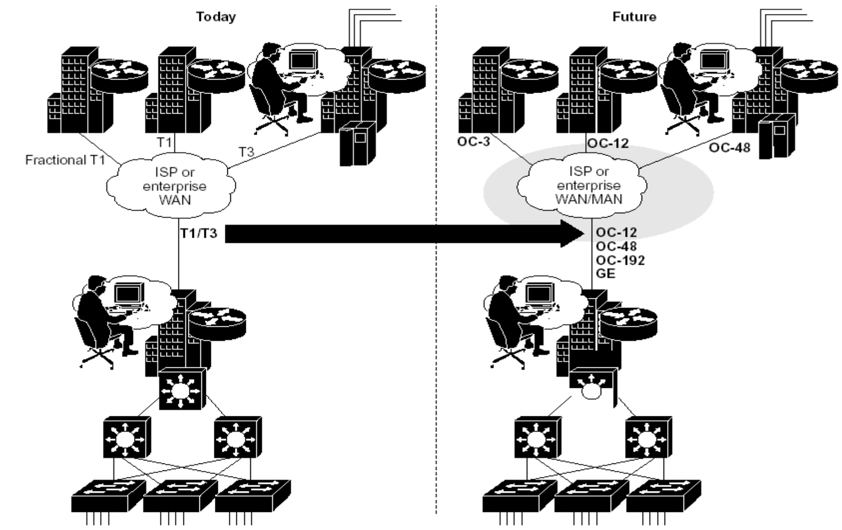
TDM and WDM Compared

SONET TDM takes synchronous and asynchronous signals and multiplexes them to a single higher bit rate for transmission at a single wavelength over fiber. Source signals may have to be converted from electrical to optical, or from optical to electrical and back to optical before being multiplexed. WDM takes multiple optical signals, maps them to individual wavelengths, and multiplexes the wavelengths over a single fiber. Another fundamental difference between the two technologies is that WDM can carry multiple protocols without a common signal format, while SONET cannot.



Additional Drivers in Metropolitan Area Networks

Bandwidth, the chief driver in the long-haul market, is also a big driver in metropolitan area, access, and large enterprise networks. In these types of networks additional applications driving demand for bandwidth include storage area networks (SANs), which make possible the server less office, consolidation of data centers, and real-time transaction processing backup.



There is also rapidly increasing demand on access networks, which function primarily to connect end users over low-speed connections, such as dial-up lines, DSL, cable, and wireless, to a local POP. These connections are typically aggregated and carried over a SONET ring, which at some point attaches to a local POP that serves as an Internet gateway for long hauls. Now, the growing demand for high-speed services is prompting service providers to transform the POP into a dynamic service-delivery center. As a result, it is increasingly likely that a customer now obtains many high-speed services directly from the POP, without ever using the core segment of the Internet.

Value of DWDM in the Metropolitan Area

DWDM is the clear winner in the backbone. It was first deployed on long-haul routes in a time of fiber scarcity. Then the equipment savings made it the solution of choice for new long-haul routes, even when ample fiber was available. While DWDM can relieve fiber exhaust in the metropolitan area, its value in this market extends beyond this single advantage. Alternatives for capacity enhancement exist, such as pulling new cable and SONET overlays, but DWDM can do more. What delivers additional value in the metropolitan market is DWDM's fast and flexible provisioning of protocol- and bit rate-transparent, data-centric, protected services, along with the ability to offer new and higher-speed services at less cost.

The need to provision services of varying types in a rapid and efficient manner in response to the changing demands of customers is a distinguishing characteristic of the metropolitan networks. With SONET, which is the foundation of the vast majority of existing MANs, service provisioning is a lengthy and complex process. Network planning and analysis, ADM provisioning, Digital Crossconnect System (DCS) reconfiguration, path and circuit verification, and service creation can take several weeks. By contrast, with DWDM equipment in place provisioning new service can be as simple as turning on another lightwave in an existing fiber pair.

Potential providers of DWDM-based services in metropolitan areas, where abundant fiber plant already exists or is being built, include incumbent local exchange carriers (ILECs), competitive local exchange carriers (CLECs), inter-exchange carriers (IXCs), Internet service providers (ISPs), cable companies, private network operators, and utility companies. Such carriers can often offer new services for less cost than older ones. Much of the cost savings is due to reducing unnecessary layers of equipment, which also lowers operational costs and simplifies the network architecture. Carriers can create revenue today by providing protocol-transparent, high-speed LAN and SAN services to large organizations, as well as a mixture of lower-speed services (Token Ring, FDDI, Ethernet) to smaller organizations. In implementing an optical network, they are ensuring that they can play in the competitive field of the future.

Requirements in the Metropolitan Area

The requirements in the metropolitan market may differ in some respects from those in the long-haul network market, yet metropolitan networks are still just a geographically distinguished segment of the global network. What happens in the core must be supported right to the edge. IP, for example, is the dominant traffic type, so interworking with this layer is a requirement, while not ignoring other traffic (TDM). Network management is now of primary concern, and protection schemes that ensure high availability are a given.

Key requirements for DWDM systems in the MAN include the following:

- Multiprotocol support
- Scalability
- Reliability and availability
- Openness (interfaces, network management, standard fiber types, electromagnetic compatibility)
- Ease of installation and management
- Size and power consumption
- Cost effectiveness

Why DWDM?

From both technical and economic perspectives, the ability to provide potentially unlimited transmission capacity is the most obvious advantage of DWDM technology. The current investment in fiber plant can not only be preserved, but optimized by a factor of at least 32. As demands change, more capacity can be added, either by simple equipment upgrades or by increasing the number of lambdas on the fiber, without expensive upgrades. Capacity can be obtained for the cost of the equipment, and existing fiber plant investment is retained.

Bandwidth aside, DWDM's most compelling technical advantages can be summarized as follows:

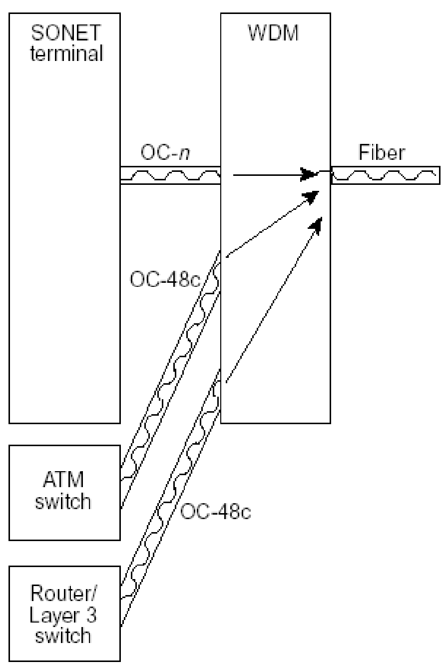
- Transparency—Because DWDM is a physical layer architecture, it can transparently support both TDM and data formats such as ATM, Gigabit Ethernet, ESCON, and Fibre Channel with open interfaces over a common physical layer.

- Scalability—DWDM can leverage the abundance of dark fiber in many metropolitan area and enterprise networks to quickly meet demand for capacity on point-to-point links and on spans of existing SONET/SDH rings.

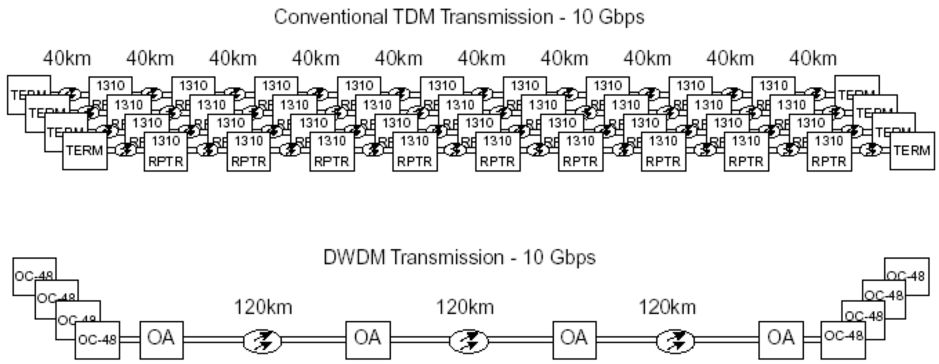
- Dynamic provisioning—Fast, simple, and dynamic provisioning of network connections give providers the ability to provide high-bandwidth services in days rather than months. In the following sections we discuss some additional advantages, including migration from SONET and reliability.

SONET with DWDM

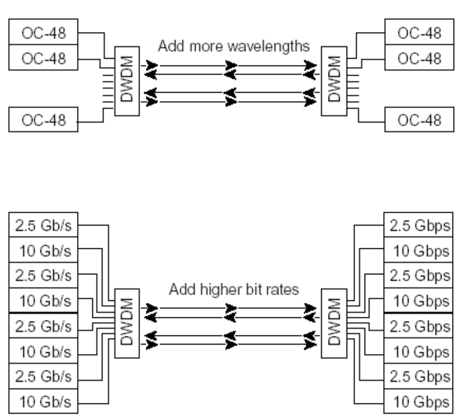
By using DWDM as a transport for TDM, existing SONET equipment investments can be preserved. Often new implementations can eliminate layers of equipment. For example, SONET multiplexing equipment can be avoided altogether by interfacing directly to DWDM equipment from ATM and packet switches, where OC-48 interfaces are common. Additionally, upgrades do not have to conform to specific bit rate interfaces, as with SONET, where aggregation of tributaries is locked into specific values.



Optical signals become attenuated as they travel through fiber and must be periodically regenerated in core networks. In SONET/SDH optical networks prior to the introduction of DWDM, each separate fiber carrying a single optical signal, typically at 2.5 Gbps, required a separate electrical regenerator every 60 to 100 km (37 to 62 mi). As additional fibers were "turned up" in a core network, the total cost of regenerators could become very large, because not only the cost of the regenerators themselves, but also the facilities to house and power them, had to be considered. The need to add regenerators also increased the time required to light new fibers The upper part of Figure 1-11 shows the infrastructure required to transmit at 10 Gbps (4 x OC-48 SR interfaces) across a span of 360 km (223 mi) using SONET equipment; the lower part of the figure shows the infrastructure required for the same capacity using DWDM. While optical amplifiers could be used in the SONET case to extend the distance of spans before having to boost signal power, there would still need to be an amplifier for each fiber. Because with DWDM all four signals can be transported on a single fiber pair (versus four), fewer pieces of equipment are required. Eliminating the expense of regenerators (RPTR) required for each fiber results in considerable savings.



A single optical amplifier can reamplify all the channels on a DWDM fiber without demultiplexing and processing them individually, with a cost approaching that of a single regenerator. The optical amplifier merely amplifies the signals; it does not reshape, retime or retransmit them as a regenerator does, so the signals may still need to be regenerated periodically. But depending on system design, signals can now be transmitted anywhere from 600 to thousands of kilometers without regeneration. In addition to dramatically reducing the cost of regenerators, DWDM systems greatly simplify the expansion of network capacity. The only requirement is to install additional or higher bit-rate interfaces in the DWDM systems at either end of the fiber. In some cases it will only be necessary to increase the number of lambdas on the fiber by deploying existing interfaces, as shown in the upper half of Figure 1-12. The existing optical amplifiers amplify the new channel without additional regenerators. In the case of adding higher bit-rate interfaces, as shown in the lower half of Figure 1-12, fiber type can become a consideration. See the “Optical Fibers” section on page 2-5 for an overview of types of optical fibers and their uses.



Although amplifiers are of great benefit in long-haul transport, they are often unnecessary in metropolitan networks. Where distances between network elements are relatively short, signal strength and integrity can be adequate without amplification. But with MANs expanding in deeper into long-haul reaches, amplifiers will become useful.

Enhancing Performance and Reliability

Today's metropolitan and enterprise networks support many mission-critical applications that require high availability, such as billing and accounting on mainframes or client-server installations in data centers. Continuous backups or reliable decentralized data processing and storage are essential. These applications, along with disaster recovery and parallel processing, have high requirements for performance and reliability. As enterprises out source data services and inter-LAN connectivity, the burden of service falls on the service provider rather than on the enterprise. With DWDM, the transport network is theoretically unconstrained by the speed of available electronics. There is no need for optical-electrical-optical (OEO) conversion when using optical amplifiers, rather than regenerators, on the physical link. Although not yet prevalent, direct optical interfaces to DWDM equipment can also eliminate the need for an OEO function. While optical amplifiers are a major factor in the ability to extend the effective range of DWDM, other factors also come into play. For example, DWDM is subject to dispersion and nonlinear effects. These effects are further discussed in the "Optical Fibers" section on page 2-5.Many components, such as the optical add/drop multiplexer (OADM), are passive and therefore continue to work, even if there is a power cut. In addition, these components tend to have a very high mean time between failures (MTBF). Protection schemes implemented on DWDM equipment and in the network designs are at least as robust as those built into SONET. All these factors contribute to better performance and lower maintenance in the optical network.

Network Management Capability

One of the primary advantages offered by SONET technology is the capability of the data communication channel (DCC). Used for operations functions, DCCs ship such things as alarms, administration data, signal control information, and maintenance messages. When SONET is transported over DWDM, DCCs continue to perform these functions between SONET network elements. In addition, a DWDM system can have its own management channel for the optical layer. For out-of-band management, an additional wavelength (for example, a 33rd wavelength in a 32-wavelength system) is used as the optical supervisory channel (OSC). For inband management, a small amount of bandwidth (for example, 8 kHz) is reserved for management on a per-channel basis.

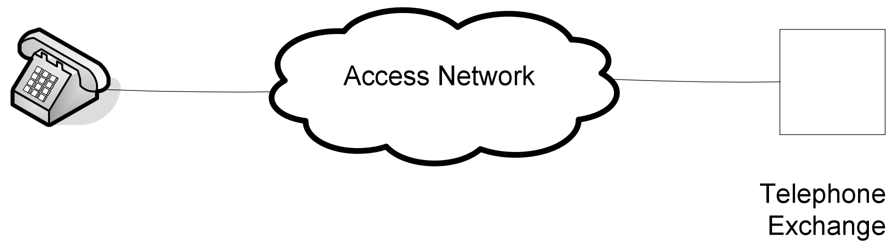
Additional Benefits

The shift in the makeup of traffic from voice to data has important implications for the design and operation of carrier networks. The introduction of cell-switching technologies such as ATM and Frame Relay demonstrates the limitations of the narrow-band, circuit-switched network design, but the limits of these technologies are being reached. Data is no longer an add-on to the voice-centric network, but is central. There are fundamentally different requirements of a data-centric network; two of these are the aggregation model and the open versus proprietary interfaces. Aggregation in a voice-centric network consists of multiplexing numerous times onto transmission facilities and at many points in the network. Aggregation in a data-centric network, by contrast, tends to happen at the edge. With OC-48 (and higher) interfaces readily available on cell and packet switches, it becomes possible to eliminate costly SONET multiplexing and digital cross-connect equipment. OC-48 connections can interface directly to DWDM equipment. Finally, service providers and enterprises can respond more quickly to changing demands by allocating bandwidth on demand. The ability to provision services rapidly by providing wavelength on demand creates new revenue opportunities such as wavelength leasing (an alternative to leasing of physical links or bit rate-limited tunnels), disaster recovery, and optical VPNs.

Chapter – 2 (Basic Telecommunication Concepts – FAQ)

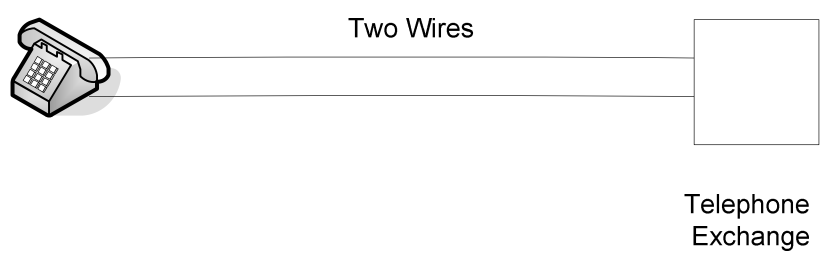
1. What is a telephone Access Network?

The network used to connect the telephone-to-telephone exchange.



2. What is Wired and Wireless?

The connection between telephone and exchange can be two copper wires or radio wave. If copper wires are used it is called wired connection.



It is required two wires to connect the telephone and exchange. It is called a local loop.

This connection also can be provided by a radio wave. It is called a Wireless Local Loop (WLL).

3. What is a fixed telephone?

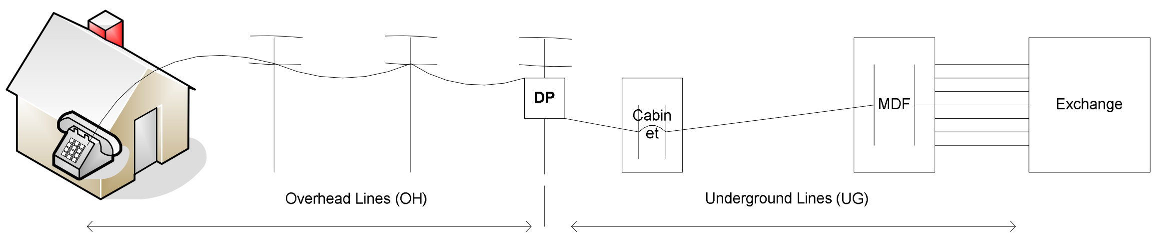
If the telephone is not moved from its position it is called a fixed telephone.

4. Who are the fixed line operators in Sri Lanka?

- Sri Lanka Telecom
- Suntel
- Bell

SLT provides both wired and wireless local loops. Suntel and Bell provides wireless local loops only, therefore they are called WLL operators.

5. Show the path of wired connection



6. What are the devices/Stages involved to a wired connection?

- Telephone exchange
- MDF
- Cabinet
- DP
- Telephone

7. What is the function of telephone exchange?

When a telephone dials a number it is received by the telephone exchange and analyses the number and connect to the relevant telephone via the remote exchange.

For the calls received from the remote exchange the originated telephone exchange checks whether the relevant telephone is free or busy. If the telephone is busy the busy tone is sent to the other exchange. If the telephone is not busy the ring back tone is sent to the other exchange.

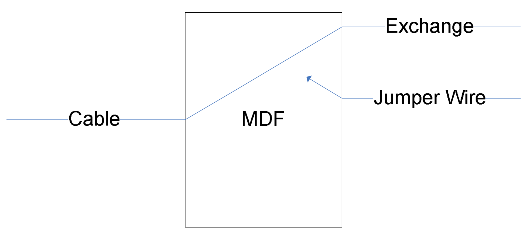
Note: These are the basic functions of a telephone exchange. There are many other functions also dine by the telephone exchange.

8. What is MDF?

MDF stands for Main Distribution Frame.

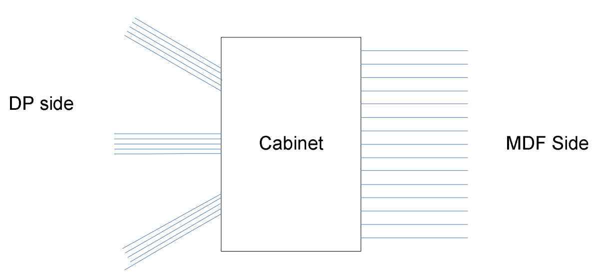
MDF is the interface between telephone exchange and telephone cables. At the time of installation of exchange all telephone channels (line cards) are terminated at the MDF.

The multi pair telephone cables are terminates of the other side.



When a new connection is given the exchange side relevant point and cable side relevant point are connected by using another small cable. This small cable is called a jumper wire. MDF is a passive device. (No need of electrical power to operate.)

9. What is a Cabinet?



50 pair, 100 pair, 200 pair etc. multi pair cables are drawn from MDF to cabinet. From the cabinet 10 pair, 5 pair cables are drown up to Distribution Points (DPs).

Primary cable - MDF to cabinet cable is called primary cable.

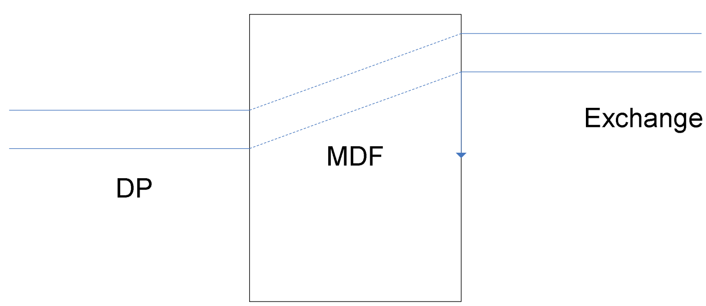
Secondary Cable - Cabinet to DP cable is called Secondary Cable.
Primary cable to Secondary cable jumpering is done at the cabinet.

10. What is a DP?

A 10 pair or 5 pair cable is laid from Cabinet to Distribution point (DP). That means, there are 10 loops or 5 loops in a DP. When a new connection is provided a pair of cable is drawn from DP to home. That means normally 10 or 5 telephone new connections can be provided from a DP.

11. What are the works to be done to provide a telephone new connection?

- DP to house cabling (Normally Over Head cabling)
- Jumpering at MDF
- The telephone exchange identifies a free line. It is connected to the Allocated loop.



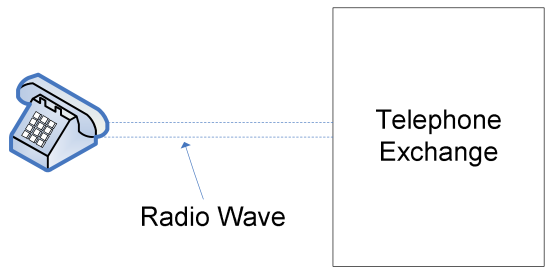
- Allocate a telephone number from the exchange and program it in the exchange.

  In order to provide a new connection all above functions have to be completed.

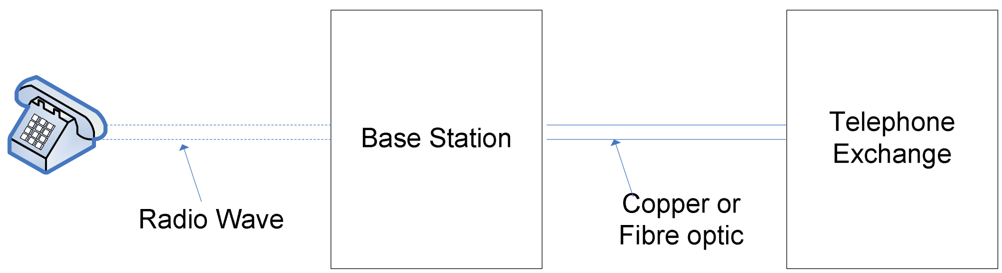
12. What is a Wireless Access Loop?

The telephone

And exchange is connected by using a radio wave.



In practical scenario, an intermediate unit is used. It is called a Base station.



Telephone is connected to Base Station by using a radio wave. The Base Station is connected to telephone exchange by using copper or fibre optic.

13. What are the components in Base Station?

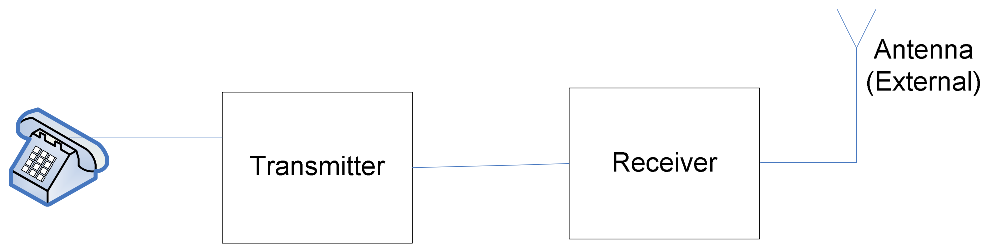
- Transmitter
- Receiver
- Antenna

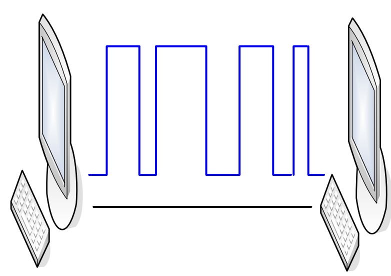
Normally Transmitter and Receiver come as one unit and it is called Base Transceiver Station (BTS).

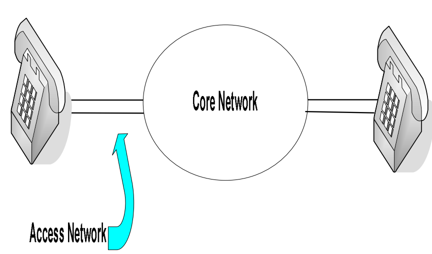
14. What are the components of Customer Premises?

- Transmitter
- Receiver
- Antenna
- Telephone

The whole set is called Customer premises Equipment (CPE).







The transmitter and Receiver need electrical power to operate. If they are in the telephone unit it should be provided the electrical power. (e.g.: Normal CDMA telephone)

15. What is a Radio Wave?

Any signal energy travels through the air (atmospheric) as a wave.

16. What are the characteristics of a wave?

It has a frequency (f), wave length (λ), and Speed (V).

The letter λ is a Greek Letter and it is called Lambda.

The relationship of those three are,

$$V=f\lambda$$

17. How a signal travels through copper cable and fibre optic cable?

Signal travels as a Wave.

18. Is there any difference between Radio Wave and Copper and Fibre Waves?

No, all belong to electromagnetic waves.

19. What is an Electromagnetic wave?

The energy travels as electric energy and magnetic energy.

20. What is Frequency?

It is one of the characteristics of a wave measured in Hertz (Hz).

Normally frequency is measured in kHz, MHz,GHz,THz. (Eg: kHz – Kilo Hertz, T - Tera - 10^{12})

k – kilo - 10^3

M – Mega - 10^6

G – Giga - 10^9

21. What is Wave Length?

It is another characteristic of a wave and measured in meters. Normally measures in millimeter (mm), micrometer (µm), nanometer (nm).

µ is a Greek letter and it is called mew.

m – milli - 1/1000 -10^{-3}

µ - micro - 10^{-6}

n – nano - 10^{-9}

Normally the electromagnetic waves in fibre are measured as a wavelength, not as frequencies.

Eg:1300nm, 1550nm

22. What is frequency spectrum?

The usable electromagnetic frequency range (Band) is called frequency spectrum.

30 km	3 km	300 m	30 m	3 m	30 cm	3 cm	3 mm		
VLF	LF	MF	HF	VHF	UHF	SHF	EHF	Infrared	Visible Light
10 kHz	100 kHz	1 MHz	10 MHz	100 MHz	1 GHz	10 GHz	100 GHz		

23. What is a transmission media?

An electromagnetic wave can be sent through

Copper, Radio, Fiber optic

They are called transmission media.

24. What are the usable frequency bands of different transmission medias?

Copper	-	from Hz to Mhz
Radio	-	from kHz to GHz
Fibre optic	-	T Hz range

25. What is guided media?

Copper and fibre optic are called guided media since the wave goes through the media. Those waves will not interfere with each other.

26. What is unguided media?

Radio waves go through air (free space). There is no physical connection between the transmitter and receiver. Therefore free space is called unguided media.

If two signals are transmitted with same frequency these two can be mixed up and both will not be able to use. (Just like the two airplanes fly at same height of same route)

Therefore the usage (allocation) of frequencies should be controlled by a particular body. In any country if is controlled by a government body. In Sri Lanka Telecommunication Regulatory Commission (TRC) controls it.

The usable frequency band also called “frequency spectrum”.

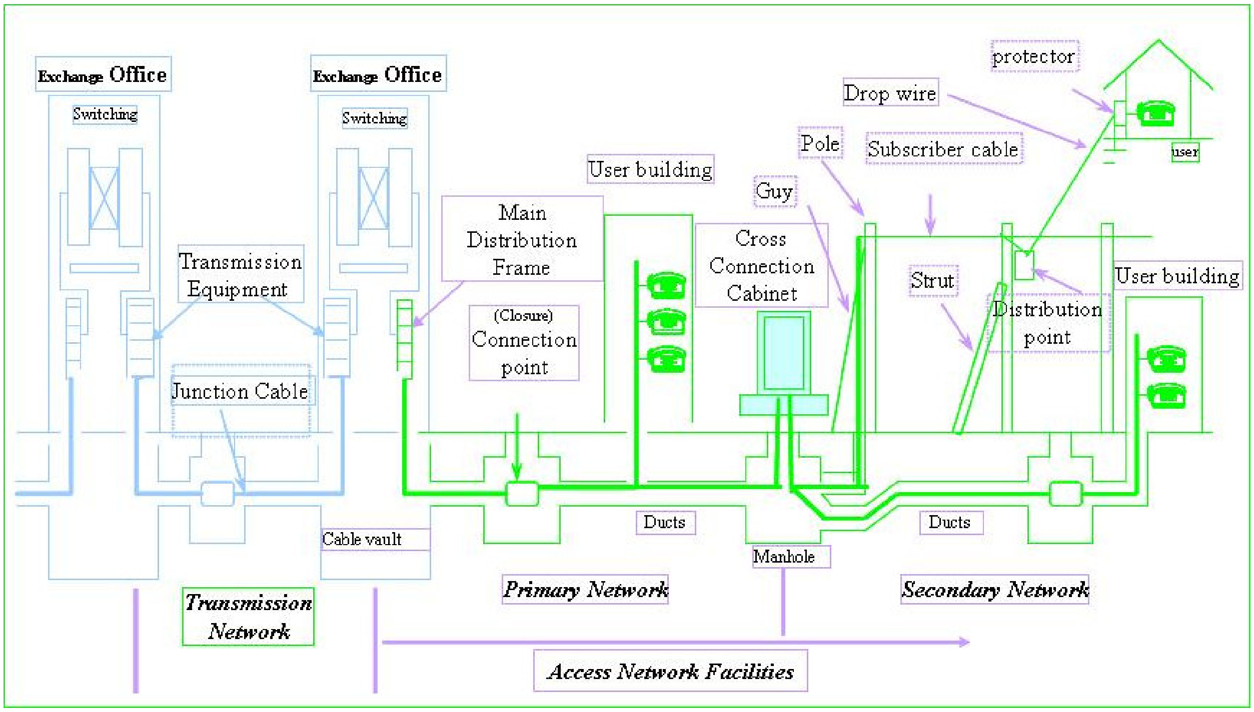
27. What are the standard frequency bands?

Different frequency ranges are given different names.

Designation		Frequency	Wavelength
ELF	Extremely low frequency	3Hz to 30Hz	100'000km to 10'000 km
SLF	Super low frequency	30Hz to 300Hz	10'000km to 1'000km
ULF	Ultra low frequency	300Hz to 3000Hz	1'000km to 100km
VLF	very low frequency	3kHz to 30kHz	100km to 10km
LF	low frequency	30kHz to 300kHz	10km to 1km
MF	medium frequency	300kHz to 3000kHz	1km to 100m
HF	high frequency	3MHz to 30MHz	100m to 10m
VHF	very high frequency	30MHz to 300MHz	10m to 1m
UHF	ultrahigh frequency	300MHz to 3000MHz	1m to 10cm
SHF	Super high frequency	3GHz to 30GHz	10cm to 1cm
EHF	extremely high frequency	30GHz to 300GHz	1cm to 1mm

Different frequency bands are suitable for different usages (applications). Different band also gives different letters. e.g.:

1-2 GHz 30-15 cm ....L Band
2-4 GHz 15-7.5 cm ....S Band
4-8 GHz 7.5-3.75 cm ....C Band
8-12 GHz 3.75-2.50 cm.......... X Band
12-18 GHz 2.5-1.67 cm ....Ku Band
18-27 GHz 1.67-1.11 cm ....K Band
27-40 GHz 1.11 cm-7.5 mm ..Ka Band


Exchange Office
Switching
Exchange Office
Switching
Transmission Equipment
Junction Cable
Main Distribution Frame
(Closure) Connection point
User building
Cross Connection Cabinet
Guy
Pole
Subscriber cable
Drop wire
protector
user
Strut
Distribution point
User building
Cable vault
Ducts
Manhole
Ducts
Transmission Network
Primary Network
Secondary Network
Access Network Facilities

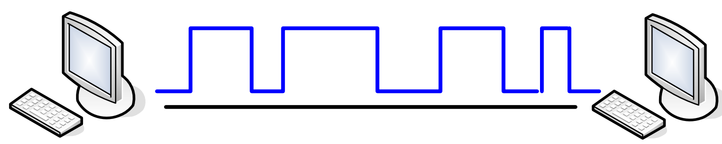
Chapter – 3 (Introduction to Core and Access Networks)

VoiceCommunication

Voice

Data Communication

Data

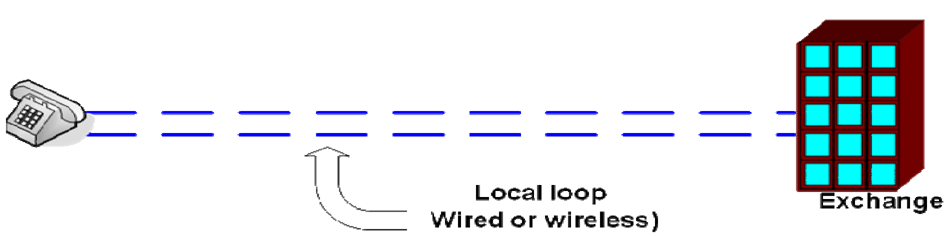


Voice Communication

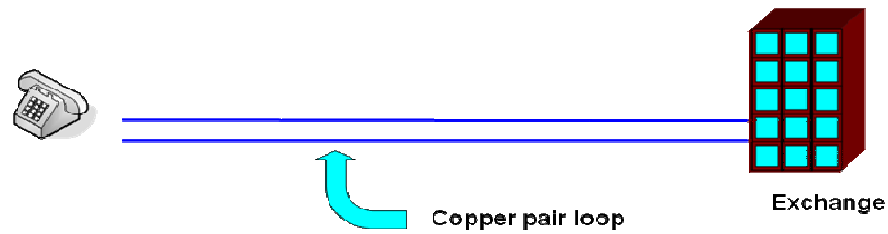
- ➢ Wired - Copper cable pair
- ➢ Wireless - Radio signal
- Fixed
- Mobile

Fixed

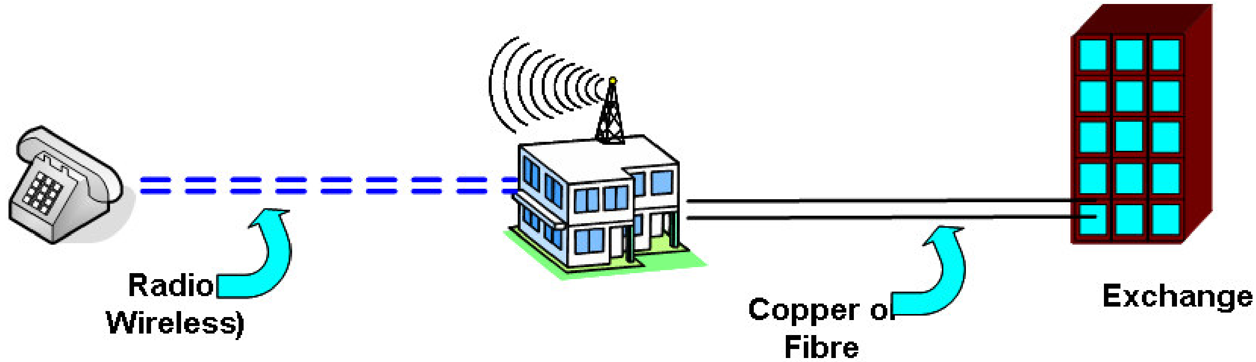
- ➢ Wired
- ➢ Wireless (WLL)



Fixed Wired

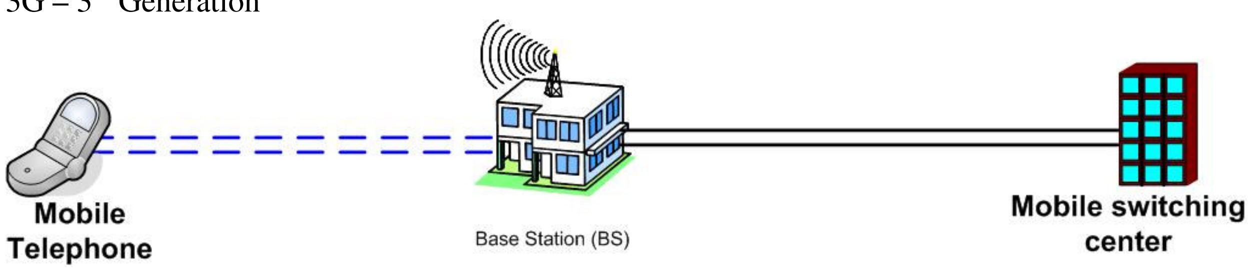


Fixed – Wireless (CDMA)



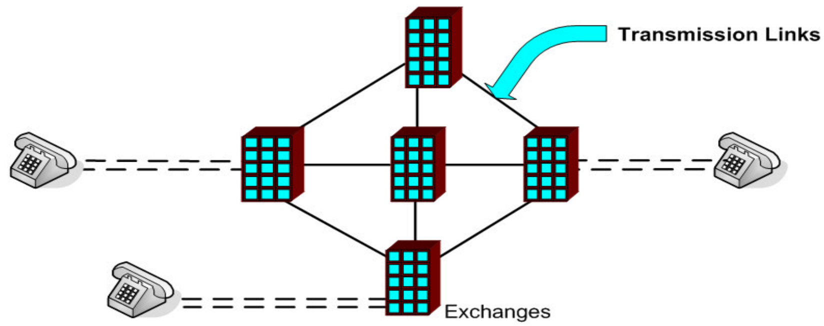
Mobile Network

1 G – 1st Generation
2G – 2nd Generation
3G – 3rd Generation

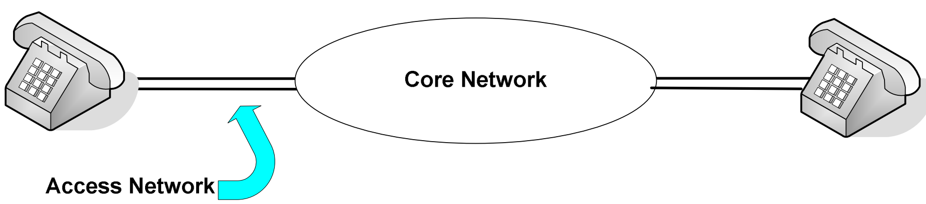


Mobile Voice
- Same as PSTN.
- Mobility Management

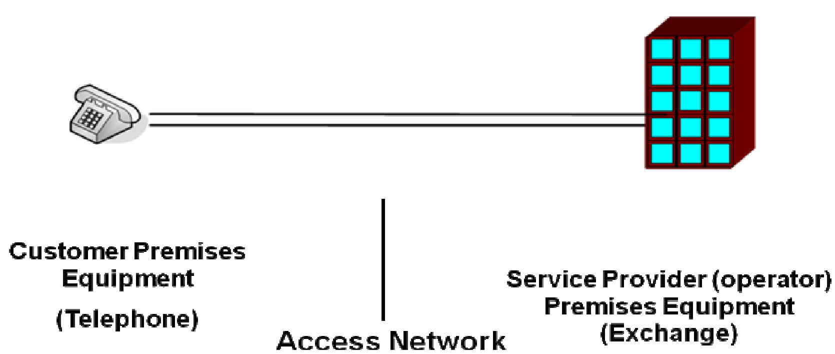
Public Switched Telephone Network (PSTN)



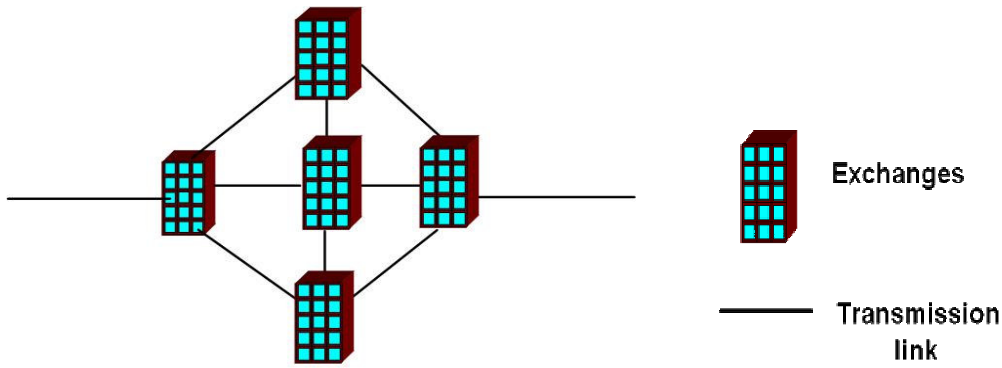
PSTN



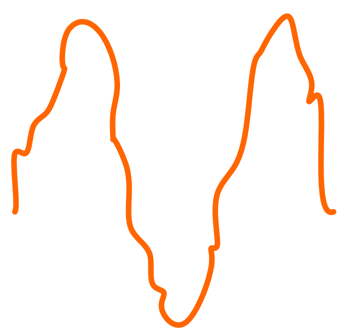
Access Network



Core Network



Voice Signal



Telephone Channel

0 – 4 kHz

Analog Signal

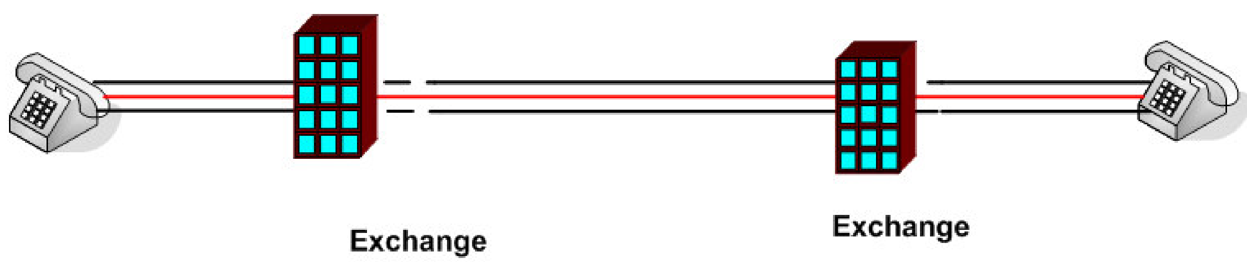
Voice Services

- Normal Telephone Service
- (Plain Old Telephone System - POTS)
- Voice Mail
- Call Forwarding
- Call Conferencing

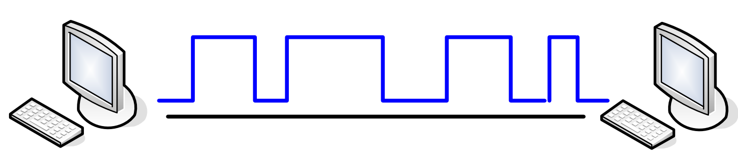
Circuit Switching

Telephone Exchange, Mobile Switching Centre

- Connect users (Telephones)
- Circuit Switching
- Real Time Connection (No Delay)



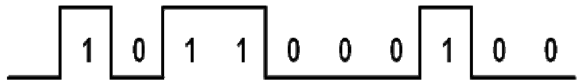
Data Communication



Data

Digital Signals
1010111011

Electrical Signal



Bits

1,0 - Bits

11011 - 5 Bits
110 - 3 Bits

Data Traveling

Bit Rate -
Measurement of Bit Speed

Bits Per Second –
How many Bits Travel in one second
1 b/s, 10 b/s

Bit Rate

- k – kilo - 1000
- M – Mega - 1000000
- G – Giga - 1000000000
- T – Tera - 1000000000000

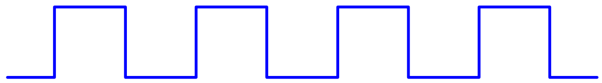
- 1 kb/s - 1000 b/s
- 1 Mb/s - 1000 Kb/s
- 1 Gb/s - 1000 Mb/s
- 1 Tb/s - 1000 Gb/s

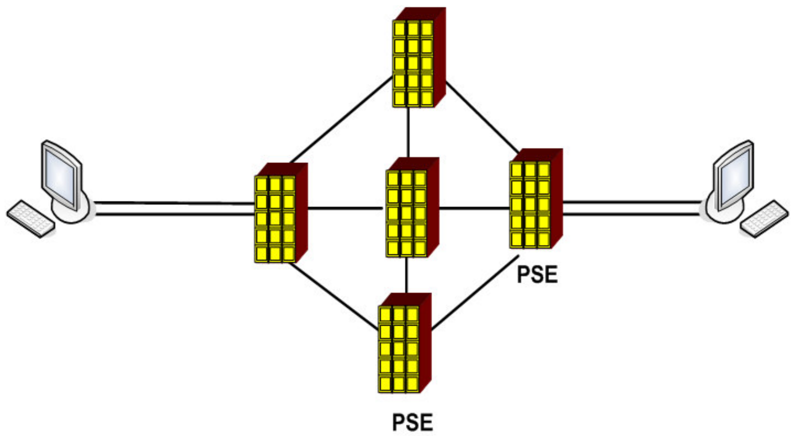
Data Services

- World Wide Web (www)
- Electronic Mail (E - mail)
- Short Message Service (SMS)
- Multi Media Services (MMS)

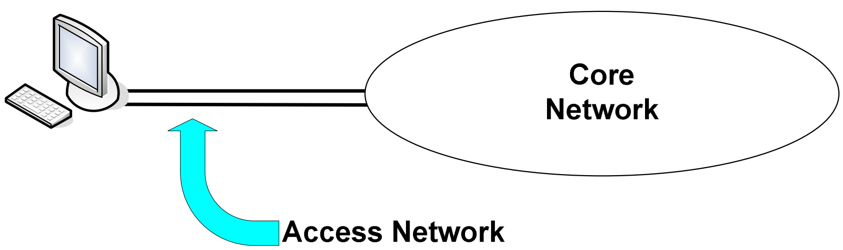
Data Packet

Group of Bits



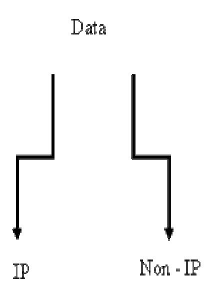


Packet Switching Exchange (PSE)



IP data & Non - IP data

- IP – Internet Protocol
- Standards defined by internet society
-

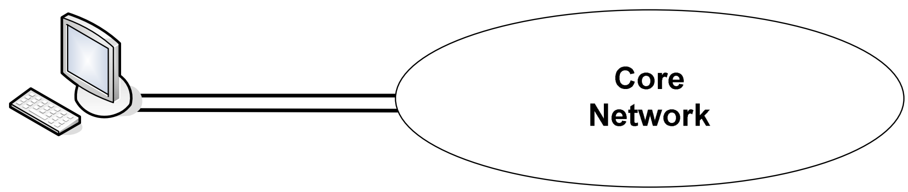


IP Data Services

- World Wide Web
- E - mail

Non – IP Data Services

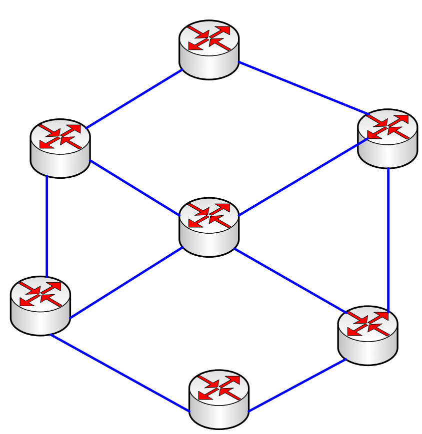
- SMS - Short Message Service
- MMS – Multimedia Message Service



- Access Network - IP
Non – IP
- Core Network - IP
Non – IP

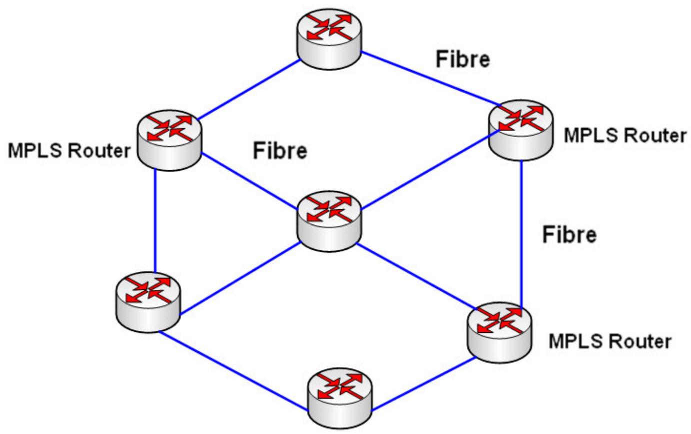
IP Core Network

PSE is called Router

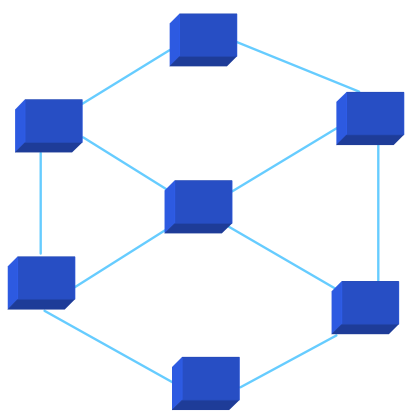


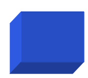
IP - MPLS Core Network

PSE is called Router



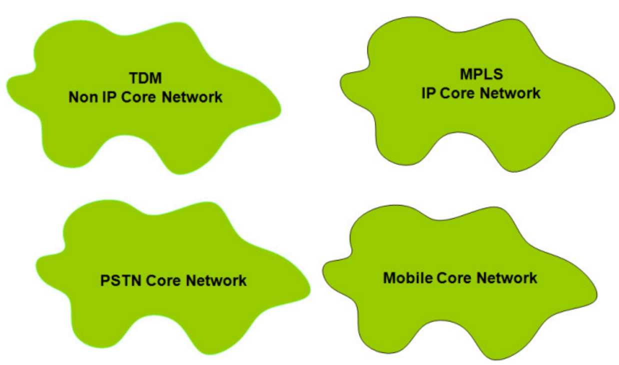
TDM Core Network





Multiplexer

TDM - Time Division Multiplexing



Can we send Data through PSTN?

- Yes

  - Shall Convert to an analog signal Similar to Voice
  - Use a Modem

Speed : 9.6 kbps, 14.4 kbps, 28.8 kbps, 56 kbps

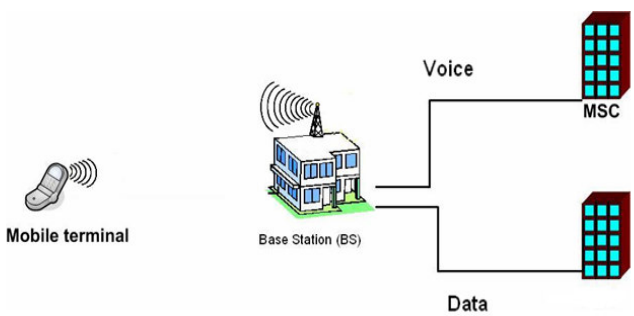
Can We send Data through Mobile Network?

- Yes
  - Same as PSTN
  - Use a Modem

- 1 G – 9.6 kb/s
- 2 G – GSM
  Circuit Switched Data (CSD) - 9.6 kb/s
  High Speed CSD (HCSD) - 43.2 kb/s

Can Mobile Network send Packet Switched Data?

- Yes
  - 2 G and Upwards



MSC – Mobile Switching Center
DSE – Data Switch Exchange

GSM Packet Switch Data

GPRS - General Packet Radio System
EDGE - Enhanced Data Rate for Global Evaluation

- GPRS – (Up to 60 kbps)
- EDGE – (Up to 236 kbps)

Can CDMA Network send Packet Switched Data ?

- CDMA Data

- Data through circuit switch (9.6 kbps)

- Data through Packet switch – (EV-DO)
  Download up to 2 Mb/s

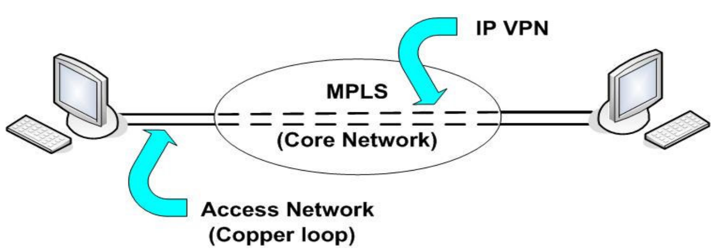
EV-DO – Evolution-Data Optimized or Evolution-Data only

3 G Packet Switched Data

- High Speed Downlink Packet Access (HSDPA)

    Downlink Speed – upto 14 Mb/s

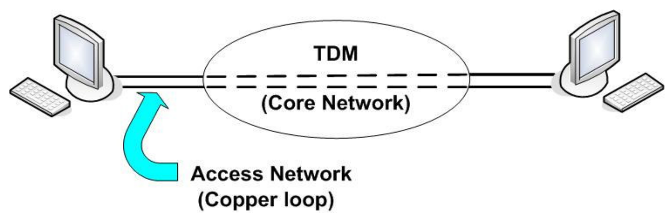
Data through MPLS core network



VPN – Virtual Private Network

- Bandwidth is shared
- Can use for IP data – upto 155 Mb/s

Data through TDM core network



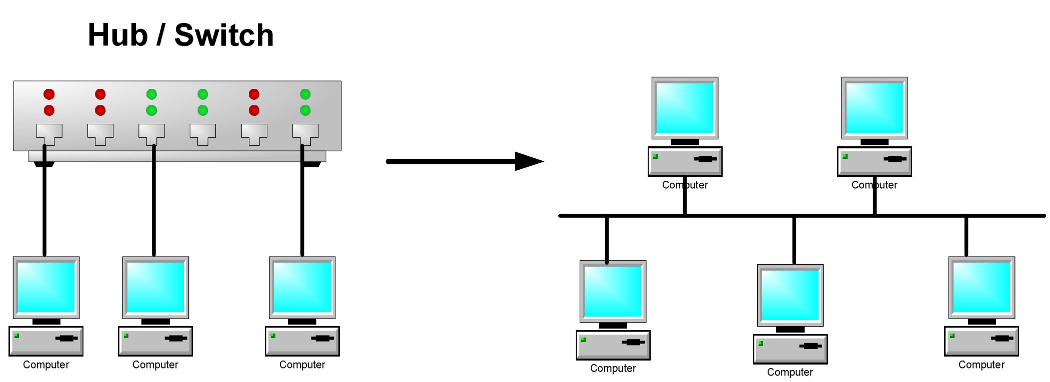
- This is Called a "Leased Line"
- Dedicated Channel
- Speed : N x 64 kb/s, N = 1, 2, …..30
  64 kb/s, 128 kb/s, 192 kb/s, 2048 kb/s

Computer Networks

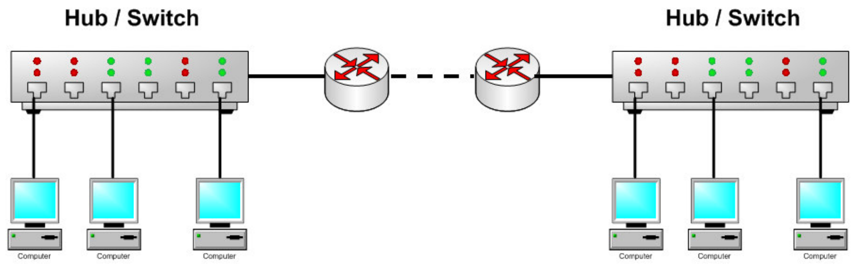
- PAN -about 10m radius
- LAN -within a building or within a campus
- MAN -several km radius
- WAN -countrywide, worldwide

- PAN - Personal Area Network
- LAN - Local Area Network
- MAN - Metropolitan Area Network
- WAN - Wide Area Network

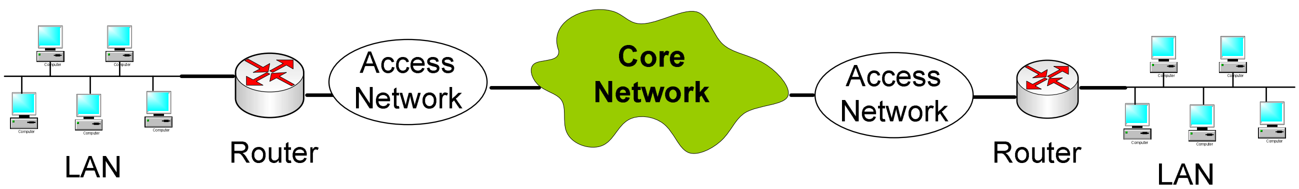
LAN



WAN



WAN link - Leased line
- IP-VPN

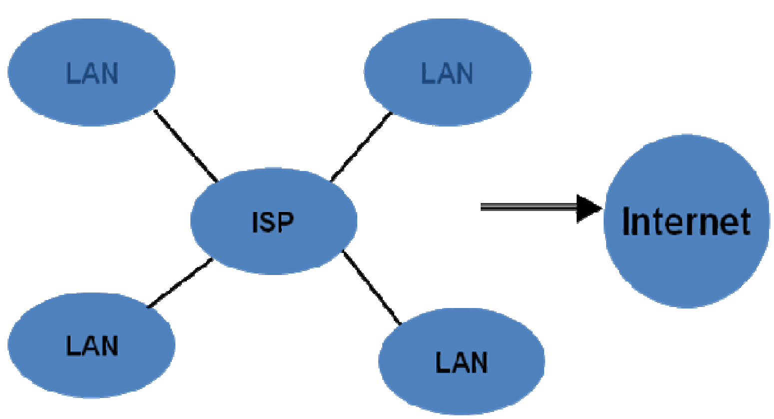


IP Address

Use to identify a computer or device.
There are two types,

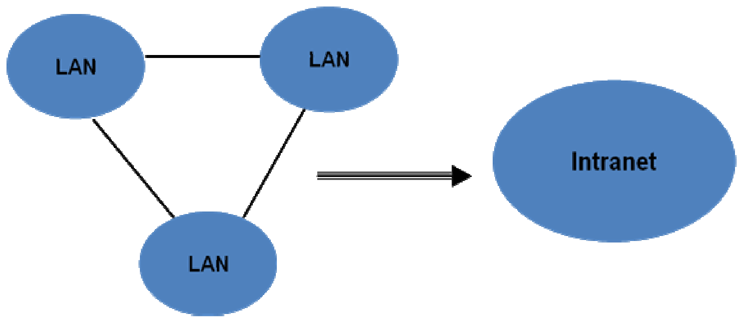
- Public IP address
- Private IP address

Internet



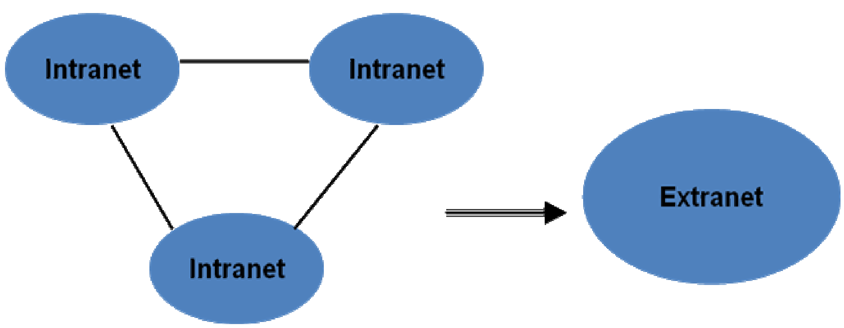
ISP – Internet Service Provider
Public IP addresses are used

Intranet



- ➢ Private IP addresses are used.
- ➢ Public cannot access.
- ➢ Closed network.

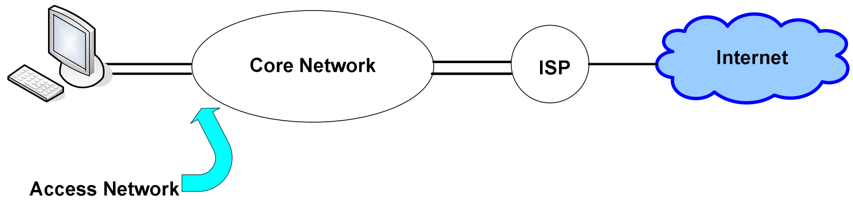
Extranet



IP Services

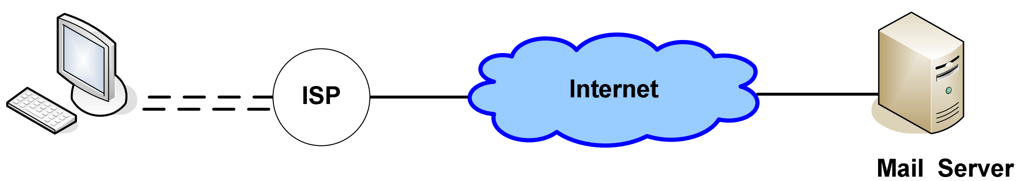
- Services access through Internet
- Services access through Intranet, Extranet
- Data travels as IP packets
- Egg : www
  E-mail

How to access Internet



Connection – PSTN - Dial-up
TDM - Leased line
MPLS - IP-VPN

IP Services – E-mail

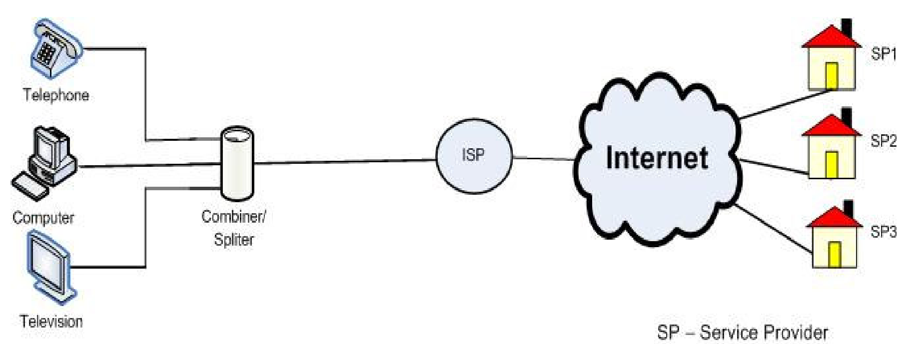


Advantages of IP Services

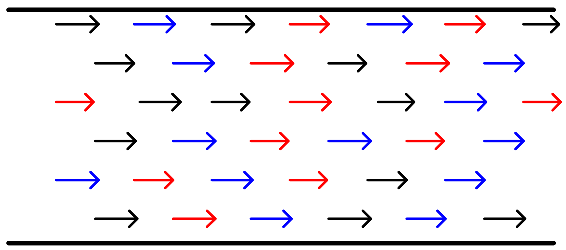
- Transmit as IP packets
- All IP packets are similar pattern. Therefore, treats
- them equally in the network.
- Can differentiate the services at computer/terminal

Other IP Services

- Since IP is very flexible, non data (e.g. Voice) also convert to IP packets
  E.g. VoIP, IPTV, Video streaming, audio streaming, Games



Broadband



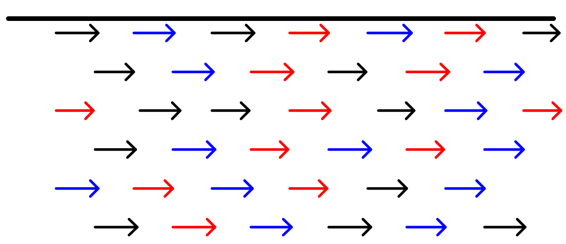
- More services together
- More Information
- More bandwidth required (Broadband required)

Narrowband and Broadband

- Narrowband – Less than 256 kb/s
- Broadband – More than 256 kb/s
- There is no User definition
- Change from Time to Time

May be after several Years, Broadband…..
More than 10 Mb/s

Converged Services

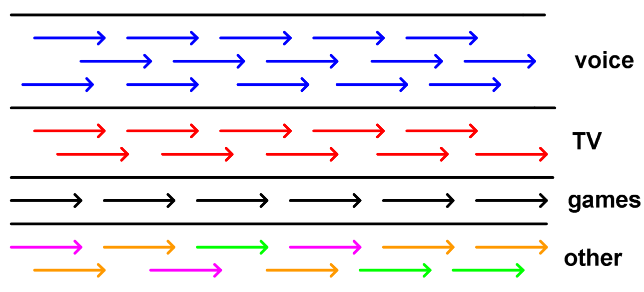


- All IP Packets travel together
- No guaranteed bandwidth per service
- This is called " Best Effort services"

Disadvantage of Best Effort Service

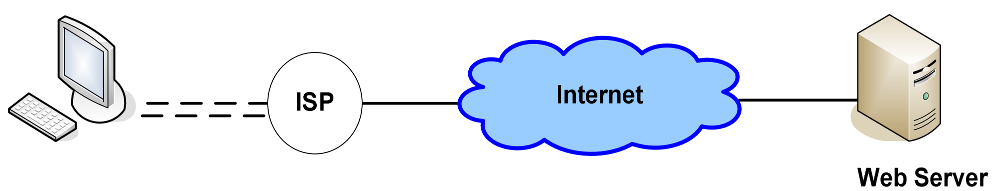
- Not prioritize the traffic
- Not suitable for interactive services such as Voice, Games etc

Quality of Services



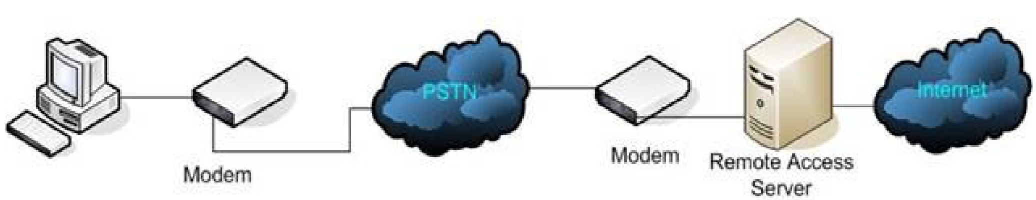
- Guaranteed bandwidth
- Prioritize packet at Router
- Less delay for high priority services

IP Services – www

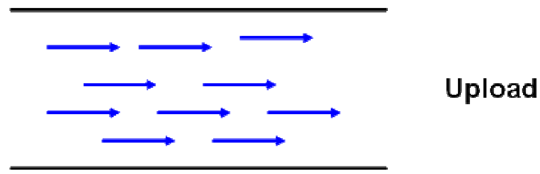


Access to WWW Service

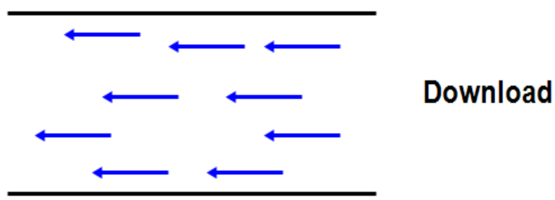
- Dial-up



Dial – up Disadvantages



OR

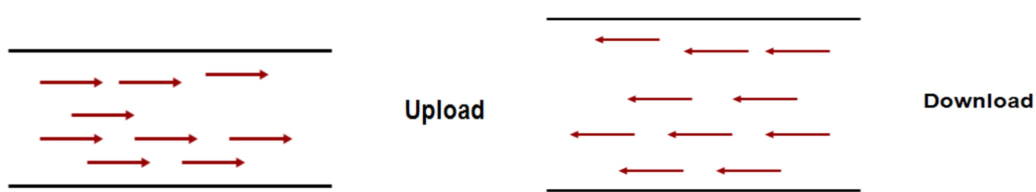


Cannot upload and download simultaneously

Dial-up Disadvantages

- Cannot access internet and telephone line simultaneously.
- Telephone charges while using internet.
- Dial and Connect to Internet
- Speed is bellowing 56 Kbps.
- Cannot Upload and Download simultaneously.

Characteristics of WWW

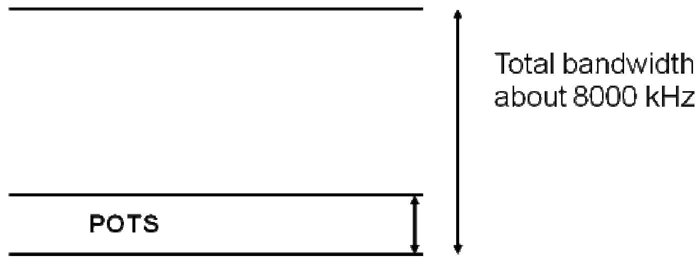


Download needs more bandwidth than upload.
Upload bandwidth $\neq$ Download bandwidth - **Asymmetric**

Copper Loop (DSL Technologies)

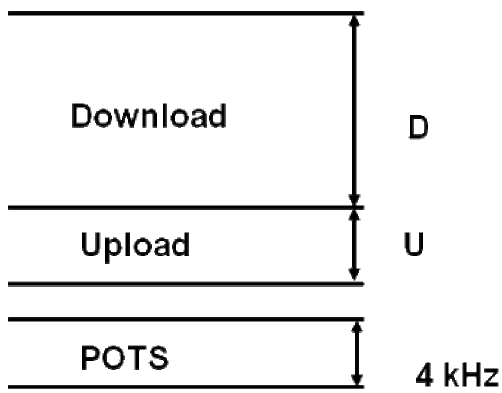
DSL – Digital Subscriber Line

Bandwidth of Copper Loop



POTS (normal telephone) uses only 4kHz.
Hence, bandwidth wastage

Asymmetric Digital Subscriber Line (ADSL)



Download has more bandwidth

- Residential Package ; upload=128 kb/s, download=512 kb/s
- Corporate Package ; upload=512 kb/s, download=2048 kb/s

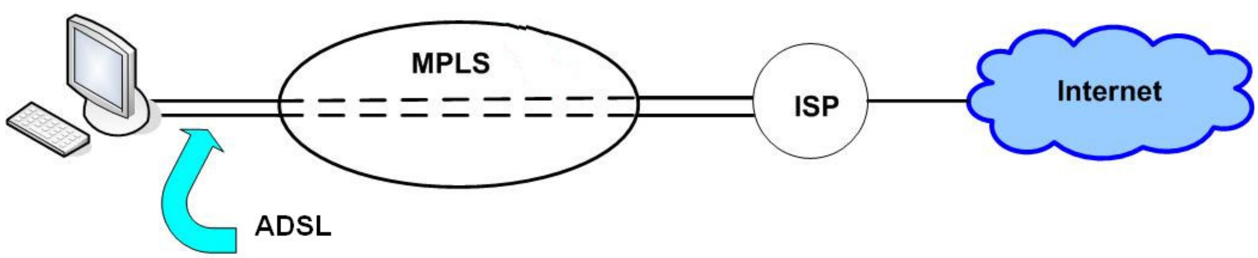
Advantages of ADSL

- Simultaneous access to Internet and telephone
- No additional telephone charges while using internet.
- 24 hour Internet connectivity
- Very High Speed

Disadvantages of ADSL

- Maximum distance – about 5 km
- Distance increase – Download speed decrease

ADSL for WWW



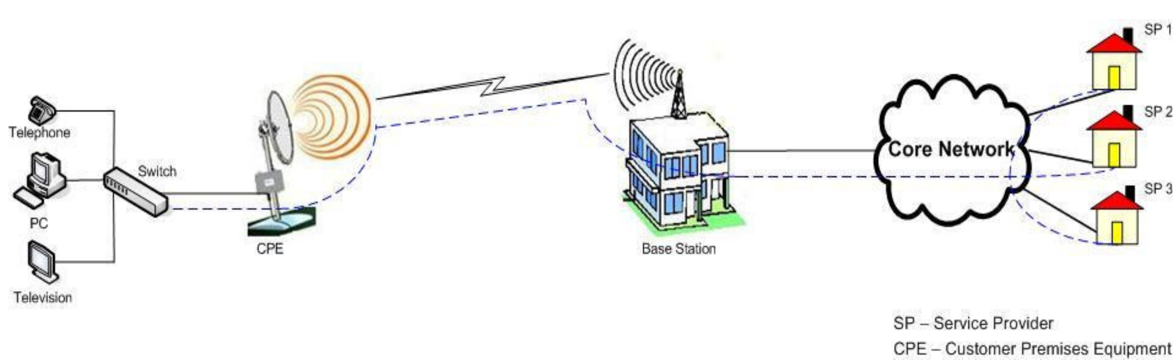
ADSL is an access technology

DSL Improvements

- ADSL2+
  Downloading speed - 16 Mbps
  Uploading Speed – 1 Mbps
- Very high speed DSL
  VDSL
  Downloading speed - 24 Mbps
- xDSL
  x = A, V ,S, H

Fixed Wireless Broadband Access

WiMAX - Worldwide Interoperability for Microwave Access



WiMAX Features

- Broadband radio access technology
- IP based data
- Converge service
- QoS
- Line of sight (LOS) about 15 km.
- Non-LOS (NLOS) about 3 km
- 3.5 GHz operating frequency

WiMAX Services

- Leased line
- WWW
- E-mail
- Voice over IP
- Games
- Video streaming
- Audio streaming

WiMAX standards

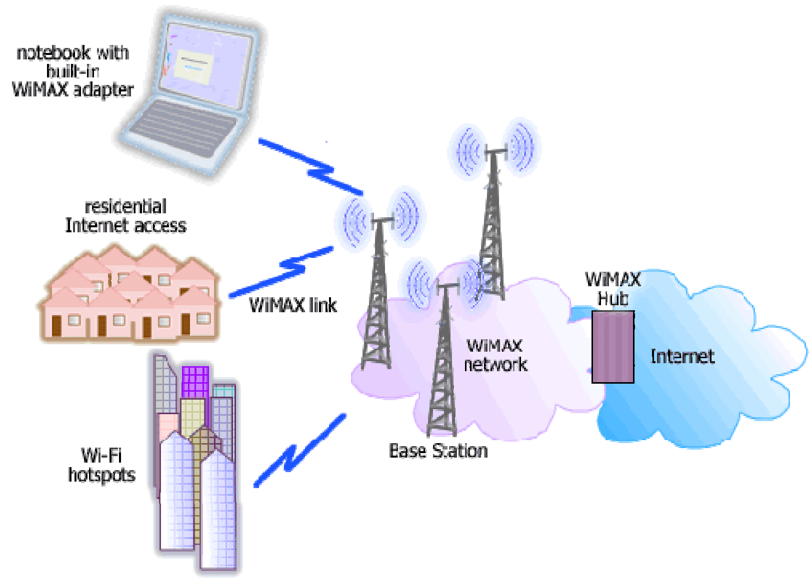
- IEEE 802.16d - Fixed WiMAX
- IEEE 802.16e - Mobile WiMAX

IEEE – Institution of Electric & Electronic Engineering

WiMAX Forum – www.wimaxforum.org

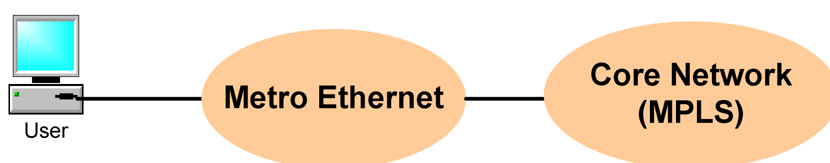
Differences between WiFi and WiMAX

- WiFi - Wireless LAN
  - Operates in 100m radius
- WiMAX - Access technology
  - LOS about 15km
  - NLOS about 3km



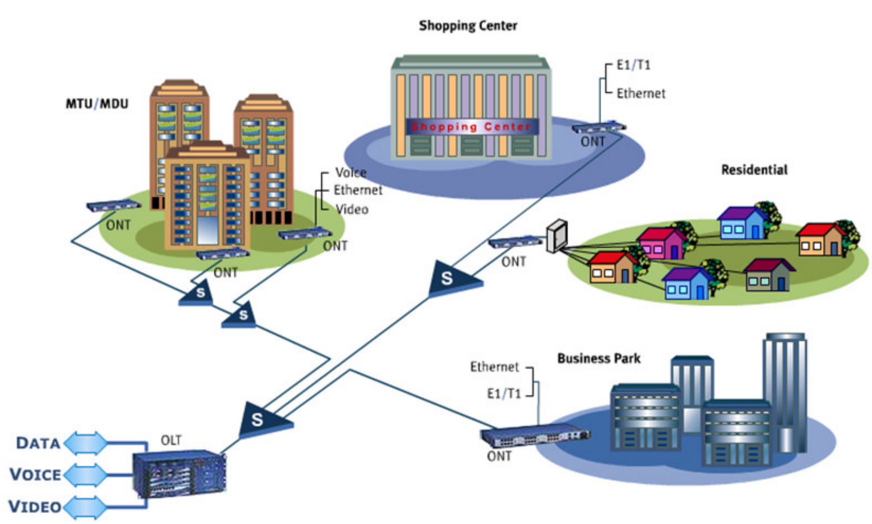
Fibre Broadband Access Technologies

Access technology – Metro Ethernet



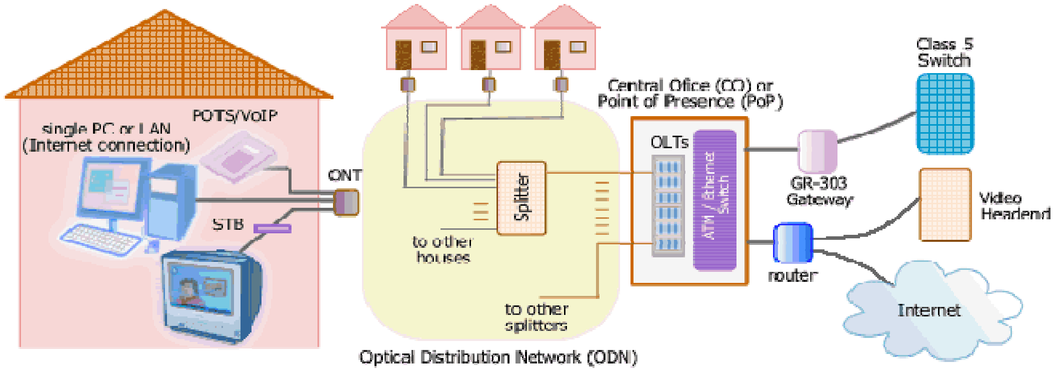
Bandwidth per user – upto 100 Mbps

Access network – Passive Optical Network (PON)



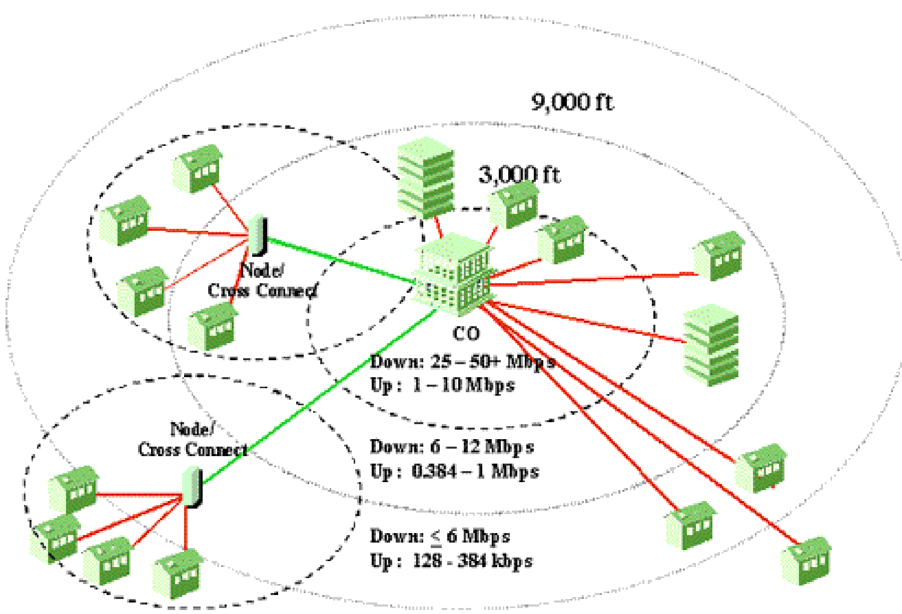
FTTH

FTTH (Fibre To The Home)



FTTN

FTTN (Fibre To The Neighborhood)



Access Technology Summary

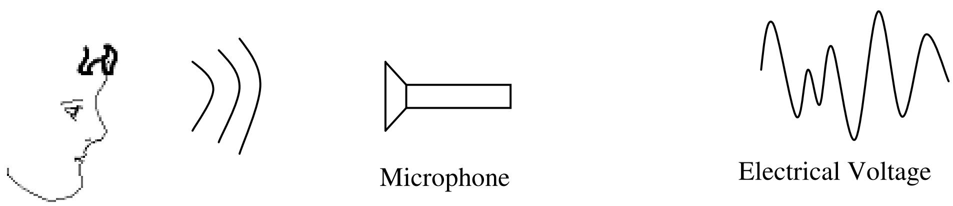
- Copper –
  - Dial-up – (PSTN)
  - ADSL (8 Mbps/1 Mbps),
  - ADSL2+ (16 Mbps),
  - VDSL (24 Mbps)

- Wireless (Fixed) –
  - CDMA/EVDO
  - WiMAX

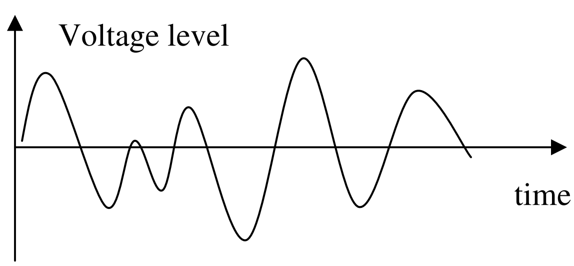
Chapter - 4 (Communication Fundamentals)

Analog Signal

A Signal is an electrical voltage or current, which varies with time. It is used to carry some information from one end to another. A typical example is a voice signal.



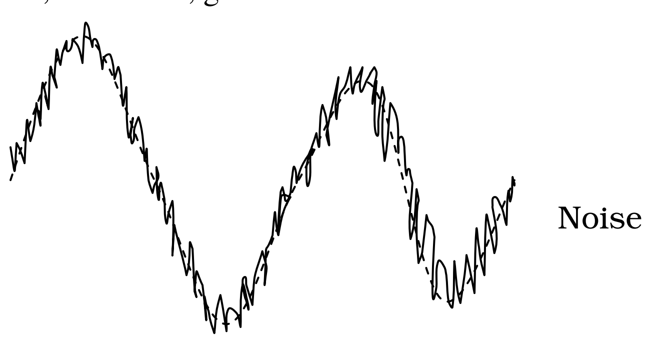
The microphone converts the sound signal to an electrical voltage.



This signal is continuously varying with time. This type of signal is called an analog signal.

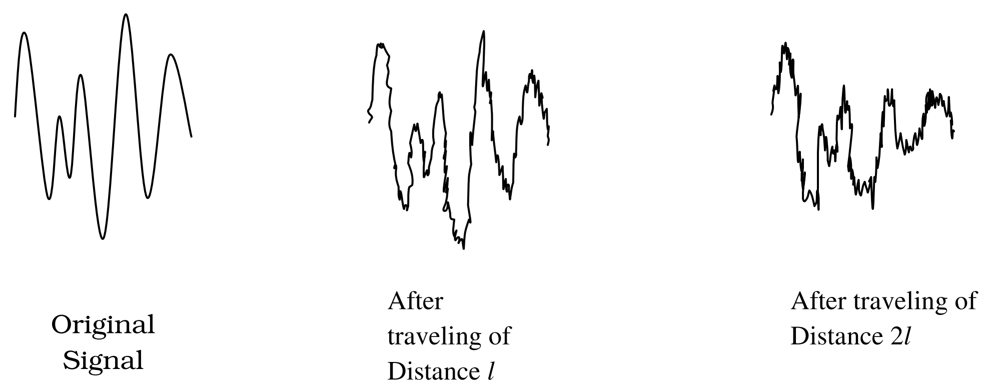
Noise

Noise is an unwanted signal. There are some freely moving electrons in the conductors. An electron movement is a current. Unwanted movement of electrons create unwanted currents. That is an unwanted signal or noise. Please note that there are some other ways of creating noise such as transistor noise, shot noise, galactic noise etc.

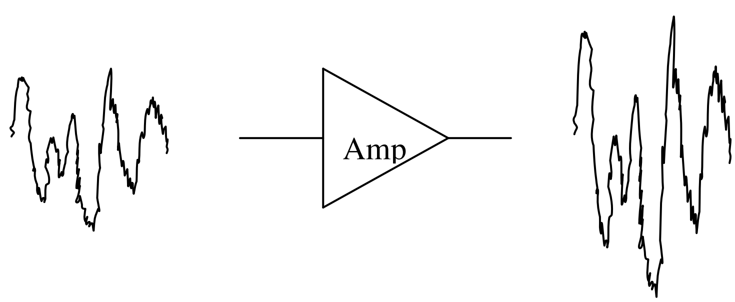


If any signal goes through a conductor it mixes with noise. The ratio of signal level and noise level is called the S/N or signal to noise ratio. The quality of a signal is measured as S/N. Higher the S/N, better the signal quality.

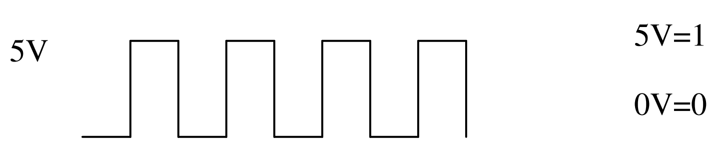
If an analog signal travels a long distance, more and more noise is added to it. Therefore, the S/N reduces and the signal quality is degraded. The biggest disadvantage of an analog signal is, the noise cannot be removed and it accumulates.



The other problem is, if the signal is amplified the noise is also amplified.



Digital Signal

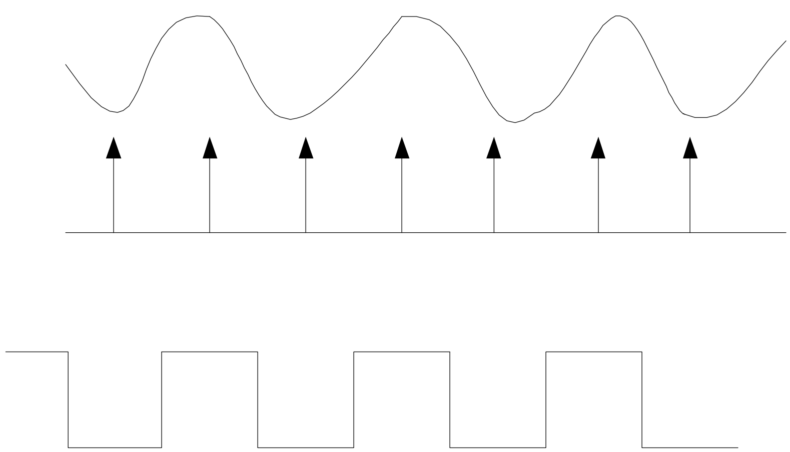


A discrete electrical signal, which has only two levels, is called a digital signal. These two levels are named as "1" and "0". Normally a digital signal has a fixed number of bits and travels within a particular duration. This is called the pulse rate or bit rate.

The digital signal also gets mixed with noise.

Original Signal After Transmission

The centre location of the digital signal can be identified by using a special bit pattern called a clock signal.



By checking the level at the centre location of each bit it can be decided whether it is a "0" or a "1". This process is called the regeneration of the signal. By this method, the original signal can be generated and noise can be completely eliminated. This is the main advantage of a digital signal over an analog signal.

But there is a possibility to change the bit from 1 to 0 or 0 to 1 due to high noise. This is called an error. However there are many methods to correct these errors. Hence at the receive end the original bit pattern can be obtained. Hence digital signal quality will not depend on the distance traveled by the signal.

How to convert an analog signal to a digital signal?

The most commonly used method is the Pulse Code Modulation. Normally all voice telephone channels use this method.

Voice telephone channel frequency band is = 0.3 kHz to 3.4 kHz.

The process can be described as follows.

(i) Sampling

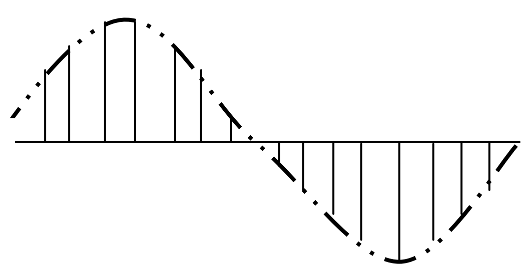
(ii) Quantizing

(iii) Encoding

What is sampling?

The samples of an analog signal are taken.

The sampled signal is called a pulse amplitude modulated signal.



It can be shown that the original signal can be constructed at the receive end using these samples.

Sampling Theorem

In order to completely reconstruct the original signal from the samples, the sample rate should be at least twice its highest frequency.

i.e. sampling rate $\geq 2 \times$ highest frequency

The highest frequency of telephone voice channel is 3.4 kHz.

Hence sampling rate $\geq 2 \times 3.4$

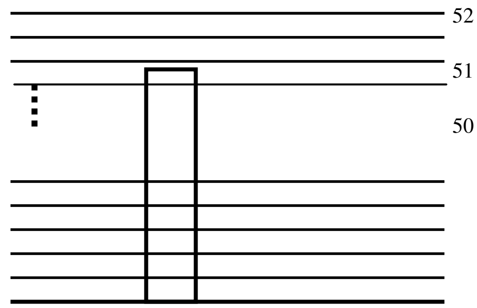
≥ 6.8 kHz

Hence a sample rate of 8 kHz is selected.

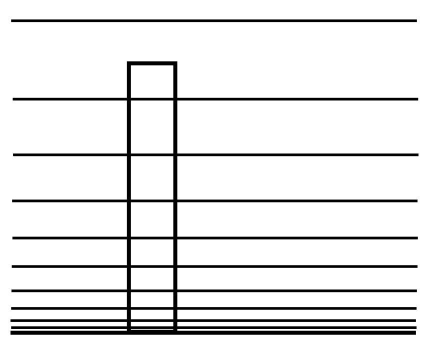
I.e. An analog signal is sampled at a rate of 8000 samples per second.

Quantizing

The samples are divided into many discrete levels. Then each sample is numbered according to their corresponding level.



There is no exact level for the above sample. The approximate level of the above sample is 50. Therefore the level of the sample is considered as 50. Hence an error will be introduced. This is called the quantizing error. This will reflect as noise at the receive end and it affects to the signal to noise ratio at the receive signal. It can be shown that, higher amplitude pulses will have high S/N and small amplitude pulses have low S/N. But we expect equal S/N for all pulses. In order to achieve this, non-linear quantizing is introduced.



It can be shown that using this method equal S/N can be obtained for all pulses.

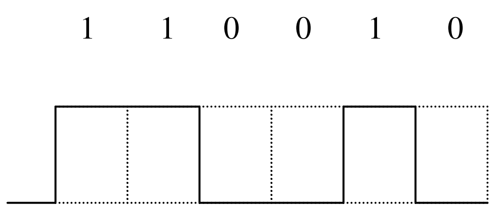
Encoding

After quantizing the corresponding level it is to be represented in some manner.

E.g. If the level is 50, it can be represented as,

Decimal	-	50
Hexa	-	32
Octal	-	62
Binary	-	110010

The 110010-bit pattern should be represented as an electrical signal, i.e. current or voltage. To represent a decimal number 10 voltage levels are required. Likewise 16, 8 and 2 voltage levels are required for hexa, octal and binary respectively. But practically representing more than two voltage levels is difficult. The most convenient and reliable method is using two levels. I.e. binary



This is called a bit stream.

Then we have to decide, how many quantizing levels are required. The more quantizing levels are used, more bits are required. It may cause to increase the bit stream and hence the bandwidth. Therefore, an optimum number of levels are to be selected. The standard number of levels is 256.

$$2^8 = 256$$

In order to represent 256 levels 8 bits are required. Hence each pulse is encoded to 8 bits.

1 sample	=	8 bits
Signal	=	8000 samples/sec
	=	8000 x 8 bits /sec

= 64000 bits/sec

= 64 kb/s

Therefore, bit rate of a digital telephone channel is 64 kb/s.

Analog Signal → 8000 samples/sec → 64kb/s

PCM

Modulation

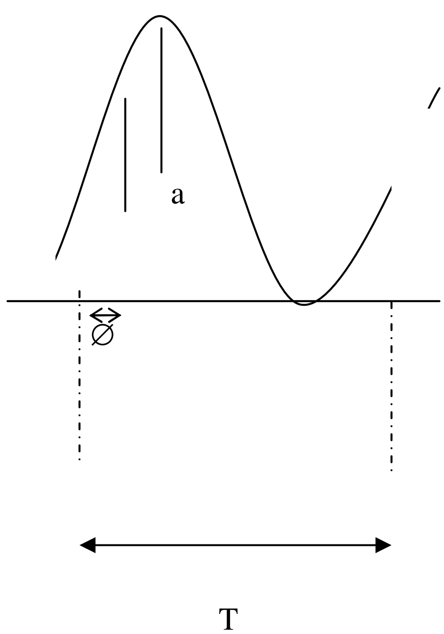
Modulation is a technique used to send information by modifying the characteristics of a basic electromagnetic signal. The basic signal is called the carrier signal.

The characteristics of a signal are amplitude, frequency and phase.

A signal can be represented by

$$a \text{ Sin } (\omega t + \varnothing)$$

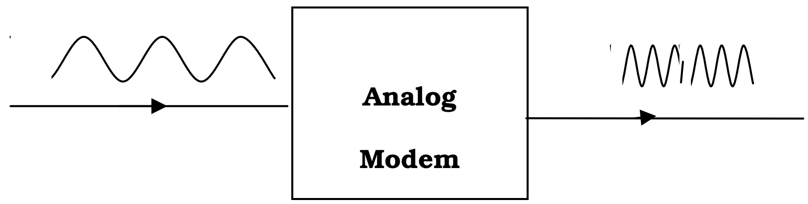
a - amplitude
ω - $2\pi f$ f - frequency
$\varnothing$ - phase



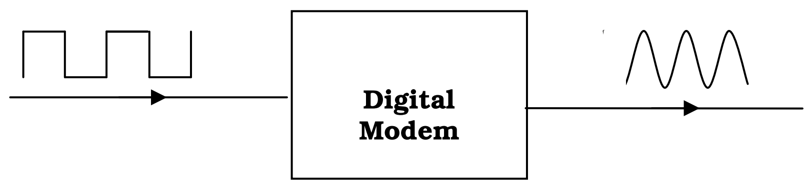
T = Period

$$f = \frac{1}{T}$$

Modulation can be used to convert a low frequency analog signal to a high frequency analog signal,

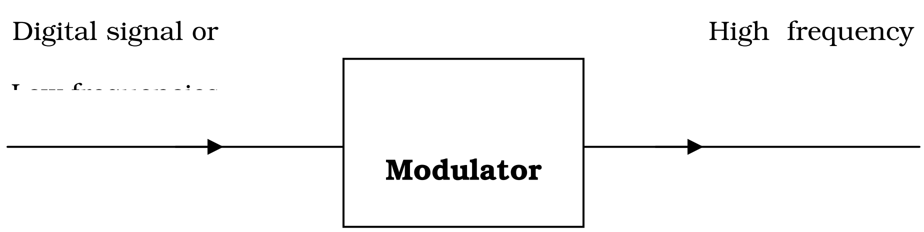


or a digital signal to an analog signal. For example a modem falls into the second category



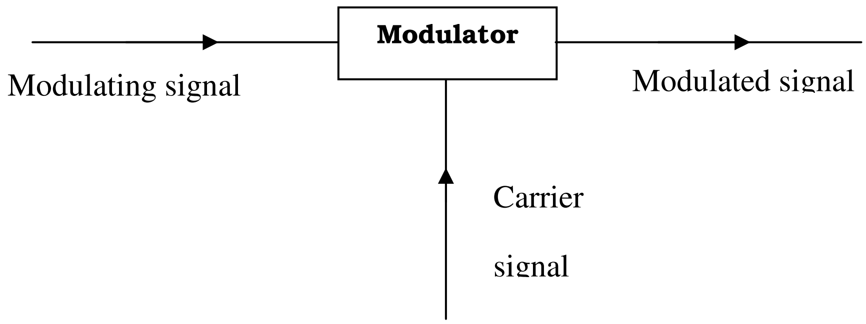
The input bit rate can be 9.6, 14.4, 19.2, 28.8, 56 kb/s. The output is an analog signal of frequency band 0.3 – 3.4 kHz.

Another application of modulation is to convert an analog or a digital signal to a very high frequency radio signal to transmit it through free space. [Broadband Radio Transmission]



[R*adio Transmission is discussed in another section*]

Modulation Process



Modulating Signal

This is the useful signal. This can be an analog signal or a digital signal. If the modulating signal is analog it is called analog modulation. If the modulating signal is digital, it is called digital modulation.

Carrier Signal

This is a high frequency analog signal.

Modulated Signal

The three characteristics of any signal are amplitude, frequency and phase. One of these characteristics are changed according to the shape of the input analog signal or the bit pattern of the input digital signal.

Modulation Methods

If the modulation signal is an analog signal, the three modulation methods are called,

- Amplitude Modulation [AM]
- Frequency Modulation [FM]
- Phase Modulation [PM]

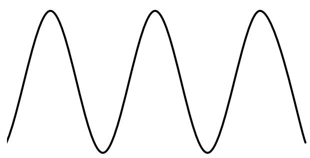
If the modulating signal is a digital signal, the three modulation methods are called,

- Amplitude Shift Keying [ASK]
- Frequency Shift Keying [FSK]

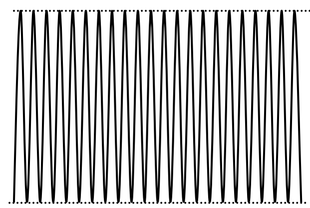
- Phase Shift Keying [PSK]

Analog Modulation

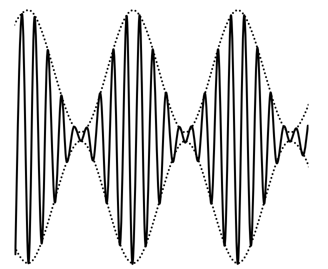
Amplitude Modulation [AM]



Modulating Signal



Carrier Signal



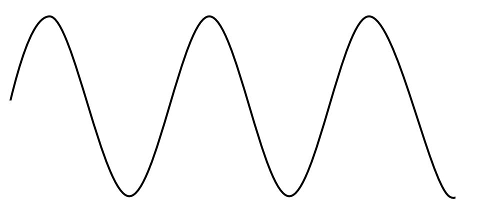
Modulated Signal

Amplitude of carrier signal varies according to the amplitude of modulating signal. The envelop of modulated signal is same as the shape of modulating signal.

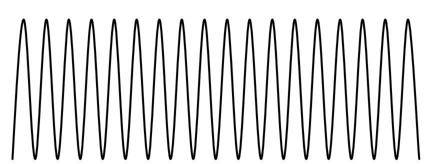
Please note that the frequency or phase of the carrier signal is not changed.

Frequency Modulation

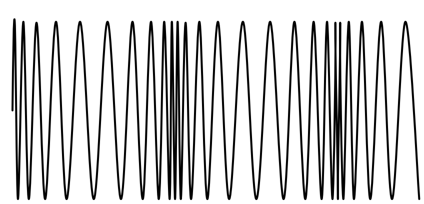
Modulating Signal



Carrier



Modulated Signal



The carrier signal frequency changes according to the amplitude of the modulating signal. When amplitude increases, the modulated carrier signal's frequency increases. If the modulating signal amplitude is negative, the frequency of the modulated carrier signal is decreased.

Please note that the amplitude and phase of the carrier signal is not changed.

Phase Modulation

Same as AM or FM. Instead of Carrier Amplitude or Frequency the carrier phase is changed.

It is not possible to show it pictorially.

Digital Modulation

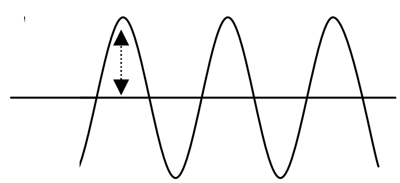
The digital signals are transmitted as 1s and 0s. The characteristic of the carrier signal is changed according to 1 or 0. That means there can be two states of amplitude, frequency or phase. The modulator switches [keying] the carrier to relevant state.

Amplitude Shift Keying [ASK]

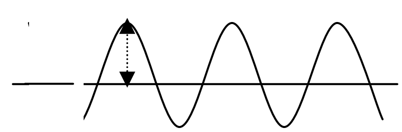
The two states are,

0 – amplitude 1 [a_1]

1 – amplitude 2 [a_2]



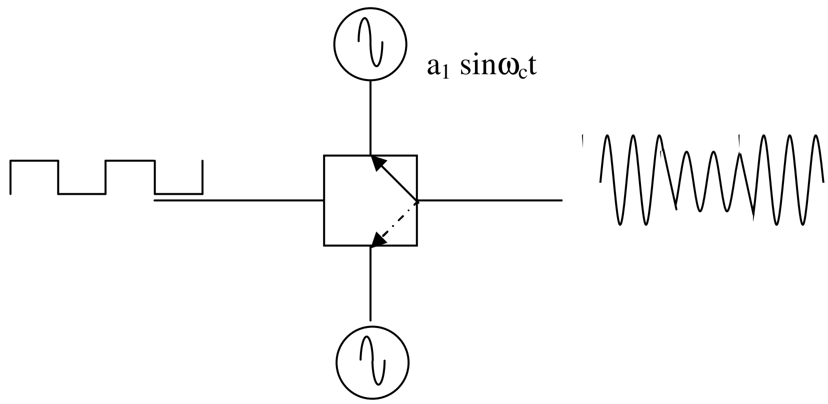
$a_1 \sin\omega_c t$



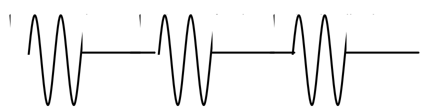
$a_2 \sin\omega_c t$

ω_c - $2\Pi f_C$ f_c – Carrier frequency

Please note that the frequency of both carrier signals are same.



If a_1 = 1V an a_2 = 0V input bit stream is 1 0 1 0 1 0, then the modulated signal pattern will be,



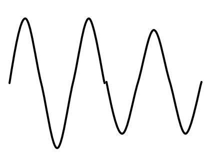
Frequency Shift Keying [FSK]

The two states are,

0 - frequency 1 [f1]

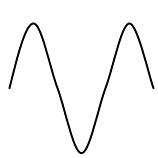
1 - frequency 2 [f2]

f_1

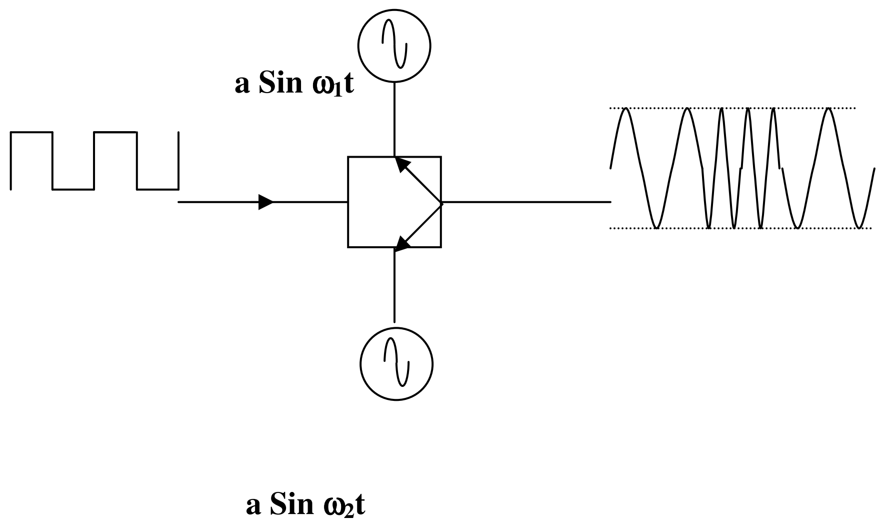


a Sin $\omega_1 t$ $\omega_1 = 2\pi f_1$

f_2



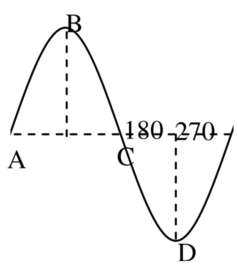
a Sin $\omega_2 t$ $\omega_2 = 2\pi f_2$



Phase Shift Keying [PSK]

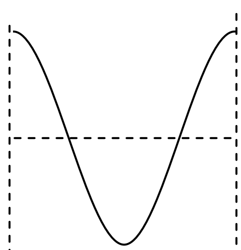
In this method, the carrier signal phase is shifted according to the input digital signal.

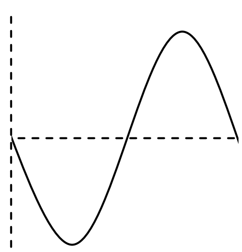
Let us first understand the phase of a signal.



The phase difference between A and B is 90^0.

In other words the point B is 90^0 phase shifted.

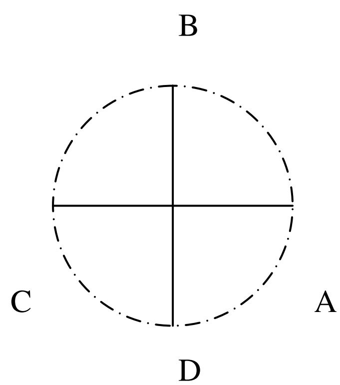




90^0 phase shifted signal.

180^0 phase shifted signal.

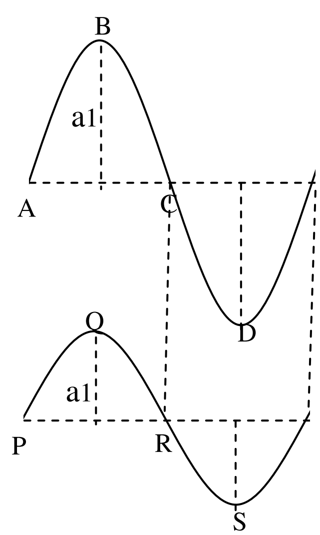
This also can be represented by using a phaser diagram.



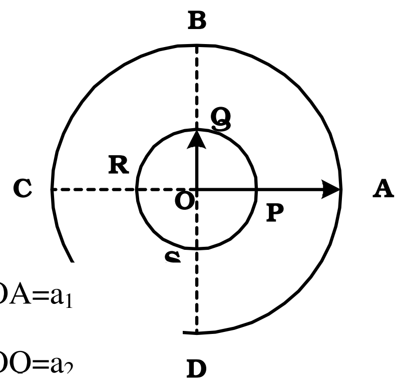
Consider two sinusoidal signals which have the same frequency but different amplitudes.

Signal 1 $a_1 \sin \omega t$

Signal 2 $a_2 \sin \omega t$



The phaser diagram can be drawn as follows.



OA=a_1

OO=a_2

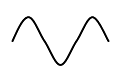
The PSK has different versions.

BPSK, QPSK, 8PSK, 16PSK etc.

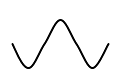
Bipolar Phase Shift Keying [BPSK]

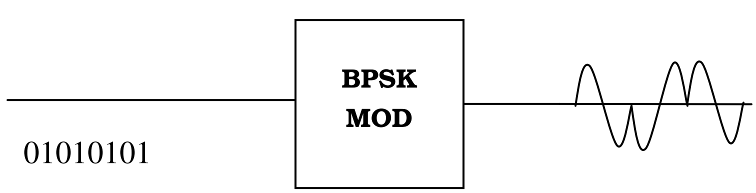
There are only two phases.

0 - no phase shift.



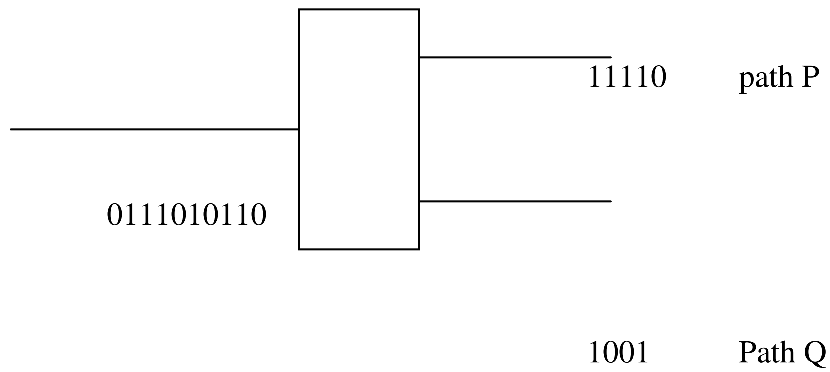
1 - 180^0 phase shift





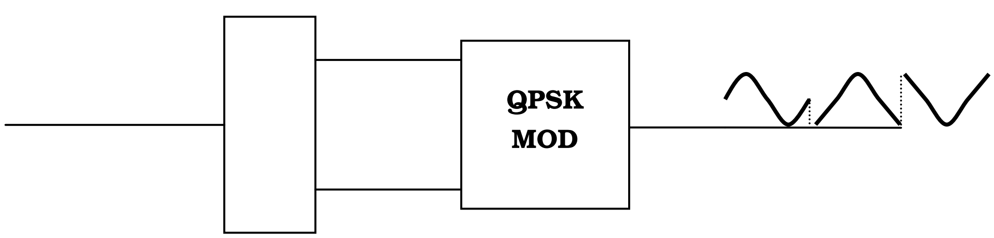
Quadrature Phase Shift Keying [QPSK or 4PSK]

In this method, first the input data stream is divided into two parallel streams.



Divider

First bit goes to P, second bit goes to Q, third bit goes to P, forth bit goes to Q and so on.



Divider

At the input of the QPSK Modulator, four types of bit combinations can be expected.

That is 00

01

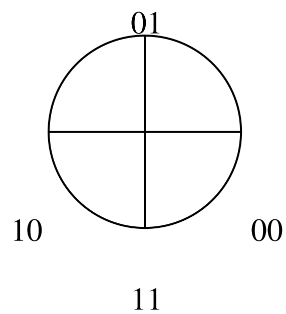
10

11

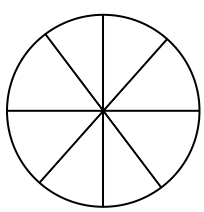
These bit combinations will have four different phases.

$00 - 0^0$, $01 - 90^0$, $10 - 180^0$, $11 - 270^0$

Phaser diagram



Similarly the 8PSK phaser diagram can be represented as follows.



Bit Combination	Phase Shift [Degrees]
000	0
001	45
010	90

Hybrid Modulation

This is a combination of ASK and PSK.

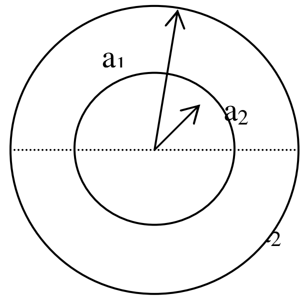
This method of modulation is called Amplitude Phase Shift Keying [APSK] or Quadrature Amplitude Modulation [QAM].

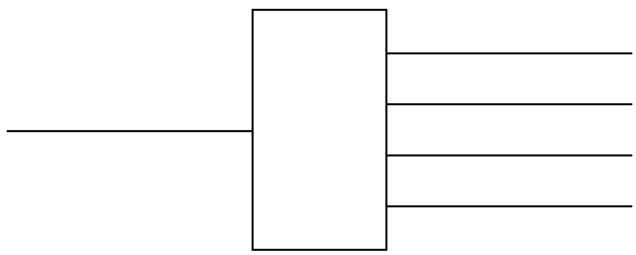
It can be 16 QAM, 64 QAM etc.

In this method two carrier signals with different amplitudes are involved.

a_1 Sin ωt $\qquad$ a_2 Sin ωt

The phaser diagram can be drawn as follows.



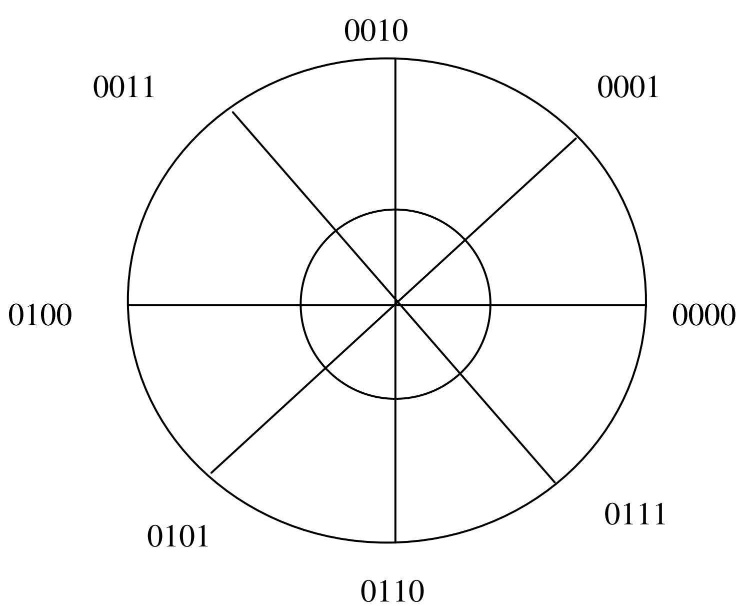


Divider

In this example, there are 16 combinations.

0000, 0001, 0010 …………….. 1111

The 16 locations of the phaser diagram is as follows.

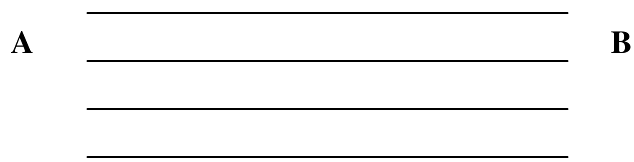


The inner circle corresponding locations represent 1000, 1001, 1010……..1111.

If the circle is divided into 16, 32 QAM can be represented.

If the circle is divided into 32, 64 QAM can be represented.

Multiplexing



Suppose we need to transmit four 64 kb/s signals from A to B. For this purpose, it is required to have four channels. Each channel needs at least 2 wires. If the length from A to B is 100m, we need 4 X 2 X 100 = 800m Copper Cable. If the length is 1000m the required length increases to 8000m.

If we can combine all four channels together without any mixing, a single pair of cable is sufficient. This type of combination (packing) of signal is called Multiplexing.

There are mainly two types of Multiplexing.

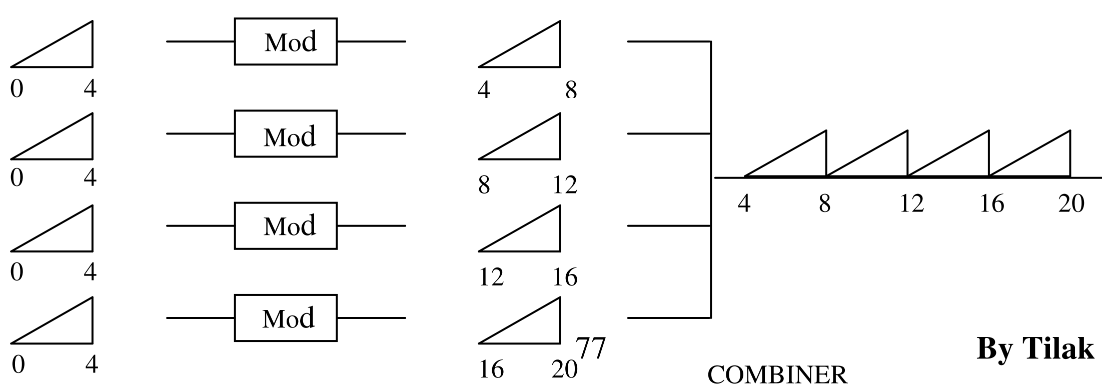
- Frequency Division Multiplexing [for Analog Signals]
- Time Division Multiplexing [for Digital Signals]

Frequency Division Multiplexing

Let us consider multiplexing of telephone channels.

One Channel - 0 – 4 kHz. [actually it is 0.3 – 3.4 kHz].

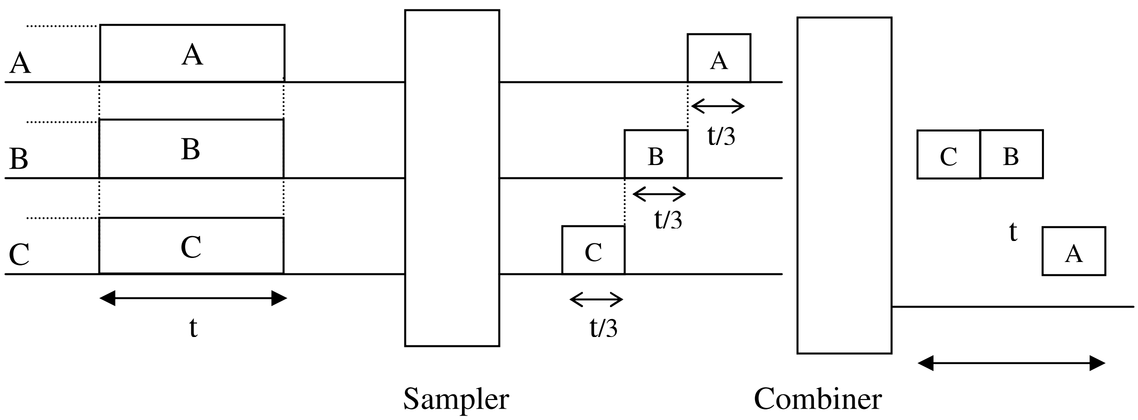
The frequency band can be shifted by modulation.



Here it can be seen that there is no interference of channels. This process is called Frequency Division Multiplexing.

Time Division Multiplexing [TDM]

Suppose we want to multiplex three Digital Signals, which have the same bit rate.



This process is called Time Division Multiplexing.

Suppose the input bit rate is n bits/sec

Time duration is t

t second $\longrightarrow$ 1 bit

1 second $\longrightarrow \frac{1}{t}$ bits $= n$ bits/sec

At the output

t second $\longrightarrow$ 3 bits

1 second $\longrightarrow \frac{3}{t}$ bits

$= 3 \text{ X } \frac{1}{t}$ bits

$=$ 3 n bits/sec.

It can be seen that in the TDM process, the output bit rate is increased.

Note : If A, B, C are single bits, the TDM method is called “bit interleaving”.

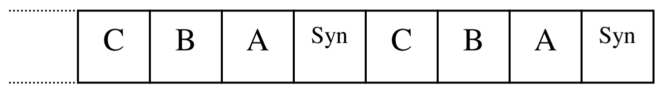
If A, B, C are each 8 bits, the TDM method is called “word interleaving”.

8 bits are also called a Byte or a Time Slot [TS].

TDM Systems

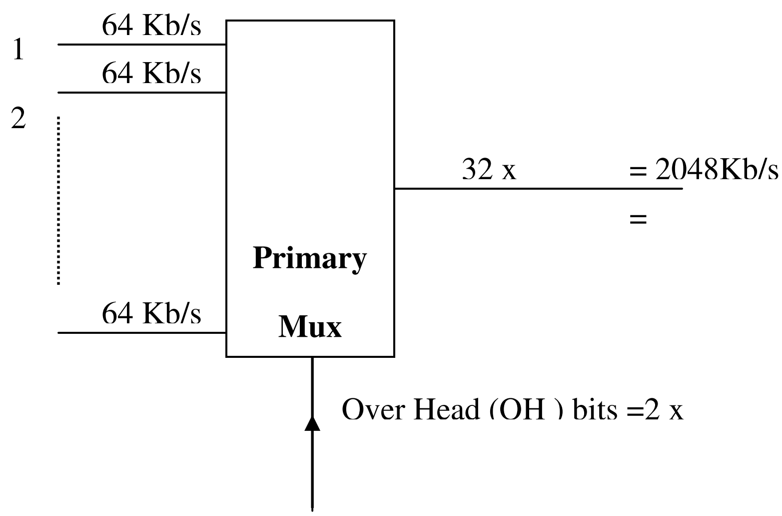
In actual systems, in addition to channel data, additional data is added. They are called the Over Head Bits. [OH bits]

E.g. Synchronization bits

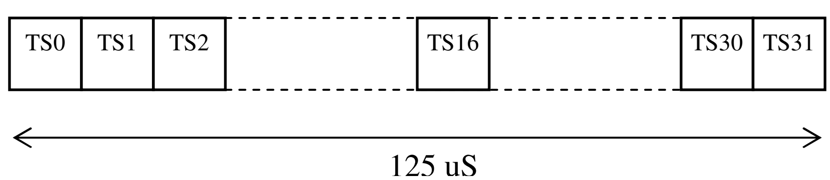


Primary Mux [E1 Channel]

By multiplexing 30 channels [each channel is 64 kb/s] the primary mux output is formed.



The frame structure of output signal is given in the figure.



Time Slot 0 [TS0] and Time Slot 16 [TS16] are overhead bits.

One Time Slot = 8 bits

Therefore, 1 frame = 8 X 32 bits

= 256 bits.

There is another Primary Mux which will multiplex 24 channels, and its output bit rate is 1.544 Mb/s. This is called a T1 channel.

Note : In Sri Lanka E1 multiplexing is used.

One TS carries data of one channel.

One channel is 64 kb/s.

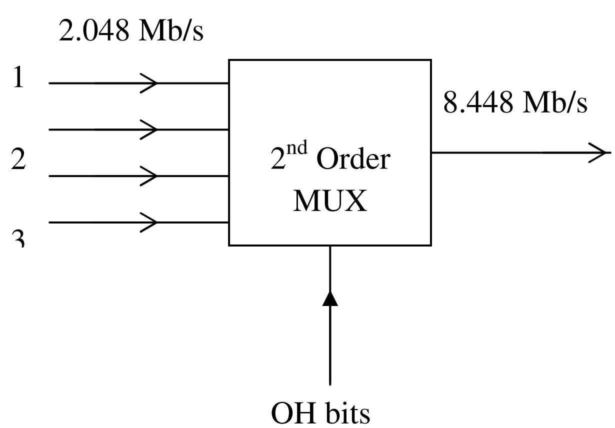
Therefore, one TS is = 64 kb/s.

If you need a 64 kb/s data channel, the data circuit provider allocates you one Time Slot. If you need 128 kb/s data circuit, two Time Slots are allocated. Similarly for 512 kb/s data channel, 8 Time Slots are allocated. If you need a 2.048 Mb/s data channel the whole E1 is allocated.

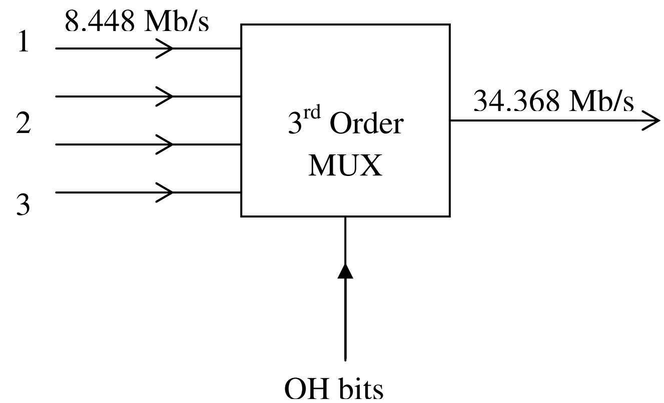
Higher Order Muxes

The primary mux is also called a 1st order mux. Four primary mux output can be again multiplexed and a 2nd order mux output is made.

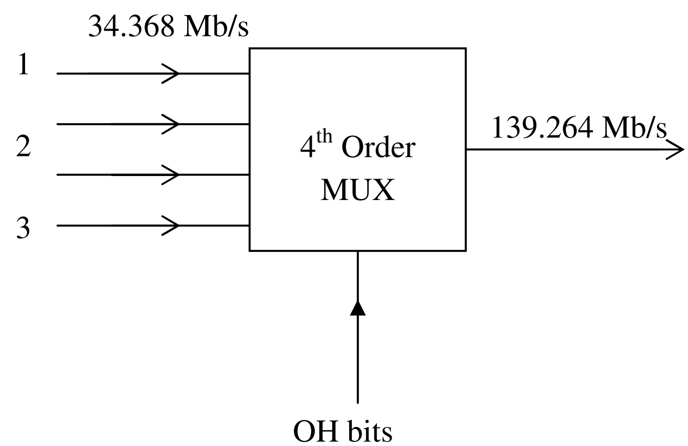
2nd Order Mux [E2 Channel]



3rd Order Mux [E3 Channel]

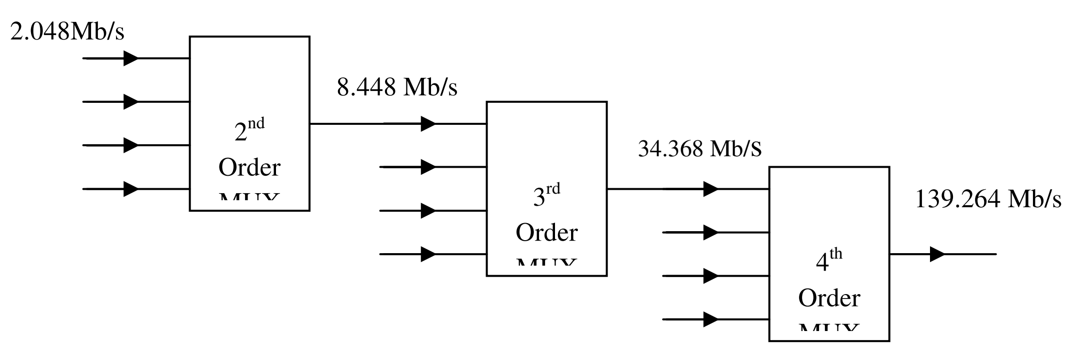


4th Order Mux [E4 Channel]



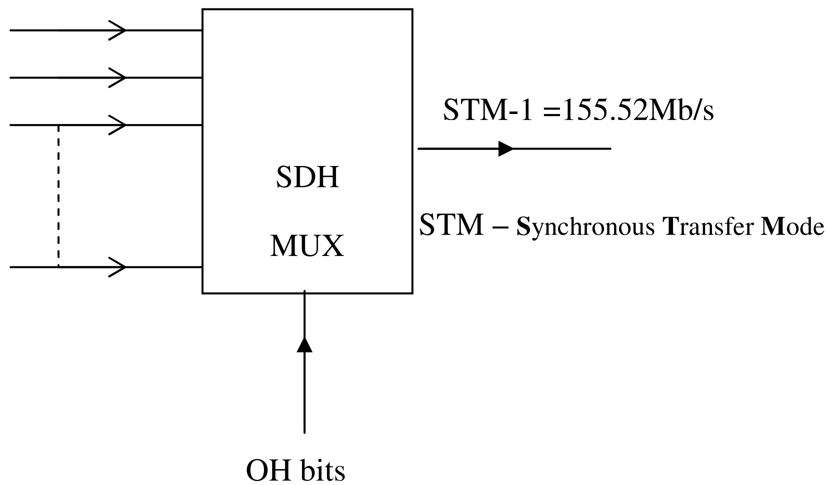
Plesiochronous Digital Hierarchy [PDH]

This is one of the digital multiplexing hierarchies.



Synchronous Digital Hierarchy [SDH]

This is the modern digital multiplexing hierarchy.



The input can be E1 or E2 or E3 or E4.

The inputs can be configured.

The output bit rate is 155.52 Mb/s.

4 X STM - 1 = STM - 4

4 X STM - 4 = STM - 16

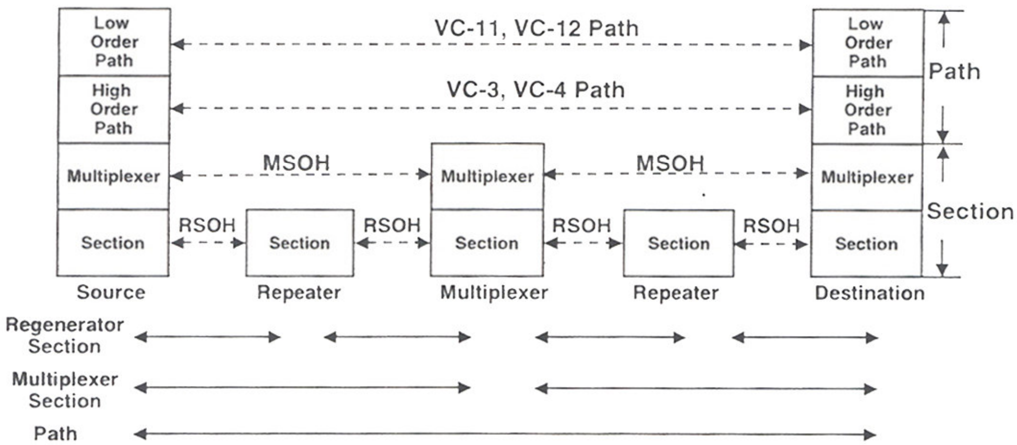
4 X STM - 16 = STM - 64

SDH/SONET Optical Transport Networks

SDH/SONET Multiplexing Structures

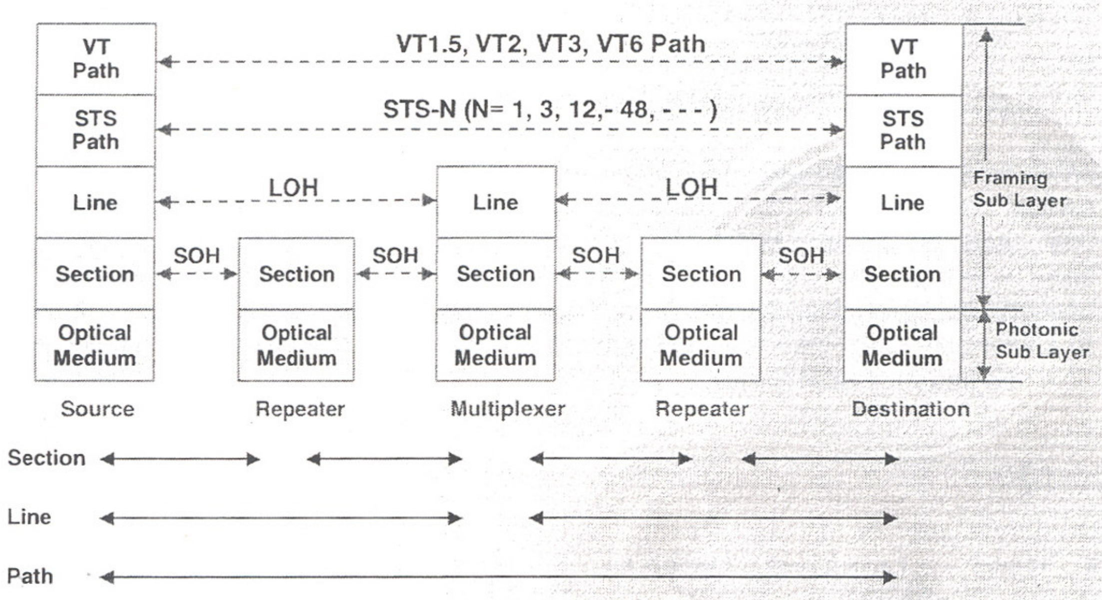
Frame Structure | Multiplexing Structure | Payload Multiplexing | FLX Series

SDH Layered Architecture



The SDH Physical Layer is divided into four sub-layers : Path (LO/HO), multiplexer and section for transporting the payload across the network within source and destination via one or more repeaters and multiplexers.

SONET Layered Architecture

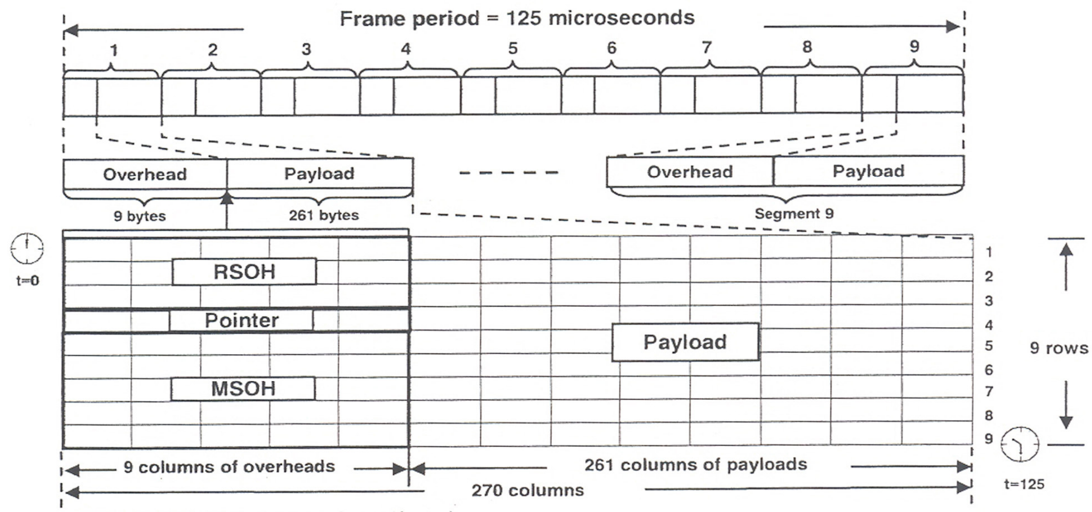


SDH/SONET Optical Transport Networks

SDH/SONET Multiplexing Structures

Frame Structure	Multiplexing Structure	Payload Multiplexing	FLX Series

SDH Frame Structure (STM-1)



STM-1 Frame Capacity (9x270=2430bytes):
- Section Overhead = 72 bytes = 8x9 (MSOH/RSOH)
- Administrative Unit Pointer = 6 bytes (Pointer)
- Negative Justification = 3 bytes (Pointer)
- Payload Without Positive Justification = 2,349 bytes = 9x261 (PAYLOAD)

Notes:

SDH Frame Structure-STM-1

SDH multiplexing combines low-speed digital signals such as 2, 34, and 140 Mb/s signals with required overhead to form a frame called Synchronous Transport Module at level one (STM-1). The Figure shows the STM-1 frame, which is created by 9 segments of 270 bytes each. The first 9 bytes of each segment carry overhead information; the remaining 261 bytes carry payload. When visualized as a block, the STM-1 frame appears as 9 rows by 270 columns of bytes. The STM-1 frame is transmitted row #1 first, with the most significant bit (MSB) of each byte transmitted first.

This formula calculates the bit rate of a framed digital signal:

Bit rate = frame rate x frame capacity

In order for SDH to easily integrate existing digital services into its hierarchy, it operates at the basic rate of 8 kHz, or 125 microseconds per frame, so the frame rate is 8,000 frames per second. The frame capacity of a signal is the number of bits contained within a single frame.

<u>Frame capacity calculations</u>:

Frame capacity = 270 bytes/row x 9 rows/frame x 8 bits/byte = 19,440 bits/frame

The bit rate of the STM-1 signal is calculated as follows:

Bit rate = 8,000 frames/second x 19,440 bits/frame = 155.52 Mb/s

<u>Section Overhead</u>

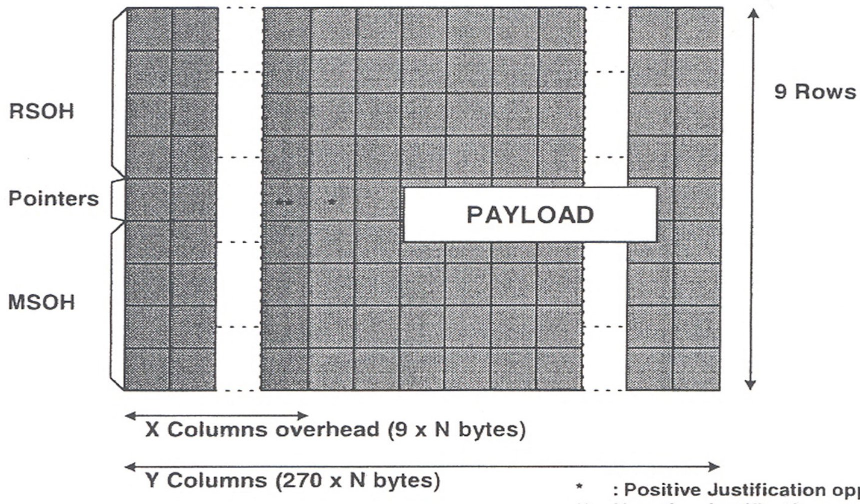
Some signal capacity is allocated in each transport frame for the “Section Overhead”. This provides the facilities(such as alarm monitoring, bit error monitoring and data communications) required to support and maintain the transportation of a VC between nodes in a synchronous network.

SDH/SONET Optical Transport Networks

SDH/SONET Multiplexing Structures

Frame Structure | Multiplexing Structure | Payload Multiplexing | FLX Series

STM-N Frame Structure



* : Positive Justification opportunity timeslots
** : Negative Justification opportunity timeslots

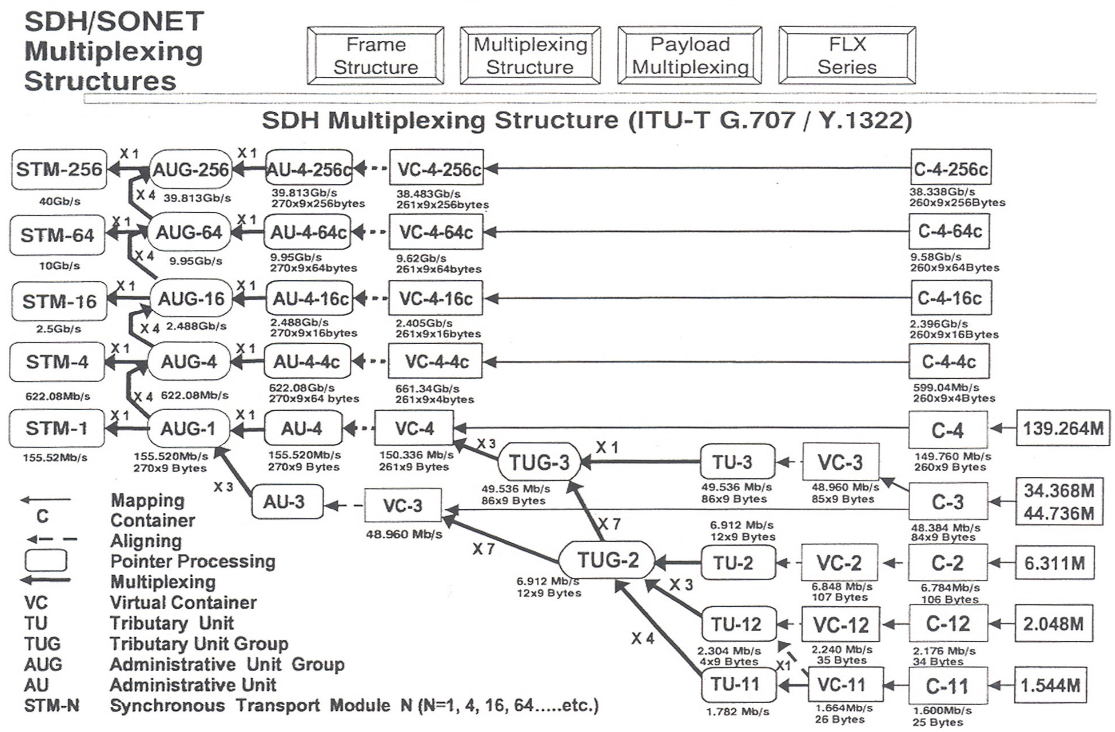
STM-4 is derived by byte by byte multiplexing of 4 x STM-1. STM-16 is derived by multiplexing 4 x STM-4, STM-64 is derived by multiplexing 4 x STM-16, and STM-256 is derived by multiplexing 4xSTM-64..
The duration of each frame is 125 μsec.

SDH/SONET Multiplexing Rates

SDH	SONET		Transmission data rate (Mb/s)				Payload data rate		User data rate (Mb/s)	Overhead Data rate (Mb/s)
Electrical/ optical	Electrical	optical	X.col.	Y.col.	Gross (Mb/s)	Row x Column (Total Bytes)	SPE (Mb/s)	Row x Column (Total Bytes)		
* STM-0	STS-1	OC-1	3	90	51.840	9 x 90=810	50.112	9 x 87=783	49.536	1.728
* STM-1	STS-3	OC-3	9	270	155.520	9 x 270=2430	150.336	9 x 261=2349	148.608	5.184
STM-3	STS-9	OC-9	27	810	466.560	9 x 810=7290	451.008	9 x 783=7047	445.824	15.552
* STM-4	STS-12	OC-12	36	1080	622.080	9 x 1080=9720	601.344	9 x 1044=9396	594.432	20.736
STM-6	STS-18	OC-18	54	1620	933.120	9 x 1620=14580	902.016	9 x 1566=14094	891.648	31.104
STM-8	STS-24	OC-24	72	2160	1244.160	9 x 2160=19440	1202.688	9 x 2088=18792	1188.864	41.472
STM-12	STS-36	OC-36	108	3240	1866.240	9 x 3240=29160	1804.032	9 x 3132=28188	1783.296	62.208
* STM-16	STS-48	OC-48	144	4320	2488.320	9 x 4320=38880	2405.376	9 x 4176=37584	2377.728	82.944
* STM-64	STS-192	OC-192	576	17280	9953.280	9 x 17280=155520	9621.504	9 x 16704=150366	9510.912	331.776
STM-256	STS-768	OC-768	2304	69120	39813.12	9 x 69120=622080	38486.016	9 x 66816=601344	38043.648	1327.104

SDH : Synchronous Digital Hierarchy. ITU -T G.707 Rec.
SONET : Synchronous Optical Network. USA Bellcore Standard.
STM-N : Synchronous Transport Module Level- N.
OC-N : Optical Carrier Level N.
STS-N : Synchronous Transport Signal Level- N.
SPE : Synchronous Payload Envelope
* : Typical SDH Systems.

SDH/SONET Optical Transport Networks



Notes:

Mapping and Multiplexing Structure (ITU-T G.707 / Y.1322)

For the purpose of transferring information, all digital signals of conventional PDH Hierarchy shown in the far right column in this diagram can be multiplexed as follows:

- Each container (ex. **C-12**) becomes a Virtual container (ex. **VC-12**) by the addition of Path Overhead bits (POH) and some stuffed bits in predetermined position. This procedure is called **Mapping**. Mapping of a non-SDH signal means increasing the frequency of the non-SDH signal to a pre-determined frequency and adding overhead for each one of the non-SDH signals
- The procedure referred to as **Aligning** relates to the process of assembling a virtual container into a tributary unit (ex. **TU-12**) in which a pointer is added to indicate the location of the first byte of the virtual container inside the tributary unit frame.
- Tributary unit group (ex. **TUG-2** and **TUG-3**) is made up by **Multiplexing** a number of Tributary units in byte interleave basis. The same procedure is applied to obtain Virtual container of higher level such as the Virtual Container 4 or VC-4.
- A virtual container-4 (VC-4) is assembled into Administrative unit-4 (**AU-4**) by adding AU-4 Pointer and becomes STM-1 by adding Section Overhead bytes (SOH).
- Another multiplexing process via VC-3 and AU-3 is mainly used in North America and Japan.
- Higher level signals of STM-N (STM-4, STM-16, STM-64 and STM-256) are configured by multiplexing STM-1 signal in byte interleave basis The bit rate of STM-N is equal to N x 155.52Mb/s where N is equal to 1, 4, 16, 64 and 256.

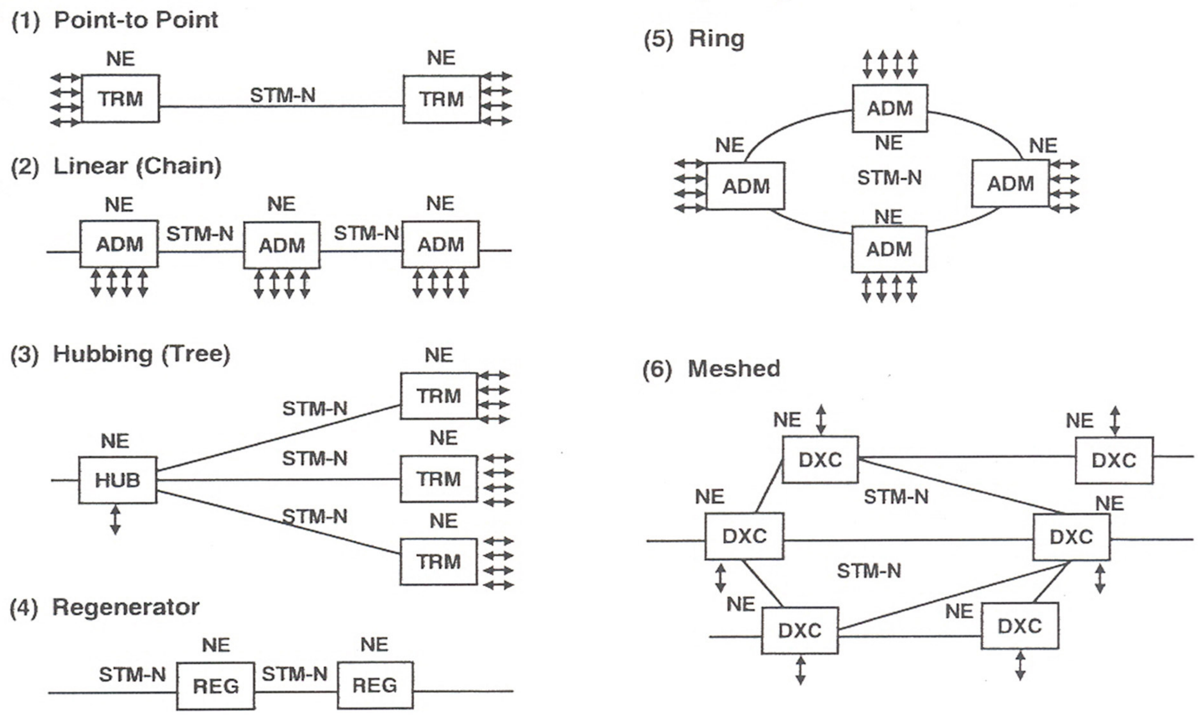
Note: Unshaded portion (VC-2, TU-2, TU-11, and VC-3) do not apply to Fujitsu SDH FLX Series

SDH/SONET Optical Transport Networks

SDH Network Configuration

Network Topologies | Network Protection | **Network Synchronization** | Network Management

SDH Network Topologies

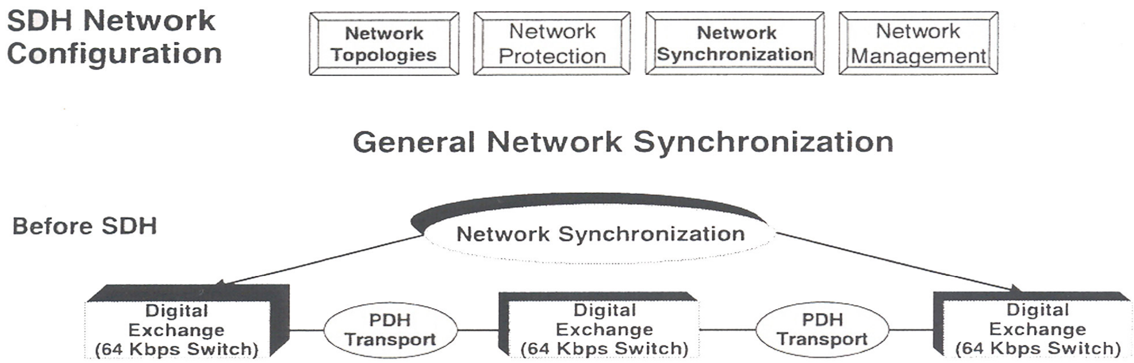


Notes:

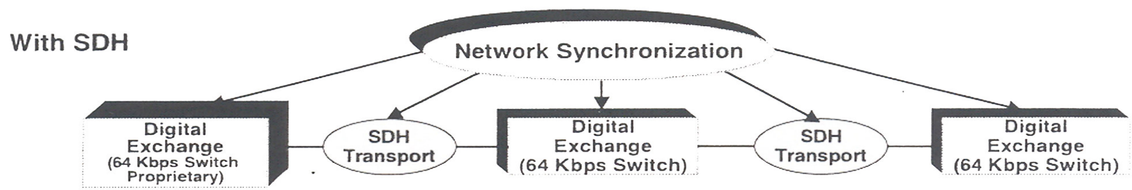
SDH Network Topologies

Besides the conventional network structure like point-to-point or simple star or tree type connection, various useful structures like simple linear (chain), ring and mesh network or combination of them is available utilizing ADM and crossconnect function of SDH node. In particular ring and mesh network has excellent survivability for line failure and are widely used in modern telecommunication network.

SDH/SONET Optical Transport Networks



- Switch & Cross Connect synchronization required.
- PDH network did not require synchronization.



- SDH Require Synchronization to Minimize Jitter and Wander.

Notes:

SDH Network Synchronization

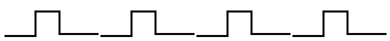
In SDH network all SDH equipment should be synchronized to network common clock as well as to interfaced digital switching system. Usually mater-slave type of clock distribution network (synchronization network) is employed. The master station with atomic oscillator (PRC) and main slave station (Transit SSU) have a duplicate configuration. Each network element of SDH system has a function to select the operating clock from several clock sources.

General Network Synchronization

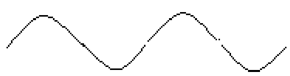
A network used to provide reference timing signals for SDH. As shown in the diagram, before SDH, the digital switching exchanges are supplied with network clock. This especially true for those exchanges operating in the toll switching centers. If a small quantity of SDH is deployed it may be possible to use existing synchronization facilities provided for the 64kbit/s switching network to synchronize the SDH network. This is especially applicable if small islands of SDH are deployed. When the interconnected SDH networks become larger, with greater than 20 Network elements per island, it will be necessary to ensure that synchronization of the SDH network is thoroughly planned.

Digital data detection

Transmits

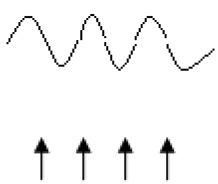


Receive

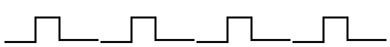


The signal is distorted due to noise.
How to detect the received binary signal

(1).identify the bit position.



Clock pulses are used for their purpose if is a very low width pulse travel.



To identify the bit position the bit rate and the clock frequencies should be effect.

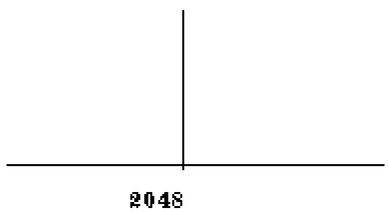
E.g.:

If the bit rate is 2.048 mb/s the clock pulse rate also should be 2.048mbps.in order to generate this rate of clock pulses clock frequency is required.
Piczo electric crystal.
This crystal should have the required frequency.
Normally the clock frequency of the crystal is slightly varying.
The clock frequency transmits also slightly varying.

The clock frequency of transmit end also slightly varying. Therefore the received bit rate is not fixed. It is slightly varying. There is a acceptance tolerance of bit rate.
This is measured in parts per million. (Ppm).

E.g.:

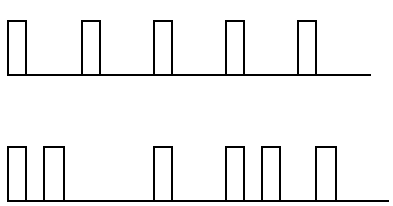
2.048 mbps has a tolerance of + or – 80 ppm.
The bit rate is $2.048*10^6$ b/s.
10^6 – Million
There 2.048 million bits.
Therefore the tolerance is = +or – 2.048 * 50 b/s.
= + or – 102.4 b/s.
The minimum acceptance bit rate is $2.048*10^6$ - 102.4 b/s.
The maximum acceptance bit rate is $2.048*10^6$ + 102.4 b/s.



Min Max

The actual bit rate can be varying between minimum and maximum.
The tolerances defined by ITU – T are
2.048 mb/s = + or – 50ppm
8 mb/s = + or – 30 ppm
34 mb/s = + or – 20 ppm
140 mb/s = + or – 15 ppm

Jitter

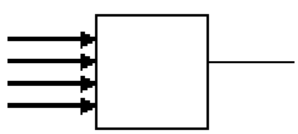


Ideal bit positions of a bit stream the bits are equal space.
Actual bit position the bits are slightly shifted due to different reasons.
This effect is called jitter.
It can be shown the jitter causes error.

Because of the above problems the bit position of incoming bit stream cannot be identified by a clock pulses of receiving end. In order to synchronize the clock pulses with incoming bit stream the clock pulses should be generated from the incoming bit stream. This is called clock recovery .normally a phase local loop circuit is used for clock recovery.

Phase lock loop circuits.
I – incoming signal
VCO- voltage controlled oscillator
Data detection
Plesiochronous digital hierarchy.
In this multiplexing hierarchy the same category of bit rates are not synchronizes.
There is a slightly variation of bit streams within the given tolerance.

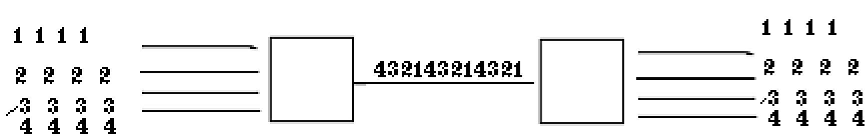
E.g.:
2.048mbps – tolerance + or – 102.4b/s.
They are called nearly synchronized or plesiochronous.



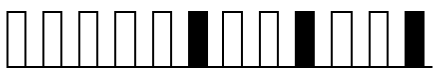
Low rate high rate

The low rate bit streams are called tribulatoriy.
The high rate bit stream is called MUX.
In European hierarchy all n^{th} order MUXes have 4 tributaries.
There are PDH hierarchy uses in the world.
European (E)
North America (T)
Japanese (J)
All systems the basic channel bit rate is 64kbps.
European
Primary Mux – joch: 1ch: E1
Primary * 4 = 2^{nd} order – 8.448mb/s – E2
2^{nd} order*4 = 3^{rd} order=

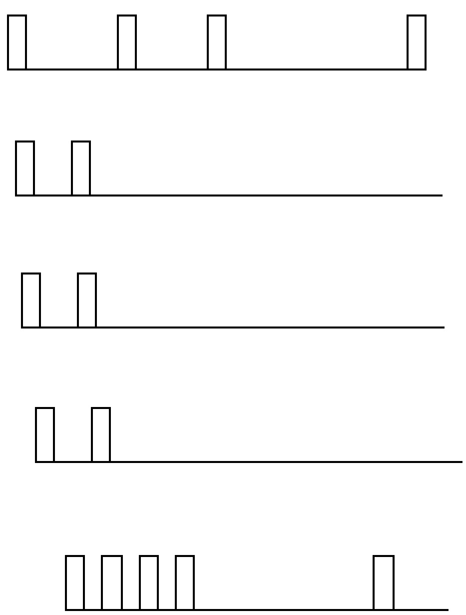
Higher order MUX



Multiplexers
E.g.:
2^{nd} order
1,2,3,4 – Bits
Bit interleaving
Justification bits

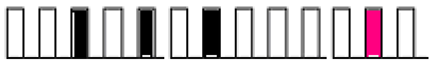


Since the tributary bit rates can slightly varying the MUX will keep room for the higher possible bit rate.
There are specified locations in the mux output frame structure for some special bits. they are called justification bits. If the input bit rate is maximum allowed bit rate all justification bits will be data bits. if the input bit rate is less than the highest possible bit rate , some of the justification bits are filled with data bits and others are filled with dummy bits. at the receive end the data bits are putting tributary after demultiplexing) and dummy bits are ignored.
In order to identify the special bit, that is whether it is a data bit or dummy bit. Another bit type is called justification control bits (JC) are introduces.
When multiplexing, firstly 4 sub frames are made from 4 tributaries. After that 4 sub frames are multiplexed by bit interleaving.
The bit order is as bellow.



123412341234

1-bit from sub frame 1
2- Bit from sub frame 2
3- Bit from sub frame 3
4-bit from sub frame 4



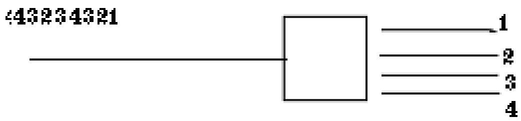
One sub frame has normally 1 JC bit and 1 Bit.

JC = 1   J = useful bit

JC = 0   J = dummy bit or vise versa

If one JC bit used, if it has an error the received Mux all tributaries get wrong data.

If J = useful bit.



Suppose only 1 JC bit per frame is used. If it gives error. The J bit is identified incorrectly. then from above example it can be shown that al tributary output will from errors after J bit.
JC and dummy J bits are overhead bits. in addition to the synchronous bit and alarm bits are included to the multiplexed frame.
Since there are 3 JC bits and the correct value of JC is considered from the majority JC value.
E.g.:

JC = 0, JC = 1, JC = 1 will consider as JC = 1

This will avoid the error situation, If JC bit subjected to error.

Introduction to FDM, OFDM, OFDMA, SOFDMA

Frequency Division Multiplexing (FDM)

WiMAX air interface is based on OFDM/OFDMA physical layer (PHY). To understand how OFDM and OFDMA work, it is useful to start with its "mother" namely FDM (Frequency Division Multiplexing).

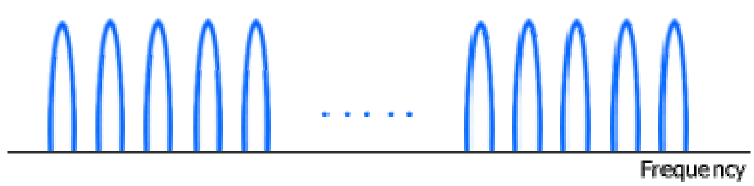


Figure. Frequency Division Multiplexing (FDM)
Spacing is put between two adjacent sub-carriers.

In FDM system, signals from multiple transmitters are transmitted simultaneously (at the same time slot) over multiple frequencies. Each frequency range (sub-carrier) is modulated separately by different data stream and a spacing (guard band) is placed between sub-carriers to avoid signal overlap.

Orthogonal Frequency Division Multiplexing (OFDM)

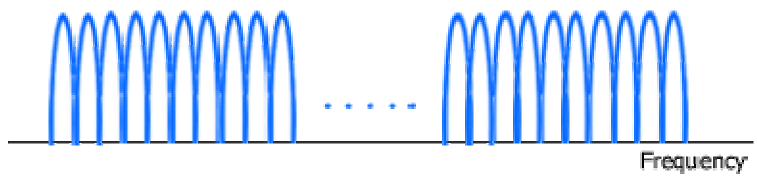


Figure. Orthogonal Frequency Division Multiplexing (OFDM)
Sub-carriers are closely spaced until overlap.
Like FDM, OFDM also uses multiple sub-carriers but the sub-carriers are closely spaced to each other without causing interference, removing guard bands between adjacent sub-carriers. This is possible because the frequencies (sub-carriers) are orthogonal, meaning the peak of one sub-carrier coincides with the null of an adjacent sub-carrier.
In an OFDM system, a very high rate data stream is divided into multiple parallel low rate data streams. Each smaller data stream is then mapped to individual data sub-carrier and modulated using some sorts of PSK (Phase Shift Keying) or QAM (Quadrature Amplitude Modulation). i.e. BPSK, QPSK, 16-QAM, 64-QAM.
OFDM needs less bandwidth than FDM to carry the same amount of information which translates to higher spectral efficiency. Besides a high spectral efficiency, an OFDM system such as WiMAX is more resilient in NLOS environment. It can efficiently overcome interference and frequency-selective fading caused by multipath because equalizing is done on a subset of sub-carriers instead of a single broader carrier. The effect of ISI (Inter Symbol Interference) is suppressed by virtue of a longer symbol period of the parallel OFDM sub-carriers than a single carrier system and the use of a cyclic prefix (CP).
OFDM has been in theory for decades but just entered real world applications in recent years thanks to the availability of modern chips that can handle complex digital signal processing. Wireline and wireless, fixed and mobile communications or networking technologies have chosen OFDM to achieve higher data rate (what is called broadband). Examples of such technologies are: ADSL, HomePlug AV, WiMedia UWB, Wi-Fi (802.11a/g), WiMAX.

Orthogonal Frequency Division Multiple Access (OFDMA)

Like OFDM, OFDMA employs multiple closely spaced sub-carriers, but the sub-carriers are divided into groups of sub-carriers. Each group is named a sub-channel. The sub-carriers that form a sub-channel need not be adjacent. In the downlink, a sub-channel may be intended for different receivers. In the uplink, a transmitter may be assigned one or more sub-channels.

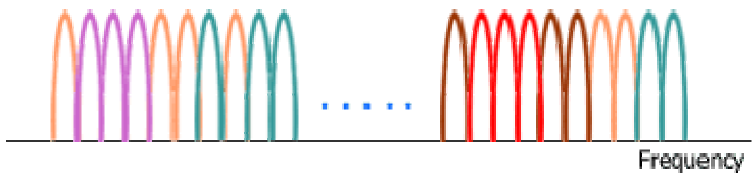


Figure. Orthogonal Frequency Division Multiple Access
Sub-carriers with the same color represent a sub-channel.

Scalable OFDMA (SOFDMA)

SOFDMA (S-OFDMA) adds scalability to OFDMA. It scales the FFT size to the channel bandwidth while keeping the sub-carrier frequency spacing constant across different channel bandwidths. Smaller FFT size is given to lower bandwidth channels, while larger FFT size to wider channels. By making the sub-carrier frequency spacing constant, SOFDMA reduces system complexity of smaller channels and improves performance of wider channels.
As a reminder, IFFT (Inverse Fast Fourier Transform) is used in a WiMAX transmitter to create an OFDM waveform from modulated data streams, while FFT (Fast Fourier Transform) is used in a WiMAX receiver to demodulate the data streams. The FFT size equals the number of sub-carriers, e.g. in a OFDM/OFDMA system with 256 sub-carriers, the FFT size is 256.

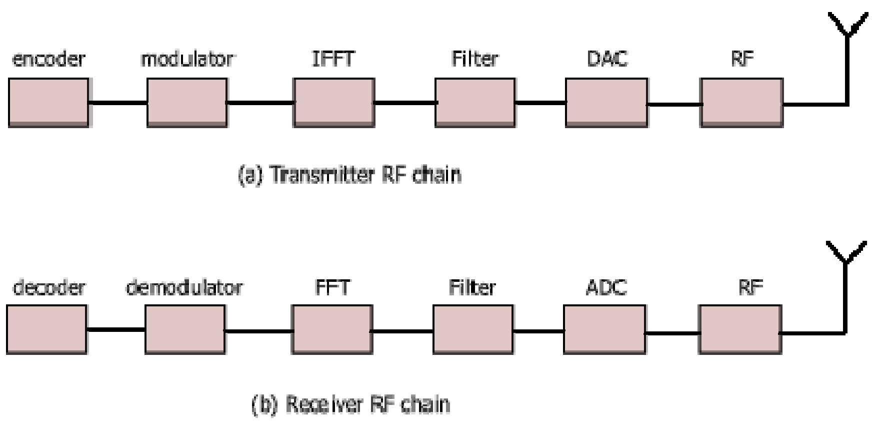


Figure. Transmitter and Receiver RF chains in WiMAX
Basic building blocks of a Tx RF chain and a Rx RF chain.

SOFDMA is the OFDMA mode used in Mobile WiMAX. It supports channel bandwidths ranging from 1.25 MHz to 20 MHz. With bandwidth scalability, Mobile WiMAX technology can comply with various frequency regulations worldwide and flexibly address diverse operator or ISP requirements, that's whether for providing only basic Internet service or a broadband service bundle.

OFDM and OFDMA Symbol Structure

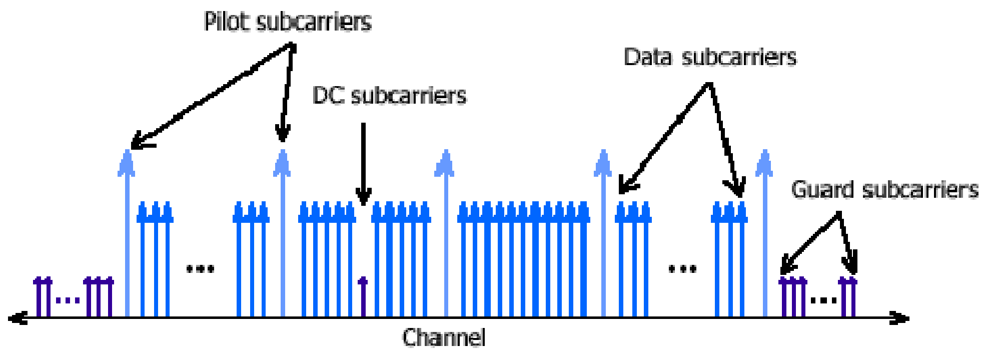


Figure. OFDM symbol structure in WiMAX

Both OFDM and OFDMA symbols are structured in similar way. Each symbol consists of:
- data sub-carriers (OFDM) or sub-channels (OFDMA) that carry data (information),
- pilot sub-carriers as reference frequencies and for various estimation purposes,
- DC sub-carrier as the center frequency, and
- guard sub-carriers or guard bands for keeping the space between OFDM/OFDMA signals.

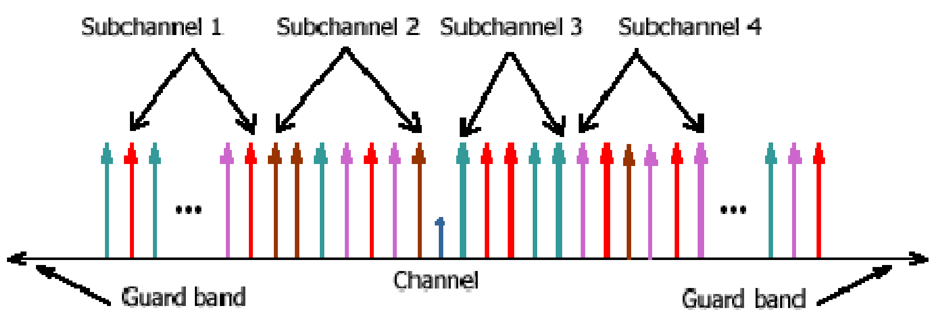


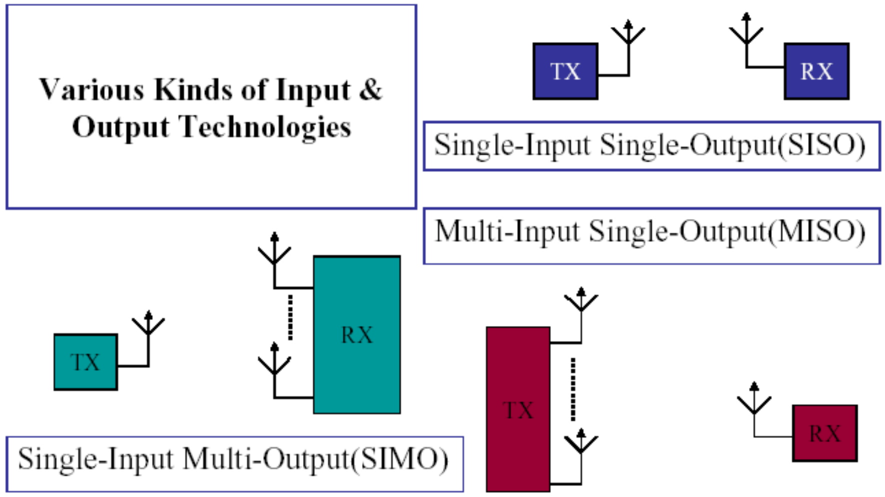
Figure. OFDMA symbol structure in WiMAX sub-carriers of the same color represent a sub-channel.

Multiple Antennas

Smart antenna and self-tracking antenna are several forms of intelligent antenna which improve the performance of radio systems by either the use of intelligent antenna techniques or employing intelligence in the spatial processing.
Issues of the Multiple Antenna Research
The following items list the issues of the multiple antenna research aims to improve the performance of radio communications.

- Intelligent antenna
- Smart antenna
- Multiple-input multiple-output (MIMO)
- Beamforming
- Diversity combining
- Diversity scheme
- Space–time code
- Spatial multiplexing
- Space-division multiple access (SDMA)
- Advanced MIMO communications

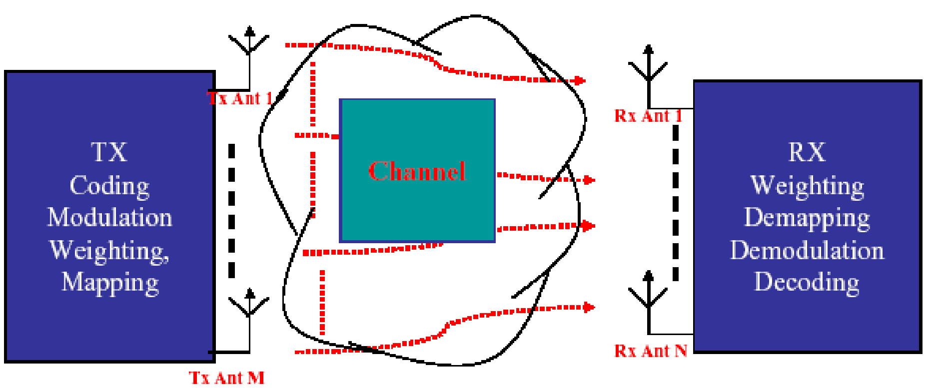
Non-MIMO Systems



What is MIMO?

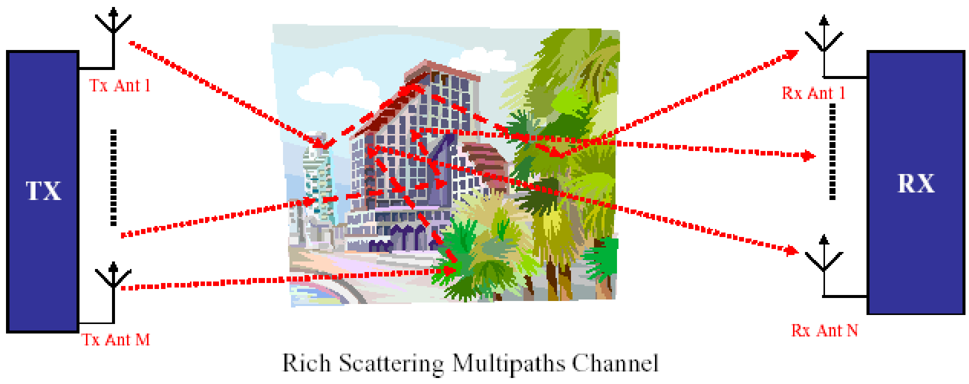
RX

- Multiple antennas at both ends of a wireless link promises significant improvements in terms of spectral efficiency and link reliability.
- Known as multiple-input multiple-output (MIMO) technology.



Rich Scattering Multipaths Channel

Generic MIMO System

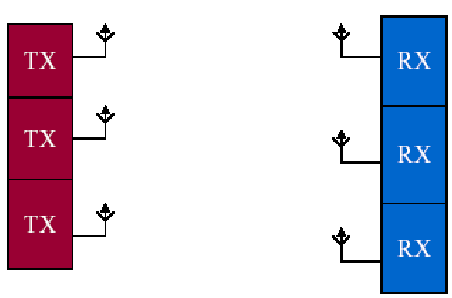


Advantages of Multiple Antennas

- **Array gain**: Signal to thermal noise ratio is improved. Increased coverage.
- **Diversity gain**: Mitigates fading through spatial diversity. Improved quality.
- **Spatial multiplexing**: Increased spectral efficiency.
- **Note:** Diversity gain and spatial multiplexing may be mutually conflicting goals.

Array gain

- Increase in average received SNR obtained by coherently combing the incoming/ outgoing signals
- H is required at transmitter/receiver.
- Increase Coverage

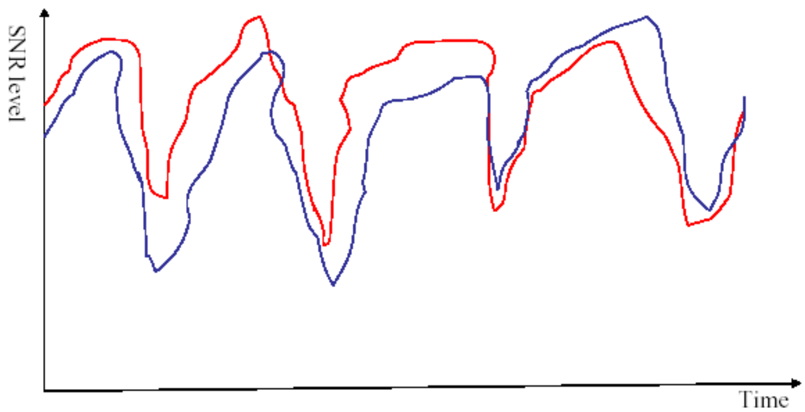


Diversity Gain

- To combat small scale fading caused by multi-path effects.
- The basic principle of diversity is that if several replicas of the information signal are received through independently fading links (branches) , then with high probability at least one or more of these links will not be in a fade links (branches) .

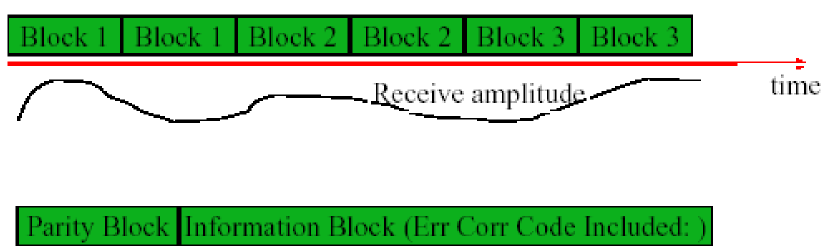
Diversity Gain

• Provide the receiver with multiple identical copies of a given signal to combat fading.
• Replicas of the transmitted waveform are provided across time



Time Diversity (Temporal)

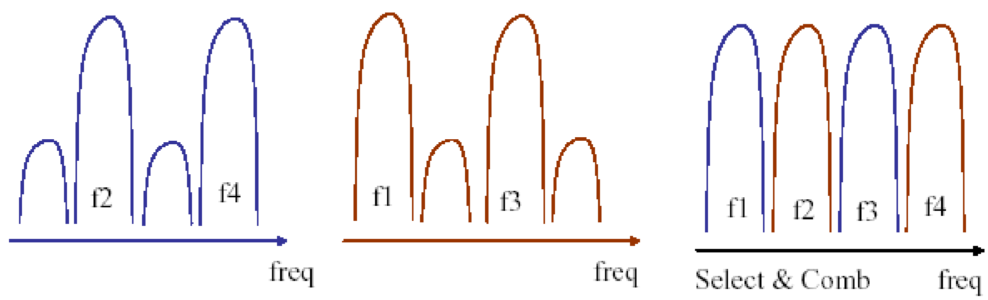
• Replicas of the transmitted waveform are provided across time by a combination of channel coding and time interleaving strategies.
• Channel must provide sufficient variations in time. ie, coherence time of the channel Is small compared to the desired interleaving symbol duration.



Frequency Diversity

• Replicas of the transmitted waveform are provided in the form of redundancies in the frequency domain.
• Coherence bandwidth of the channel must be small compared to the bandwidth of the

signal.(so, different parts of the relevant spectrum will suffer independent fades.



Spatial Diversity (Antenna Diversity)

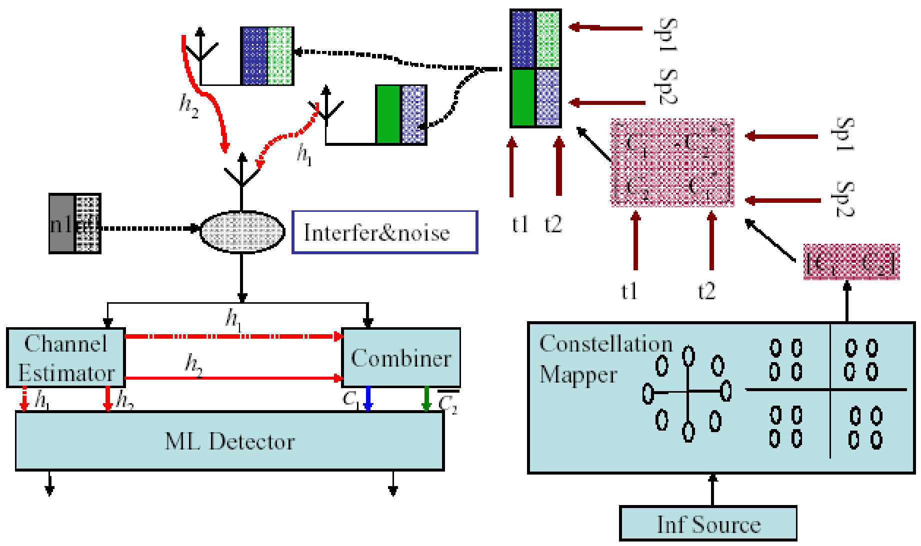
- Replicas of the transmitted waveform are provided across different antennas of the receiver.
- The antenna spacing must be larger than the coherence distance in order to ensure independent fades across different antennas.
- Various Methods: Selection diversity, Maximal Ratio Receiver Diversity and Equal Gain Diversity.
- The special type of spatial diversity is space-time coded diversity.

Space-Time Coding (STC)

- Unlike traditional spatial diversity: information is encoded by multiple transmitter antennas across both space and time (space and frequency).
- Space-Time Coding can be categorized into two classes.
- A: **Space-Time Block Codes:** These codes are transmitted using an orthogonal block structure which enables simple decoding at the receiver. But, it can give only diversity gain.
- B: **Space-Time Trellis Codes:** These are convolution codes extended to the case of multiple transmit and receive antennas. It can give not only diversity gain but also coding gain. But, complex decoding.

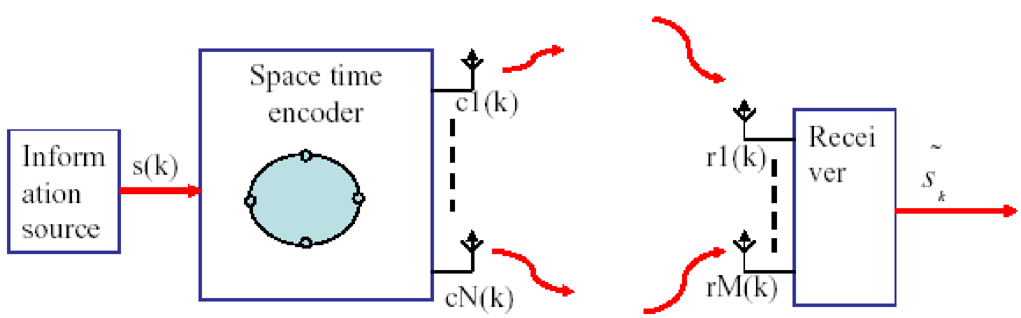
Space-Time Block Coding (STBC)

Invented by Alamouti. Two transmit and N receive antennas. No need Channel state information (CSI) at transmitter but perfect CSI is needed in receiver. H unchanged over two consecutive symbols.



Space-Time Trellis Coding (STTC)

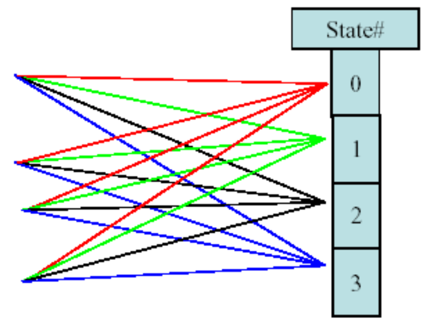
• Basic Diagram of STTC System.



Space-Time Trellis Coding

- 4 PSK STTC: Two transmit STTC with 4 states.
- Initial state of encoder is zero. Output and next state will be depend on the input bits.
- Spectral efficiency = 2 bits/s/Hz

Input bits	00	01	10	11
State 0,Output for Ant1 & Ant 2	00	01	02	03
State 1,Output for Ant1 & Ant 2	10	11	12	13
State 2,Output for Ant1 & Ant 2	20	21	22	23
State 3,Output for Ant1 & Ant 2	30	31	32	33

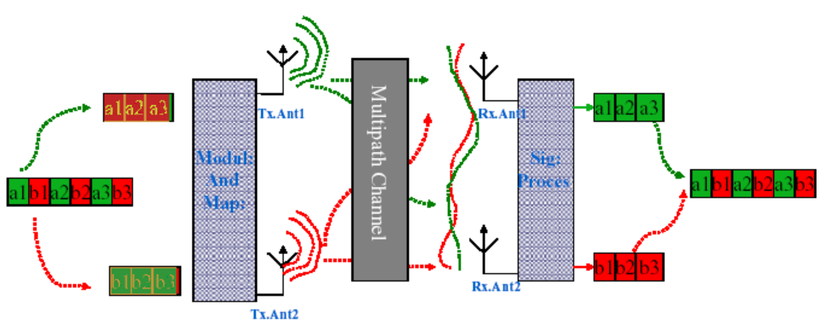


Spatial Multiplexing (BLAST)

- Exploits the rich scattering wireless channel allowing the receiver antennas to detect the different signals simultaneously transmitted by the transmit antennas in the same frequency.
- Capacity gain linearly increase (Num of Ant)
- Extra bandwidth & Total Tx power (No need)
- Different from that of STC method, (STC improve information protection and increase diversity gain).

Basic Spatial Multiplexing Scheme

Basic spatial multiplexing scheme with two TX and two RX antennas.



Rich Scattering Multipaths Channel

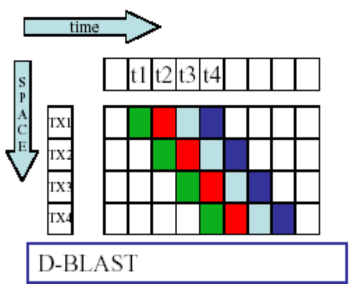
Num of Rx Ant >= Num of Tx Ant

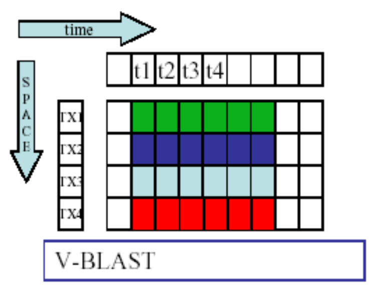
Operation of Spatial Multiplexing Systems

• The transmitted data stream is demultiplexed into *Nt* lower rate streams.
• After coding and modulation processing, each demultiplexed sub-stream is ransmitted Simultaneously from the *Nt* transmit antennas.
• The sub-stream are co-channel signals, that is , they have the same frequency band.
• At the receiver, each antenna observes a superposition of the transmitted signals, separates them into constituents data streams, and multiplexed them in order to recover the original data stream.

Different Kinds of Spatial Multiplexing

• Diagonal Bell Laboratories Layered Space-Time architecture **(D-BLAST),** Approach to the theoretical capacity limit of MIMO systems. Need complex coding processing (time & sub-layers (antennas))
• Vertical Bell Laboratories Layered Space-Time architecture **(V-BLAST)**, Simplified version. No coding required although channel coding may be applied to individual antennas (sub-layers).





Advantages & Disadvantages

NO extra transmit power and bandwidth required. (**ALL Schemes**)

	Adv	Dis Adv
STBC	• Simple, Diversity Gain	• No Coding Gain
STTC	• Diversity & Coding Gain	• Complex
VBLAST	• Simple SPM, No need coding . High throughput	• No Diversity Gain, Rx Ant>=Tx Ant
DBLAST	• No need RxAnt>=TxAnt Approach to Theoretical Limits Capacity Throughput	• Complex Coding in Time and Space.

By Tilak De Silva

Beamforming

Beamforming is a signal processing technique used in sensor arrays for directional signal transmission or reception. This spatial selectivity is achieved by using adaptive or fixed receive/transmit beampattern. The improvement compared with an omnidirectional reception/transmission is known as the receive/transmit gain (or loss).

Beamforming can be used for both radio or sound waves. It has found numerous applications in radar, sonar, seismology, wireless communications, radio astronomy, speech, and biomedicine. Adaptive beamforming is used to detect and estimate the signal-of-interest at the output of a sensor array by means of data-adaptive spatial filtering and interference rejection.

Beamforming techniques

Beamforming takes advantage of interference to change the directionality of the array. When transmitting, a beamformer controls the phase and relative amplitude of the signal at each transmitter, in order to create a pattern of constructive and destructive interference in the wavefront. When receiving, information from different sensors is combined in such a way that the expected pattern of radiation is preferentially observed.

For example in sonar, to send a sharp pulse of underwater sound towards a ship in the distance, simply transmitting that sharp pulse from every sonar projector in an array simultaneously fails because the ship will first hear the pulse from the speaker that happens to be nearest the ship, then later pulses from speakers that happen to be the further from the ship. The beamforming technique involves sending the pulse from each projector at slightly different times (the projector closest to the ship last), so that every pulse hits the ship at exactly the same time, producing the effect of a single strong pulse from a single powerful projector. The same thing can be carried out in air using loudspeakers, or in radar/radio using antennas.

In passive sonar, and in reception in active sonar, the beamforming technique involves combining delayed signals from each hydrophone at slightly different times (the hydrophone closest to the target will be combined after the longest delay), so that every signal reaches the output at exactly the same time, making one loud signal, as if the signal came from a single, very sensitive hydrophone. Receive beamforming can also be used with microphones or radar antennas.

With narrow-band systems the time delay is equivalent to a "phase shift", so in this case the array of antennas, each one shifted a slightly different amount, is called a phased array. A narrow band system, typical of radars, is one where the bandwidth is only a small fraction of the centre frequency. With wide band systems this approximation no longer holds, which is typical in sonars.

In the receive beamfomer the signal from each antenna may be amplified by a different "weight." Different weighting patterns (e.g., Dolph-Chebyshev) can be used to achieve the desired sensitivity patterns. A main lobe is produced together with nulls and sidelobes. As well as controlling the main lobe width (the beam) and the sidelobe levels, the position of a null can be

controlled. This is useful to ignore noise or jammers in one particular direction, while listening for events in other directions. A similar result can be obtained on transmission.
For the full mathematics on directing beams using amplitude and phase shifts, see the mathematical section in phased array.
Beamforming techniques can be broadly divided into two categories.

- conventional (fixed) beamformers or switched beam smart antennas
- adaptive beamformers or adaptive array smart antennas
  - Desired signal maximization mode
  - Interference signal minimization or cancellation mode

Conventional beamformers use a fixed set of weightings and time-delays (or phasings) to combine the signals from the sensors in the array, primarily using only information about the location of the sensors in space and the wave directions of interest. In contrast, adaptive beamforming techniques, generally combine this information with properties of the signals actually received by the array, typically to improve rejection of unwanted signals from other directions. This process may be carried out in the time or frequency domains.
As the name indicates, an adaptive beamformer is able to adapt automatically its response to different situations. Some criterion has to be set up to allow the adaption to proceed such as minimising the total noise output. Because of the variation of noise with frequency, in wide band systems it may be desirable to carry out the process in the frequency domain.
Beamforming can be computationally intensive. Sonar phased array has a data rate slow enough that it can be processed in real-time in software, which is flexible enough to transmit and/or receive in several directions at once. In contrast, radar phased array has a data rate so fast that it usually requires dedicated hardware processing, which is hard-wired to transmit and/or receive in only one direction at a time. However, newer field programmable gate arrays are fast enough to handle radar data in real-time, and can be quickly re-programmed like software, blurring the hardware/software distinction.
Sonar beamforming requirements Sonar itself has many applications, such as wide-area-search-and-ranging, underwater imaging sonars such as side-scan sonar and acoustic cameras.
Sonar beamforming implementation is similar in general technique but varies significantly in detail compared to electromagnetic system beamforming implementation. Sonar applications vary from 1 Hz to as high as 2 MHz, and array elements may be few and large, or number in the hundreds yet very small. This will shift sonar beamforming design efforts significantly between demands of such system components as the "front end" (transducers, preamps and digitizers) and the actual beamformer computational hardware downstream. High frequency, focused beam, multi-element imaging-search sonars and acoustic cameras often implement fifth-order spatial processing that places strains equivalent to Aegis radar demands on the processors.
Many sonar systems, such as on torpedoes, are made up of arrays of up to 100 elements that must accomplish beamsteering over a 100 degree field of view and work in both active and passive modes.

Sonar arrays are used both actively and passively in 1, 2, and 3 dimensional arrays.

- 1 dimensional "line" arrays are usually in multi-element passive systems towed behind ships and in single or multi-element side scan sonar.
- 2 dimensional "planar" arrays are common in active/passive ship hull mounted sonars and some side-scan sonar.
- 3 dimensional spherical and cylindrical arrays are used in 'sonar domes' in the modern submarine and ships.

Sonar differs from radar in that in some applications such as wide-area-search all directions often need to be listened to, and in some applications broadcast to, simultaneously. Thus a multibeam system is needed. In a narrowband sonar receiver the phases for each beam can be manipulated entirely by signal processing software, as compared to present radar systems that use hardware to 'listen' in a single direction at a time.

Sonar also uses beamforming to compensate for the significant problem of the slower propagation speed of sound as compared to that of electromagnetic radiation. In side-look-sonars, the speed of the towing system or vehicle carrying the sonar is moving at sufficient speed to move the sonar out of the field of the returning sound "ping". In addition to focusing algorithms intended to improve reception, many side scan sonars also employ beam steering to look forward and backward to "catch" incoming pulses that would have been missed by a single sidelooking beam.

Beamforming schemes

- A conventional beamformer can be a simple beamformer also known as delay-and-sum beamformer. All the weights of the antenna elements can have equal magnitudes. The beamformer is steered to a specified direction only by selecting appropriate phases for each antenna.
- Null-steering beamformer
- Frequency domain beamformer

Beamforming history in cellular standards

Beamforming techniques used in cellular phone standards have advanced through the generations to make use of more complex systems to achieve higher density cells, with higher throughput.

- Passive mode: (almost) non-standardized solutions
  - Wideband Code Division Multiple Access (WCDMA) supports direction of arrival (DOA) based beamforming
- Active mode: mandatory standardized solutions
  - 2G — Transmit antenna selection as an elementary beamforming
  - 3G — WCDMA: Transmit antenna array (TxAA) beamforming

Chapter 5 – (Telegraphs & Long-Distance Communications)

Approaches to long-distance communications

- Courier: physical transport of the message
  -Messenger pigeons, pony express, FedEx

- Telegraph: message is transmitted across a network using signals
  -Drums, beacons, mirrors, smoke, flags, semaphores…

  -Electricity, light

- Telegraph delivers message much sooner

Optical (Visual) Telegraph

- Claude Chappe invented optical telegraph in the 1790's
- Semaphore mimicked a person with outstretched arms with flags in each hand
- Different angle combinations of arms & hands generated hundreds of possible signals
- Code for enciphering messages kept secret
- Signal could propagate 800 km in 3 minutes!

Electric Telegraph

- William Sturgeon Electro-magnet (1825)
  - Electric current in a wire wrapped around a piece of iron generates a magnetic force
- Joseph Henry (1830)
  - Current over 1 mile of wire to ring a bell
- Samuel Morse (1835)
  - Pulses of current deflect electromagnet to generate dots & dashes
  - Experimental telegraph line over 40 miles (1840)
- Signal propagates at the speed of light!!!
  - Approximately 2 x 108 meters/second in cable

Digital Communications

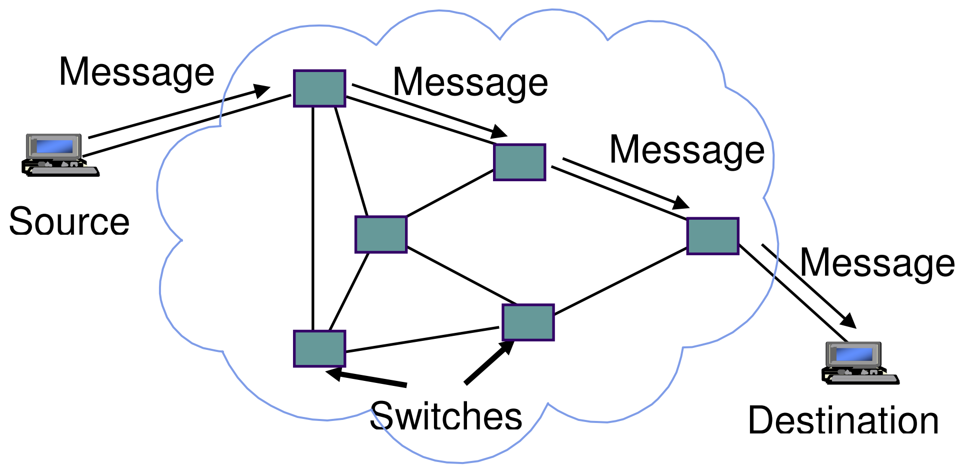
- Morse code converts text message into sequence of dots and dashes
- Use transmission system designed to convey dots and dashes

	Morse Code		Morse Code		Morse Code		Morse Code
A	· —	J	· — — —	S	· · ·	2	· · — — —
B	— · · ·	K	— · —	T	—	3	· · · — —
C	— · — ·	L	· — · ·	U	· · —	4	· · · · —
D	— · ·	M	— —	V	· · · —	5	· · · · ·
E	·	N	— ·	W	· — —	6	— · · · ·
F	· · — ·	O	— — —	X	— · · —	7	— — · · ·
G	— — ·	P	· — — ·	Y	— · — —	8	— — — · ·
H	· · · ·	Q	— — · —	Z	— — · ·	9	— — — — ·
I	· ·	R	· — ·	1	· — — — —	0	— — — — —

Electric Telegraph Networks

- Electric telegraph networks exploded
  - Message switching & Store-and-Forward operation
  - Key elements: Addressing, Routing, Forwarding

- Optical telegraph networks disappeared



Elements of Telegraph Network Architecture

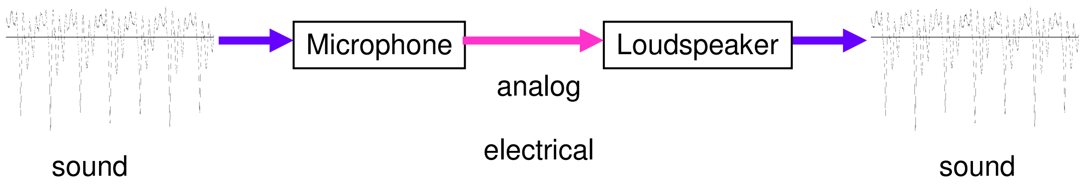
- Digital transmission
  - Text messages converted into symbols (dots/dashes, zeros/ones)
  - Transmission system designed to convey symbols
- Multiplexing
  - *Framing* needed to recover text characters
- Message Switching
  - Messages contain source & destination *addresses*
  - *Store-and-Forward:* Messages forwarded hop-by-hop across network
  - *Routing* according to destination address

Bell's Telephone

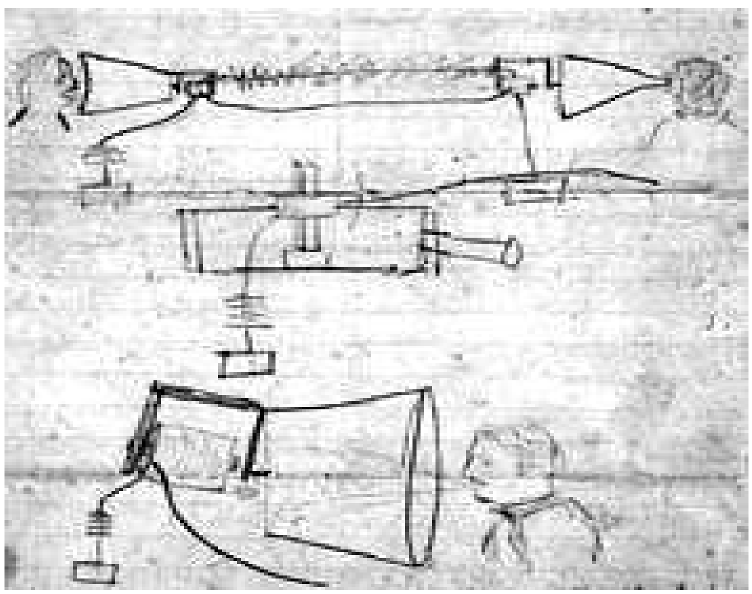
- Alexander Graham Bell (1875) working on harmonic telegraph to multiplex telegraph signals
- Discovered voice signals can be transmitted directly
  - Microphone converts voice pressure variation (sound) into *analogous* electrical signal
  - Loudspeaker converts electrical signal back into sound

- Telephone patent granted in 1876
- Bell Telephone Company founded in 1877

Signal for "ae" as in cat

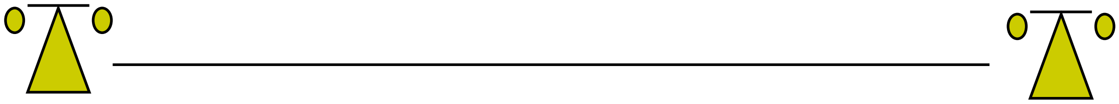


Bell's Sketch of Telephone



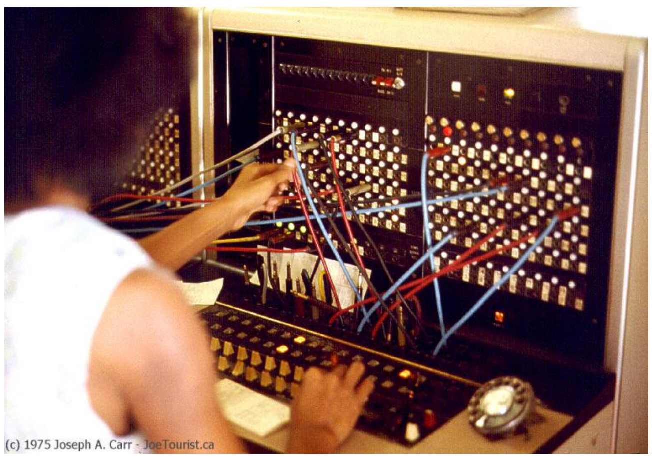
Signaling

- Signaling required to establish a call
  - Flashing light and ringing devices to alert the called party of incoming call
  - Called party information to operator to establish calls



Signaling + voice signal transfer

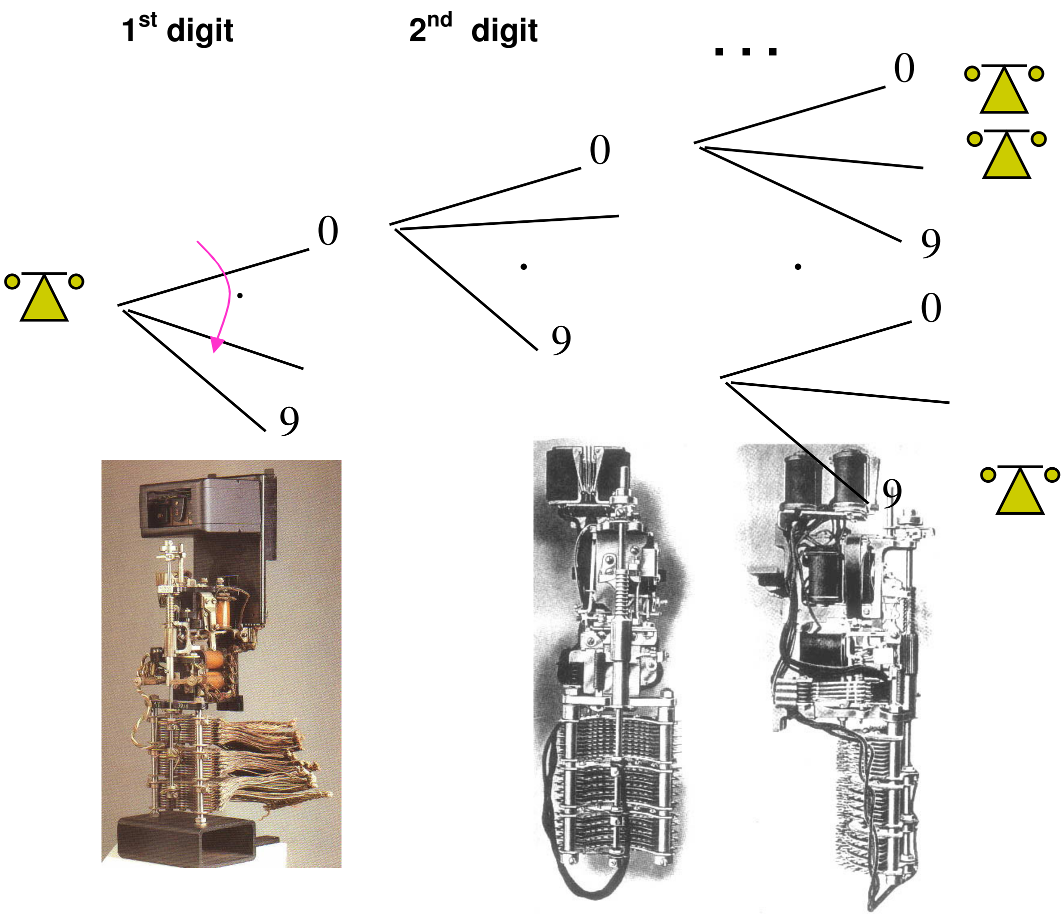
Manual Telephone Switching



In the mid-1870's Alexander Graham Bell invented the telephone; a wired system for two way voice communication between remote locations. You spoke into a unit at one location and you voice was heard at the other location, immediately. This system was somewhat limited in that it only allowed communication with one fixed location, so it was an obvious advance to have lines going to other locations. Initially, this is what happened - each telephone had lines going to many other telephones, which meant a lot of wires and there were practical limitations as to the number of phones one could connect to.Soon, the idea of central switching was developed. Each telephone connected to a central hub (the exchange) and from there, the operator would connect your call to another subscriber. Thus switchboards were developed where upon lifting your receiver, an operator was alerted and you would tell her who you wished to speak to (telephone operators were almost exclusively women in the early days). The operator would then take a wire from your socket on her switchboard and plug you into the other persons socket. When you completed the call, you would hang up

Strowger Exchange

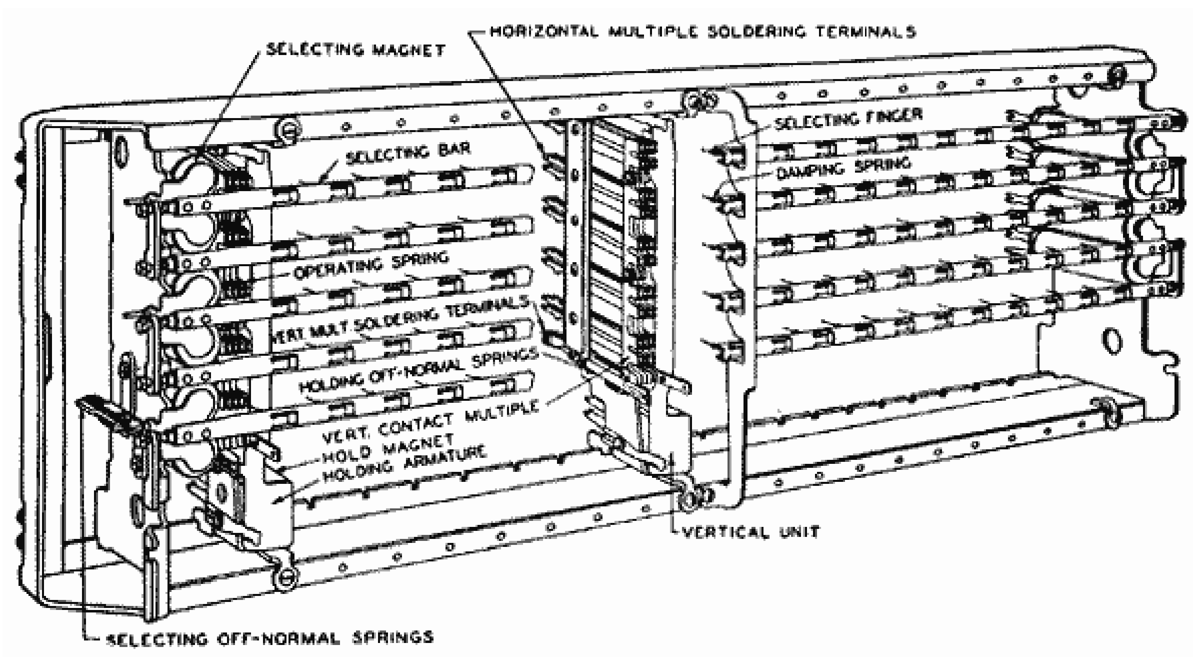
- Human operators intelligent & flexible
  - But expensive and not always discreet
- Strowger invented automated switch in 1888
  - Each current pulse advances wiper by 1 position
  - User dialing controls connection setup
- Decimal telephone numbering system
- Hierarchical network structure simplifies routing
  - Area code, exchange (CO), station number



Almon B. Strowger was an undertaker in Kansas City, USA. The story goes that there was a competing undertaker locally whose wife was an operator at the local (manual) telephone exchange. Whenever a caller asked to be put through to Strowger, calls were deliberately put through to his competitor. This obviously frustrated Strowger greatly and he set about devising a system for doing away with the human part of the equation !

Strowger developed a system of automatic switching using an electromechanicalcal switch based around around electromagnets and pawls. With the help of his nephew (Walter S. Strowger) he produced a working model in 1888 (US Patent No. 447918 10/6/1891). In this selector, a moving wiper (with contacts on the end) moved up to and around a bank of many other contacts, making a connection with any one of them.
Strowger did not invent the idea of automatic switching; it was first invented in 1879 by Connolly & McTigthe but Strowger was the first to put it to effective use. Together with Joseph B. Harris and Moses A. Meyer, Strowger formed his company 'Strowger Automatic Telephone Exchange' in October 1891.
In the late 1890's Almon B. Strowger retired and eventually died in 1902. In 1901, Joseph Harris licenced the Strowger selectors to the Automatic Electric Co. (AE); the two companies merged in 1908. The company still exists today as AG Communications Systems (www.agcs.com), having undergone various corporate changes and buyouts along the way.

Crossbar Exchange



A

crossbar switch is one of the principal architectures used to construct switches of many types. Originally the term was used literally, for a matrix switch controlled by a grid of crossing metal bars, and later was broadened to matrix switches in general. Crossbar switches are sometimes referred to as "cross-point switches," "crosspoint switches," or "matrix switches." The other principal switch architectures are that of a memory switch or a crossover switch.

General properties

Crossbar switches have a characteristic matrix of switches between the inputs to the switch and the output of the switch. If the switch has M inputs and N outputs, then a crossbar has a matrix with M x N cross-points or places where the "bars" "cross". A given crossbar is a single layer, non-blocking switch. Collections of crossbars can be used to implement multiple layer and/or blocking switches.

Implementations

Historically, a crossbar switch consisted of metal bars associated with each input and output, controlling movable contacts at each cross-point. In the later part of the 20th Century these literal crossbar switches declined and the term came to be used figuratively for rectangular array switches in general.
As you might imagine, this was very labour intensive and as the polularity of the telephone grew, the number of operators employed by the Post Office grew; large switching centres (exchanges) could have many tens of operators, each with their own switchboard. Note that in the early days, each telephone user was known as the 'subscriber' and that term is still used today.

Electronic switching

Research on electronic switching[2] started soon after the Second World War, but commercial fully electronic exchanges began to emerge only about 30 years later. However, electronic techniques proved economic for common controls much earlier. ln electromechanical exchanges, common controls mainly use switches and relays which were originally designed for use in switching networks. In common controls, they are operated much more frequently and wear out earlier. In contrast, the life of an electronic device is almost independent of its frequency of operation. This gave an incentive for developing electronic common controls and resulted in electronic replacements for registers, markers, etc., having much greater reliability than their electromechanical predecessors. Advances made in computer technology were incorporated[2] and led to the development of stored—program control (SPC). This enables a digital computer to be used as a central control and perform different functions with the same hardware by executing different programs. As a result, SPC exchanges can offer a wider range of facilities than earlier systems. ln addition, the facilities provided to an individual customer can be readily altered by changing the customer’s class—of-service data stored in a central electronic memory. Moreover, since the processor’s stored data can be altered electronically, some of these facilities can be controlled by customers. Examples

Include:

l. *Call barring* (*outgoing or incoming*) The customer can prevent unauthorized calls being made and can prevent incoming calls when wishing to be left in peace.

2. *Repeat lost call* lf a called line is engaged, the caller can try again later without having to redial the full number.

3. *Reminder coils* The exchange can be instructed to call the customer at a pre—arranged time (e.g. for a wake-up call).

4. *Call diversion* The exchange can be instructed to connect calls to a different number when the customer goes away.

5. *Three-way calls* The customer can instruct the exchange to connect a third party to a call that is already in progress.

6. *Charge advice* As a result of the caller sending the appropriate instruction when starting a call, the exchange calls back at the end of the call to indicate the call duration and charge.

In order to develop a fully electronic exchange it was necessary to replace electromechanical switches in the speech path, in addition to using electronic common controls. One approach is to replace the relay contacts of the switch shown in Figure 3.l I by electronic devices multipled together. A diode could be used as a crosspoint, as shown in Figure 3.21. If point A is positive the diode is reverse biased and the crosspoint is open. If point A is negative the diode conducts and the crosspoint is closed. However, an external circuit is needed to keep point A negative for the duration of a connection. Thus, in order to implement a crosspoint, a one-bit memory is required in addition to the switching element. For example, a crossbar switch provides this memory by means of the hnger which is trapped in each crosspoint when it is operated. Research was carried out into devices which could combine both functions in a single component. Cold-cathode gas tubes and PNPN semiconductor devices have been used.

Another approach is to use a multiplex system instead of multipled elements. A frequeney-division multiplex (FDM) system could be used as a switch by bringing the two ends ofits transmission path together. Ifthe modems at one end ofthe path operate at fixed frequencies, but those at the other end can operate at the frequency of any channel, then any trunk at one side ofthe switch can be connected to any trunk at the other side. On one side ofthe switch it must be possible to tune the filters to any channel and to supply any carrier frequency to any modem. This proved too expensive to put into practice.

Nevertheless, the multiplex principle was not abandoned. A multipled-element switch with N incoming and outgoing trunks needs N2 elements. A multiplex switch of the same size needs N devices on each side, i.e. a total of only 2N. Thus, if N is sufhciently large the devices used can be much more costly than those of a multipled-element switch. A time-division multiplex (TDM) system, as shown in Figure 2.8, can be used. If any of the N receiving gates is operated by a train of pulses coincident with those applied to one of the N sending gates, then a transmission path is provided from the incoming trunk to the outgoing trunk via the common highway. For a transmission system, fixed pulse timings are used. If the pulse timings can be altered. Then any incoming trunk can be connected to any outgoing trunk, i.e. an N * N switch

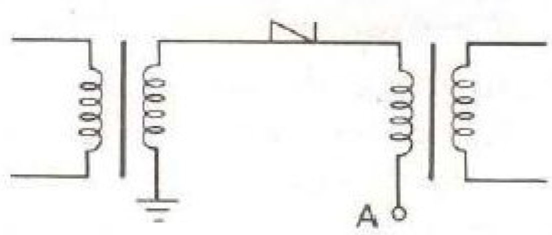


Figure 3.21 Diode crosspoint.

is obtained. Moreover, a memory which stores the appropriate pulse timings can be implemented relatively cheaply. Research on TDM switching proved successful and systems using it are now in service world wide. They are described in- Chapter 6.

Consequently, switching systems may be classified-as:

1. *Space-division {SD} system* Each connection is made over a different path in space which exists for the duration of the connection.

2. *Time- division {TD} system* Each connection is made over the same path in space, but at different instants in time.

All electromechanical exchanges are, of course, SD system. Both SD and TD electronic switching systems have been developed, but the-latter now predominate.

The distinction between SD and TD applies to the control arrangements of an exchange as well as to its switching network. For example, if an exchange uses individual registers, this is space division. lf the register function is performed by a central processor, handling each call in turn, this is time division.

Some fully electronic systems were given held trials over 30 years ago.[2] They worked, but they were too expensive to enter commercial production. The chief reason for this was that the switches (whether SD or TD) were-suitable for speech signals, but they could not handle the relatively high currents and voltages required for line feeding and ringing. These had to be provided from every customer's line circuit, instead of from a much smaller number of supervisory trunks as in electromechanical systems. It is also more difficult to provide adequate protection from high-voltage surges on external lines for electronic components than for electromechanical apparatus. Testing of the insulation resistance of a line is not possible without disconnecting it from an electronic line circuit. lf time-division switching is used, the speech path is a four—wire circuit, so this requires a hybrid transformer in the customer's line circuit. It took about another ten years for these problems to be solved in an economic fashion.

Reed-electronic systems

The solution to the problem of high-cost line circuits was to exploit the advantages of electronic control, but to revert to electromechanical switches, with a much-improved cross point, in the form of the reed relay shown in Figure 3.22. A single operating coil, as shown in Figure 3.23. can contain three such reed inserts to switch +, - and P wires, together with a fourth which acts as the one-bit memory to hold the relay operated for the duration of a connection. A matrix of reed relays is equivalent to a crossbar switch and similar trunking principles can be used.

An alternative design of crosspoint, the ferreed,[l7] uses a remanent ferrite material in its magnetic circuit instead of the latching contact. It is therefore released by a pulse of current in the opposite direction from its operating current. An improved form offerreed, known as the remreed, is used in the Bell No.l ESS system.[l 8, l 9] This system uses a *map in memory*. The busy/free conditions of all the trunks are stored in a central memory.

Stored Program Control exchange

Stored Program Control exchange (SPC) is the technical name used for telephone exchanges controlled by a computer program stored in the memory of the system. Early exchanges such as Strowger, Panel, Rotary, and Crossbar were electro mechanical and had no software control. SPC was introduced on a small scale in so called Electronic switching systems in the 1960s (the 101ESS PBX was a minor Bell System example) and on a large scale (1ESS switch was the major Bell System example) in the 1970s. SPC allowed more sophisticated Calling features. As SPC exchanges evolved, reliability and versatility increased. In the 1980s SPC completely took over the industry, hence the term lost all except historical interest.

Stored – program control

Processor architecture

In order to obtain adequate security a switching system with central processors requires a minimum of two. If two processors are used, each must carry the full traffic load if the other fails, so each must have sufficient processing capacity to do this. Two processors may be configured to operate in the following ways:

- Worker and standby
- Load sharing
- Synchronous operation.

In a system with a worker and a standby, either cold standby or hot standby may be used. The term ‘cold standby’ is really a misnomer; the spare processor must be switched on and working. However, its memories are not updated, so there will be disruption of calls when change-over takes place. In hot standby, the spare processor is constantly updated with details of all calls; it can therefore assume control without disruption.

In load sharing, both processors are working independently. Thus, at any time, they will be performing different tasks for different calls. This makes it difficult for the remaining processor to take over from the failed one without disruption of traffic. Contention between the processors must be prevented and each must update the other

Stored-program control (SPC) is a term used to describe computer-controlled telecommunication systems. When SPC was originally conceived, computers were still large and expensive machines and their programs were generally written in low-level languages to achieve run-time efficiencies. Integrated circuits had not yet appeared on the scene and the idea of using computers for controlling a telephone exchange rather than as number-crunching data processors was a novel one.

Today, of course, VLSI, microprocessors and dense memory chips are commonplace hardware building blocks and software engineering is emerging as a recognised discipline from the various software crises through which computing has passed in the intervening period. SPC remains as a slightly anachronistic term from me early days, but the application of computers and software-based control to telecommunications is at the vanguard of present day concerns in information technology.

Characteristics SPC systems

SPC systems in telecommunication networks share many features with other computer-controlled systems, particularly those involved with real-time applications such as process control, however. Telecommunications presents the following combination of requirements which is not commonly faced in other fields:

(a) Very high reliability and Availability.

SPC systems are required to provide service for 24 hours a day, year in, year out. During a lifetime which may span two or more decades, they are expected to be out of action for at most a few hours.

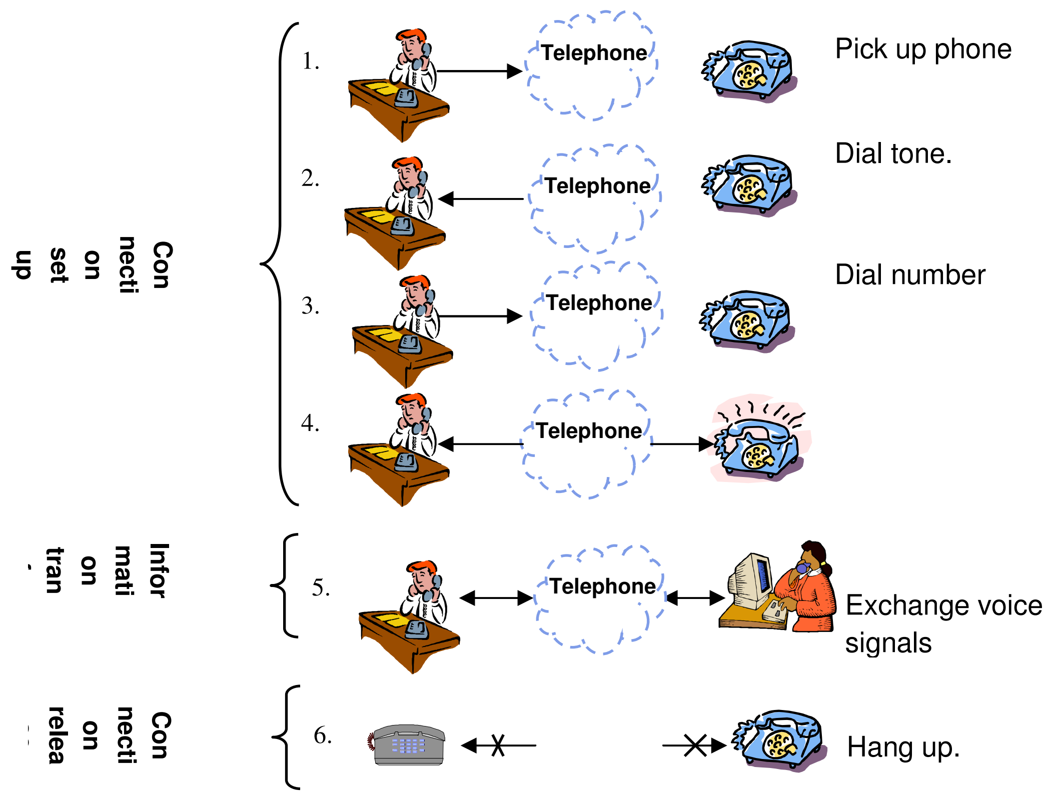
(b) Very large peripheral connectivity.

Telephone networks support many millions of customers and a single telephone exchange can have tens of thousands of customer terminals.

(c) High degree of event-driven concurrency.

SPC systems make extensive use of time-multiplexed operation in all areas of the network, to maximise the cost-effectiveness of equipment. A large telephone exchange for example, must handle hundreds of processing-intensive call setups simultaneously and thousands of calls may be in speech at the same time.

Three Phases of a Connection



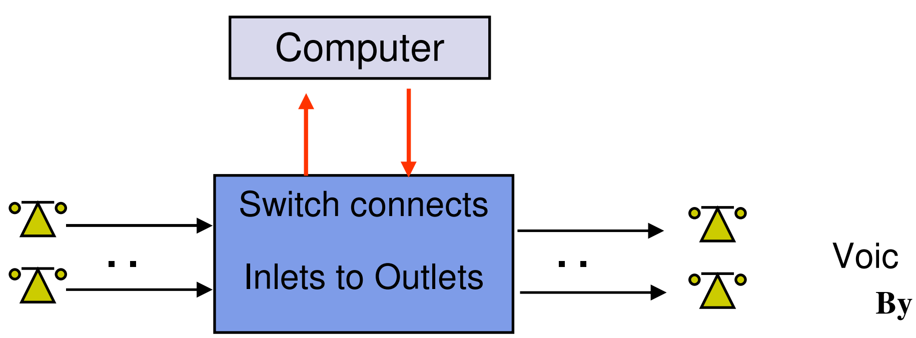
Computer Connection Control

- A computer controls connection in telephone switch
- Computers exchange *signaling messages* to:
  - Coordinate set up of telephone connections

  - To implement new services such as caller ID, voice mail, etc.

  - To enable *mobility and roaming in* cellular networks

- "Intelligence" inside the network
- A separate *signaling network* is required



Digitization of Telephone Network

- Pulse Code Modulation digital voice signal
  - Voice gives 8 bits/sample x 8000 samples/sec = 64x103 bps

- Time Division Multiplexing for digital voice
  - T-1 multiplexing (1961) : 24 voice signals = 1.544x106 bps

- Digital Switching (1980s)
  - Switch TDM signals without conversion to analog form

- Digital Cellular Telephony (1990s)

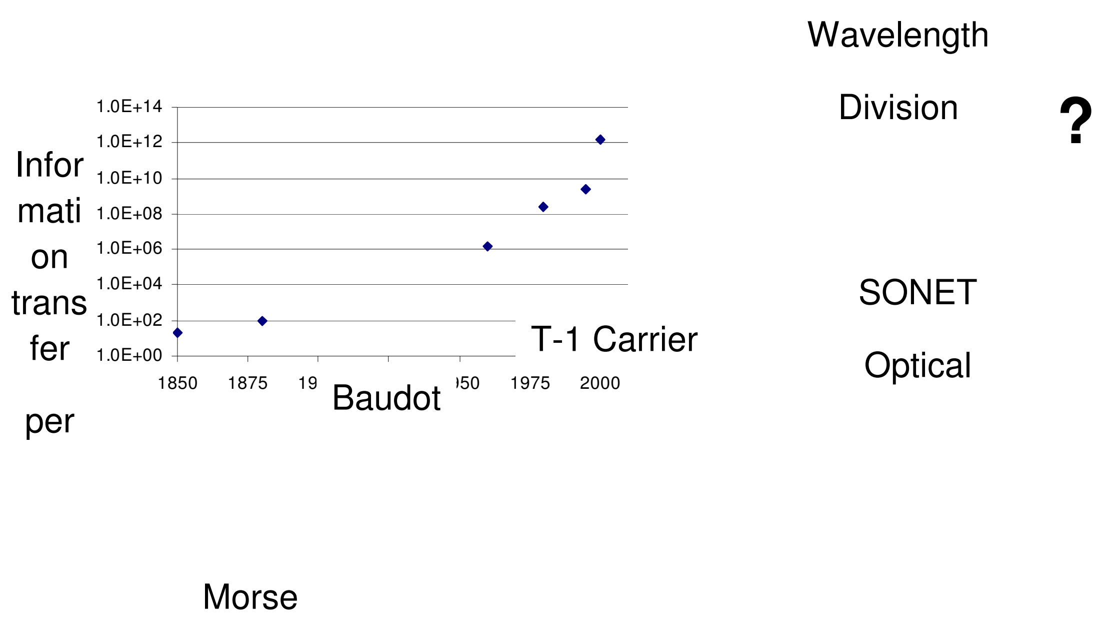
- Optical Digital Transmission (1990s)

  - One OC-192 optical signal = 10x109 bps

  - One optical fiber carries 160 OC-192 signals = 1.6x1012 bps!

All digital transmission, switching, and control

Digital Transmission Evolution



Elements of Telephone Network Architecture

- Digital transmission & switching
  - Digital voice; Time Division Multiplexing

- Circuit switching
  - User signals for call setup and tear-down
  - Route selected during connection setup
  - End-to-end connection across network
  - Signaling coordinates connection setup

- Hierarchical Network
  - Decimal numbering system
  - Hierarchical structure; simplified routing; scalability

- Signaling Network
  - Intelligence inside the network

Signalling

Signaling Links

SS7 messages are exchanged between network elements over 56 or 64 kilobit per second (kbps) bi-directional channels called signaling links. Signaling occurs out-of-band on dedicated channels rather than in-band on voice channels. Compared to in-band signaling, out-of-band signaling provides:

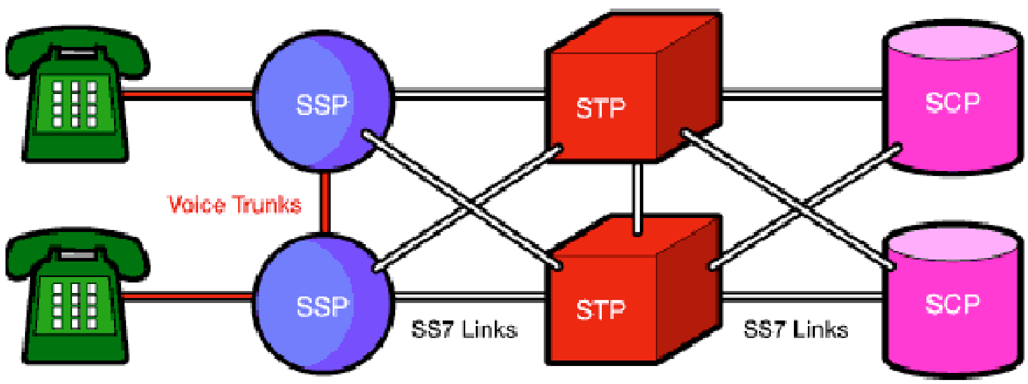
- Faster call setup times (compared to in-band signaling using multi-frequency (MF) signaling tones)
- More efficient use of voice circuits
- Support for Intelligent Network (IN) services, which require signaling to network elements without voice trunks (e.g., database systems)
- Improved control over fraudulent network usage

Signaling Points

Each signaling point in the SS7 network is uniquely identified by a numeric point code. Point codes are carried in signaling messages exchanged between signaling points to identify the source and destination of each message. Each signaling point uses a routing table to select the appropriate signaling path for each message.

There are three kinds of signaling points in the SS7 network.

- **SSP** (Service Switching Point)
- **STP** (Signal Transfer Point)
- **SCP** (Service Control Point)



SSPs are switches that originate, terminate or tandem calls. An SSP sends signaling messages to other SSPs to setup, manage and release voice circuits required to complete a call. An SSP may also send a query message to a centralized database (an SCP) to determine how to route a call (e.g., a toll-free 1-800/888 call in North America). An SCP sends a response to the originating SSP containing the routing number(s) associated with the dialed number. An alternate routing number may be used by the SSP if the primary number is busy or the call is unanswered within a specified time. Actual call features vary

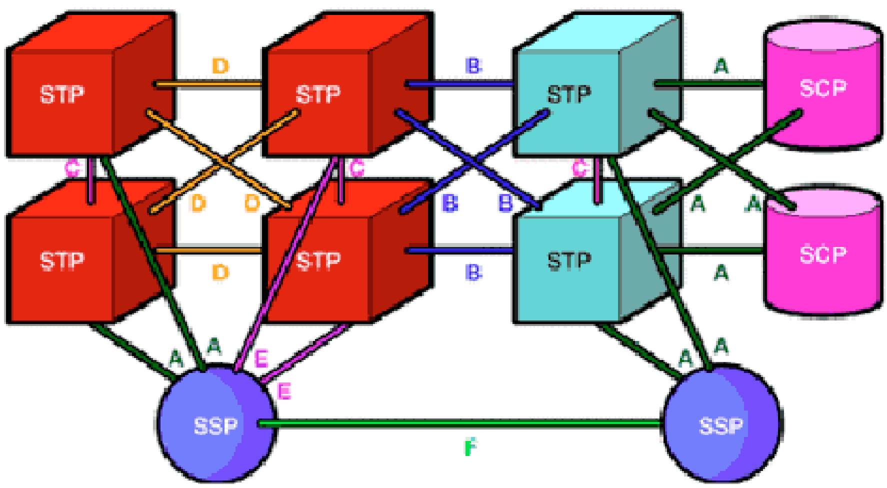
from network to network and from service to service.

Network traffic between signaling points may be routed via a packet switch called an STP. An STP routes each incoming message to an outgoing signaling link based on routing information contained in the SS7 message. Because it acts as a network hub, an STP provides improved utilization of the SS7 network by eliminating the need for direct links between signaling points. An STP may perform global title translation, a procedure by which the destination signaling point is determined from digits present in the signaling message (e.g., the dialed 800 number, calling card number or mobile subscriber identification number).

An STP can also act as a "firewall" to screen SS7 messages exchanged with other networks. Because the SS7 network is critical to call processing, SCPs and STPs are usually deployed in mated pair configurations in separate physical locations to ensure network-wide service in the event of an isolated failure. Links between signaling points are also provisioned in pairs. Traffic is shared across all links in the linkset. If one of the links fails, the signaling traffic is rerouted over another link in the linkset. The SS7 protocol provides both error correction and retransmission capabilities to allow continued service in the event of signaling point or link failures.

SS7 Signaling Link Types

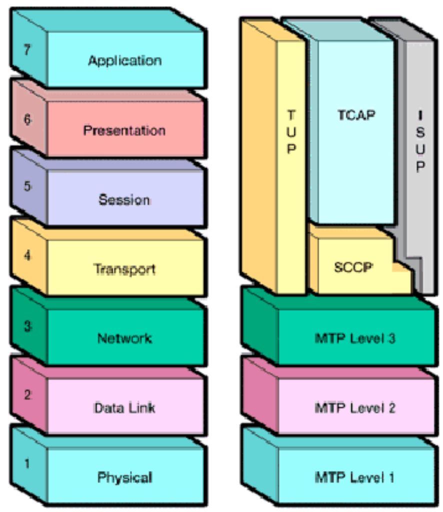
Signaling links are logically organized by link type ("A" through "F") according to their use in the SS7 signaling network.



A Link:	An "A" (access) link connects a signaling end point (e.g., an SCP or SSP) to an STP. Only messages originating from or destined to the signaling end point are transmitted on an "A" link.
B Link:	A "B" (bridge) link connects one STP to another. Typically, a quad of "B" links interconnect peer (or primary) STPs (e.g., the STPs from one network to the STPs of another network). The distinction between a "B" link and a "D" link is rather arbitrary. For this reason, such links may be referred to as "B/D" links.
C Link:	A "C" (cross) link connects STPs performing identical functions into a mated pair. A "C" link is used only when an STP has no other route available to a destination signaling point due to link failure(s). Note that SCPs may also be deployed in pairs to improve reliability; unlike STPs however, mated SCPs are not interconnected by signaling links.
D Link:	A "D" (diagonal) link connects a secondary (e.g., local or regional) STP pair to a primary (e.g., inter-network gateway) STP pair in a quad-link configuration. Secondary STPs within the same network are connected via a quad of "D" links. The distinction between a "B" link and a "D" link is rather arbitrary. For this reason, such links may be referred to as "B/D" links.
E Link:	An "E" (extended) link connects an SSP to an alternate STP. "E" links provide an alternate signaling path if an SSP's "home" STP cannot be reached via an "A" link. "E" links are not usually provisioned unless the benefit of a marginally higher degree of reliability justifies the added expense.
F Link:	An "F" (fully associated) link connects two signaling end points (i.e., SSPs and SCPs). "F" links are not usually used in networks with STPs. In networks without STPs, "F" links directly connect signaling points.

SS7 Protocol Stack

The hardware and software functions of the SS7 protocol are divided into functional abstractions called "levels." These levels map loosely to the **Open Systems Interconnect** (OSI) 7-layer model defined by the International Standards Organization (ISO).



Message Transfer Part

The Message Transfer Part (MTP) is divided into three levels.

The lowest level, **MTP Level 1**, is equivalent to the OSI Physical Layer. MTP Level 1 defines the physical, electrical and functional characteristics of the digital signaling link. Physical interfaces defined include E-1 (2048 kb/s; 32 64 kb/s channels), DS-1 (1544 kb/s; 24 64kb/s channels), V.35 (64 kb/s), DS-0 (64 kb/s) and DS-0A (56 kb/s).

MTP Level 2 ensures accurate end-to-end transmission of a message across a signaling link. Level 2 implements flow control, message sequence validation and error checking. When an error occurs on a signaling link, the message (or set of messages) is retransmitted. MTP Level 2 is equivalent to the OSI Data Link Layer.

MTP Level 3 provides message routing between signaling points in the SS7 network. MTP Level 3 re-routes traffic away from failed links and signaling points and controls traffic when congestion occurs. MTP Level 3 is equivalent to the OSI Network Layer.

ISDN User Part (ISUP)

The ISDN User Part (ISUP) defines the protocol used to set-up, manage and release trunk circuits that carry voice and data between terminating line exchanges (e.g., between a calling party and a called party). ISUP is used for both ISDN and non-ISDN calls. However, calls that originate and terminate at the same switch do not use ISUP signaling.

Telephone User Part (TUP)

In some parts of the world (i.e., China and Brazil), the Telephone User Part (TUP) is used to support basic call setup and tear-down. TUP handles analog circuits only. In many countries, ISUP has replaced TUP for call management.

Signaling Connection Control Part (SCCP)

SCCP provides connectionless and connection-oriented network services and global title translation (GTT) capabilities above MTP Level 3. A global title is an address (e.g., a dialed 800 number, calling card number or mobile subscriber identification number) that is translated by SCCP into a destination point code and subsystem number. A subsystem number uniquely identifies an application at the destination signaling point. SCCP is used as the transport layer for TCAP-based services.

Transaction Capabilities Applications Part (TCAP)

TCAP supports the exchange of non-circuit related data between applications across the SS7 network using the SCCP connectionless service. Queries and responses sent between SSPs and SCPs are carried in TCAP messages. For example, an SSP sends a TCAP query to determine the routing number associated with a dialed 800/888 number and to check the personal identification number (PIN) of a calling card user. In mobile networks (IS-41 and GSM), TCAP carries Mobile Application Part (MAP) messages sent between mobile switches and databases to support user authentication, equipment identification and roaming.

Operations, Maintenance and Administration Part (OMAP) and ASE

OMAP and ASE are areas for future definition. Presently, OMAP services may be used to verify network routing databases and to diagnose link problems.

Message Transfer Part

The Message Transfer Part (MTP) is divided into three levels:

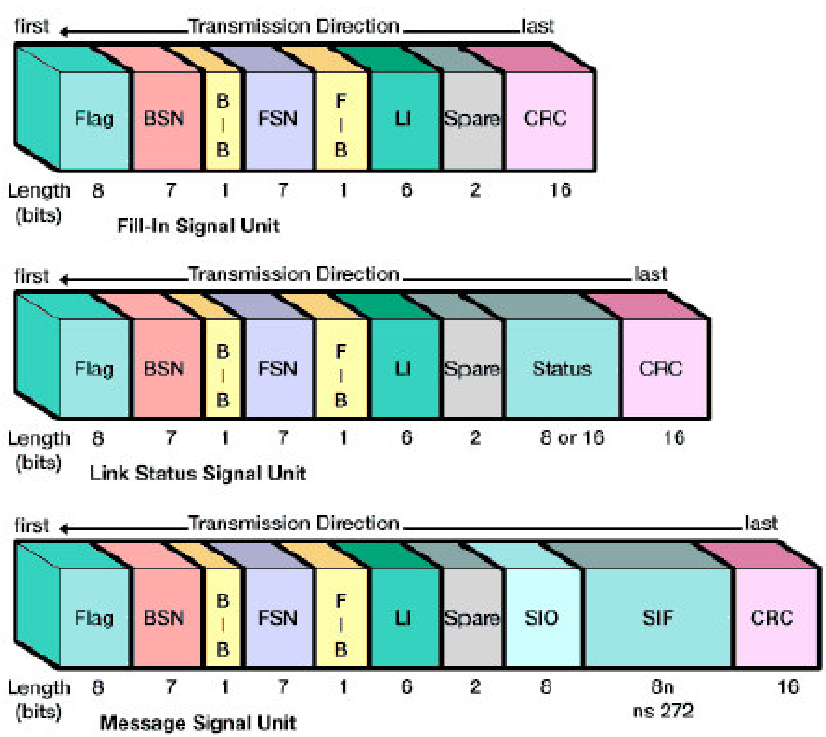
MTP Level 1

The lowest level, MTP Level 1, is equivalent to the OSI Physical Layer. MTP Level 1 defines the physical, electrical and functional characteristics of the digital signaling link. Physical interfaces defined include E-1 (2048 kb/s; 32 64 kb/s channels), DS-1 (1544 kb/s; 24 64 kp/s channels), V.35 (64 kb/s), DS-0 (64 kb/s) and DS-0A (56 kb/s).

MTP Level 2

MTP Level 2 ensures accurate end-to-end transmission of a message cross a signaling link. Level 2 implements flow control, message sequence validation and error checking. When an error occurs on a signaling link, the message (or set of messages) is retransmitted. MTP Level 2 is equivalent to the OSI Data Link Layer.

An SS7 message is called a signal unit (SU). There are three kinds of signal units: Fill-In Signal Units (FISUs), Link Status Signal Units (LSSUs) and Message Signal Units (MSUs).



FISUs are transmitted continuously on a signaling link in both directions unless other signal units (MSUs or LSSUs) are present. FISUs carry basic Level 2 information only (i.e., acknowledgment of signal unit receipt by a remote signaling point). Because a CRC checksum is calculated for each FISU, signaling link quality is checked continuously by both signaling points at either end of the link. (Note: In the ITU-T Japan variant, signaling link quality is checked by the continuous transmission of flag octets (8-bit bytes) rather than FISUs; FISUs are sent only at predefined timer intervals (e.g., once every 150 milliseconds).

LSSUs carry one or two octets (8-bit bytes) of link status information between signaling points at either end of a link. The link status is used to control link alignment and to indicate the status of a signaling point (e.g., local processor outage) to the remote signaling point.

MSUs carry all call control, database query and response, network management and network maintenance data in the signaling information field (SIF). MSUs have a routing label, which allows an originating signaling point to send information to a destination signaling point across the network.

The value of the LI (Length Indicator) field determines the signal unit type:

LI Value	Signal Unit Type
0	Fill-In Signal Unit (FISU)
1.2	Link Status Signal Unit (LSSU)
3.63	Message Signal Unit (MSU)

The 6-bit LI can store values between zero and 63. If the number of octets that follow the LI and precede the CRC is less than 63, the LI contains this number. Otherwise, the LI is set to 63. An LI of 63 indicates that the message length is equal to *or greater than* 63 octets (up to a maximum of 273 octets). The maximum length of a signal unit is 279 octets: 273 octets (data) + one octet (flag) + one octet (BSN + BIB) + one octet (FSN + FIB) + one octet (LI + two bits spare) + two octets (CRC).

Flag

The flag indicates the beginning of a new signal unit and implies the end of the previous signal unit (if any). The binary value of the flag is 0111 1110. Before transmitting a signal unit, MTP Level 2 removes "false flags" by adding a zero-bit after any sequence of five one-bits. Upon receiving a signal unit and stripping the flag, MTP Level 2 removes any zero-bit following a sequence of five one-bits to restore the original contents of the message. Duplicate flags are removed between signal units.

BSN (Backward Sequence Number)

The BSN is used to acknowledge the receipt of signal units by the remote signaling point. The BSN contains the sequence number of the signal unit being acknowledged. (See description under FIB below.)

BIB (Backward Indicator Bit)

The BIB indicates a negative acknowledgment by the remote signaling point when toggled. (See description under FIB below.)

FSN (Forward Sequence Number)

The FSN contains the sequence number of the signal unit. (See description under FIB below.)

FIB (Forward Indicator Bit)

The FIB is used in error recovery like the BIB. When a signal unit is ready for transmission, the signaling point increments the FSN (forward sequence number) by one (FSN = 0..127). The CRC (cyclic redundancy check) checksum value is calculated and appended to the forward message. Upon receiving the message, the remote signaling point checks the CRC and copies the value of the FSN into the BSN of the next available message scheduled for transmission back to the initiating signaling point. If the CRC is correct, the backward message is transmitted. If the CRC is incorrect, the remote signaling point indicates negative acknowledgment by toggling the BIB prior to sending the backward message. When the originating signaling point receives a negative acknowledgment, it retransmits all forward messages, beginning with the corrupted message, with the FIB toggled.

Because the 7-bit FSN can store values between zero and 127, a signaling point can send up to 128 signal units before requiring acknowledgment from the remote signaling point. The BSN indicates the last in-sequence signal unit received correctly by the remote signaling point. The BSN acknowledges all previously received signal units as well. For example, if a signaling point receives a signal unit with BSN = five followed by another with BSN = ten (and the BIB is not toggled), the latter BSN implies successful receipt of signal units six through nine as well.

SIO (Service Information Octet)

The SIO field in an MSU contains the 4-bit subservice field followed by the 4-bit service indicator. FISUs and LSSUs do not contain an SIO.

The *subservice* field contains the network indicator (e.g., national or international) and the message priority (zero to three with three being the highest priority). Message priority is considered only under congestion conditions, not to control the order in which messages are transmitted. Low priority messages may be discarded during periods of congestion. Signaling link test messages receive a higher priority than call setup messages.

The *service indicator* specifies the MTP user (Figure. 6), thereby allowing the decoding of the information contained in the SIF.

Service Indicator	MTP User
0	Signaling Network Management Message (SNM)
1	Maintenance Regular Message (MTN)
2	Maintenance Special Message (MTNS)
3	Signaling Connection Control Part (SCCP)
4	Telephone User Part (TUP)
5	ISDN User Part (ISUP)
6	Data User Part (call and circuit-related messages)
7	Data User Part (facility registration/cancellation messages)

SIF (Signaling Information Field)

The SIF in an MSU contains the routing label and signaling information (e.g., SCCP, TCAP and ISUP message data). LSSUs and FISUs contain neither a routing label nor an SIO as they are sent between two directly connected signaling points. For more information about routing labels, refer to the description of MTP Level 3 below.

CRC (Cyclic Redundancy Check)

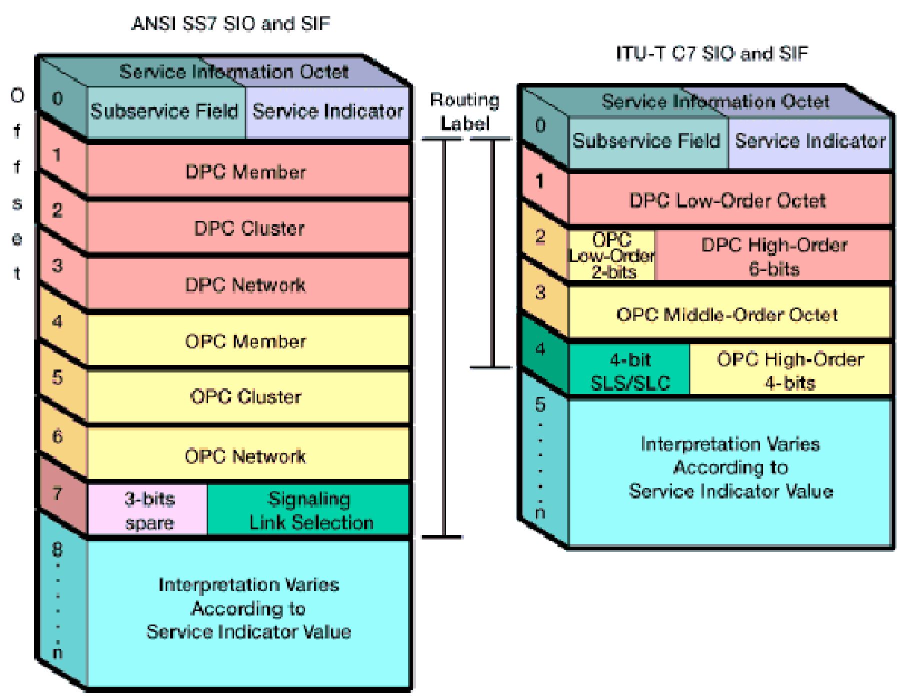
The CRC value is used to detect and correct data transmission errors. For more information, see the description for BIB above.

MTP Level 3

MTP Level 3 provides message routing between signaling points in the SS7 network. MTP Level 3 is equivalent in function to the OSI Network Layer.

MTP Level 3 routes messages based on the routing label in the signaling information field (SIF) of message signal units. The routing label is comprised of the destination point code (DPC), originating point code (OPC) and signaling link selection (SLS) field. Point codes are numeric addresses that uniquely identify each signaling point in the SS7 network. When the destination point code in a message indicates the receiving signaling point, the message is distributed to the appropriate user part (e.g., ISUP or SCCP) indicated by the service indicator in the SIO. Messages destined for other signaling points are transferred provided that the receiving signaling point has message transfer capabilities (like an STP). The selection of outgoing link is based on information in the DPC and SLS.

An ANSI routing label uses seven octets; an ITU-T routing label uses four octets (Figure. 7).



ANSI point codes use 24 bits (three octets); ITU-T point codes typically use 14 bits. For this reason, signaling information exchanged between ANSI and ITU-T networks must be routed through a gateway STP, protocol converter or other signaling point that has both an ANSI and an ITU-T point code. (Note: China uses 24-bit ITU-T point codes, which are incompatible with both ANSI and other ITU-T networks). Interaction between ANSI and ITU-T networks is further complicated by different implementations of higher-level protocols and procedures.

An ANSI point code consists of network, cluster and member octets (e.g., 245-16-0). An octet is an 8-bit byte that can contain any value between zero and 255. Telcos with large networks have unique network identifiers while smaller operators are assigned a unique cluster number within networks one through four (e.g., 1-123-9). Network number zero is not used; network number 255 is reserved for future use.

ITU-T point codes are pure binary numbers, which may be stated in terms of zone, area/network and signaling point identification numbers. For example, the point code 5557 (decimal) may be stated as 2-182-5 (binary 010 10110110 101).

Signaling Link Selection (SLS)

The selection of outgoing link is based on information in the DPC and Signaling Link Selection field. The SLS is used to:

Ensure message sequencing. Any two messages sent with the same SLS will always arrive at the destination in the same order in which they were originally sent. Allow equal load sharing of traffic among all available links. In theory, if a user part sends messages at regular intervals and assigns the SLS values in a round-robin fashion, the traffic level should be equal among all links (within the combined linkset) to that destination.

In ANSI networks, the size of the SLS field was originally five bits (32 values). In configurations with two links in each linkset of a combined linkset (totaling four links), eight SLS values were assigned to each link to allow an equal balance of traffic.

A problem arose when growing networks provisioned linksets beyond four links. With a 5 bit SLS, a configuration with five links in each linkset of a combined linkset (totaling 10 links) results in an uneven assignment of three SLS values for eight links and four SLS values for the remaining two links. To eliminate this problem, both ANSI and Bellcore moved to adopt an 8-bit SLS (256 values) to provide better loadsharing across signaling links.

In ITU-T implementations, the SLS is interpreted as the signaling link code in MTP messages. In ITU-T Telephone User Part message, a portion of the circuit identification code is stored in the SLS field. MTP Level 3 re-routes traffic away from failed links and signaling points and controls traffic when congestion occurs. However, a detailed discussion of this topic is outside the scope of this tutorial. MTP Levels 2 and 1 can be replaced by ATM (Asynchronous Transfer Mode), a simple broadband protocol that uses fixed-length 53 octet cells. MTP Level 3 interfaces to ATM using the Signaling ATM Adaptation Layer (SAAL). This interface is currently an area of ongoing study.

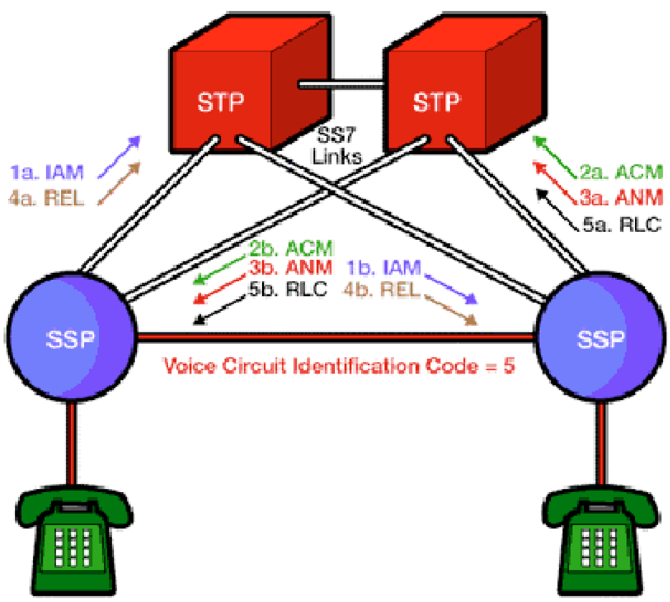
ISDN User Part

The ISDN User Part (ISUP) defines the protocol and procedures used to set-up, manage and release trunk circuits that carry voice and data calls over the public switched telephone network (PSTN). ISUP is used for both ISDN and non-ISDN calls. Calls that originate and terminate at the same switch do not use ISUP signaling.

Basic ISUP Call Control

Figure 8 depicts the ISUP signaling associated with a basic call.

1. When a call is placed to an out-of-switch number, the originating SSP transmits an ISUP **initial address message** (IAM) to reserve an idle trunk circuit from the originating switch to the destination switch (**1a**). The IAM includes the originating point code, destination point code, **circuit identification code** (circuit "5" in Fig. 8), dialed digits and, optionally, the calling party number and name. In the example below, the IAM is routed via the home STP of the originating switch to the destination switch (**1b**). Note that the same signaling link(s) are used for the duration of the call unless a link failure condition forces a switch to use an alternate signaling link.



The destination switch examines the dialed number, determines that it serves the called party and that the line is available for ringing. The destination switch rings the called party line and transmits an ISUP address complete message (ACM) to the originating switch (2a) (via its home STP) to indicate that the remote end of the trunk circuit has been reserved. The STP routes the ACM to the originating switch (2b), which rings the calling party's line and connects it to the trunk to complete the voice circuit from the calling party to the called party.

In the example shown above, the originating and destination switches are directly connected with trunks. If the originating and destination switches are not directly connected with trunks, the originating switch transmits an IAM to reserve a trunk circuit to an intermediate switch. The intermediate switch sends an ACM to acknowledge the circuit reservation request and then transmits an IAM to reserve a trunk circuit to another switch. This process continues until all trunks required to complete the voice circuit from the originating switch to the destination switch are reserved.

3. When the called party picks up the phone, the destination switch terminates the ringing tone and transmits an ISUP answer message (ANM) to the originating switch via its home STP (**3a**). The STP routes the ANM to the originating switch (3b), which verifies that the calling party's line is connected to the reserved trunk and, if so, initiates billing.

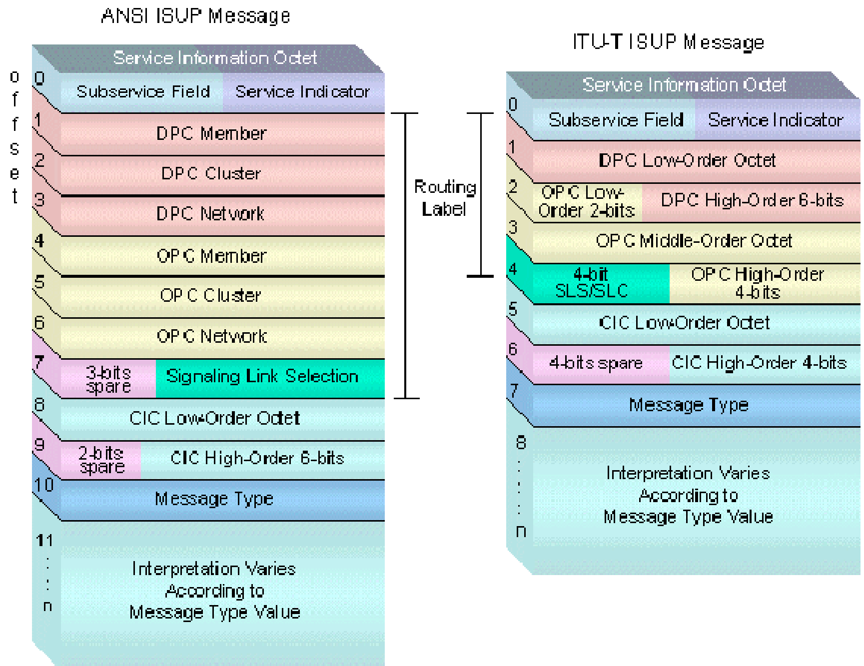
4. If the calling party hangs-up first, the originating switch sends an ISUP release message (REL) to release the trunk circuit between the switches (4a). The STP routes the REL to the destination switch (4b). If the called party hangs up first, or if the line is busy, the destination switch sends an REL to the originating switch indicating the release cause (e.g., normal release or busy).

Upon receiving the REL, the destination switch disconnects the trunk from the called party's line, sets the trunk state to idle and transmits an ISUP release complete message (RLC) to the originating switch (5a) to acknowledge the release of the remote end of the trunk circuit. When the originating switch receives (or generates) the RLC (5b), it terminates the billing cycle and sets the trunk state to idle in preparation for the next call.

ISUP messages may also be transmitted during the connection phase of the call (i.e., between the ISUP Answer (ANM) and Release (REL) messages.

ISUP Message Format

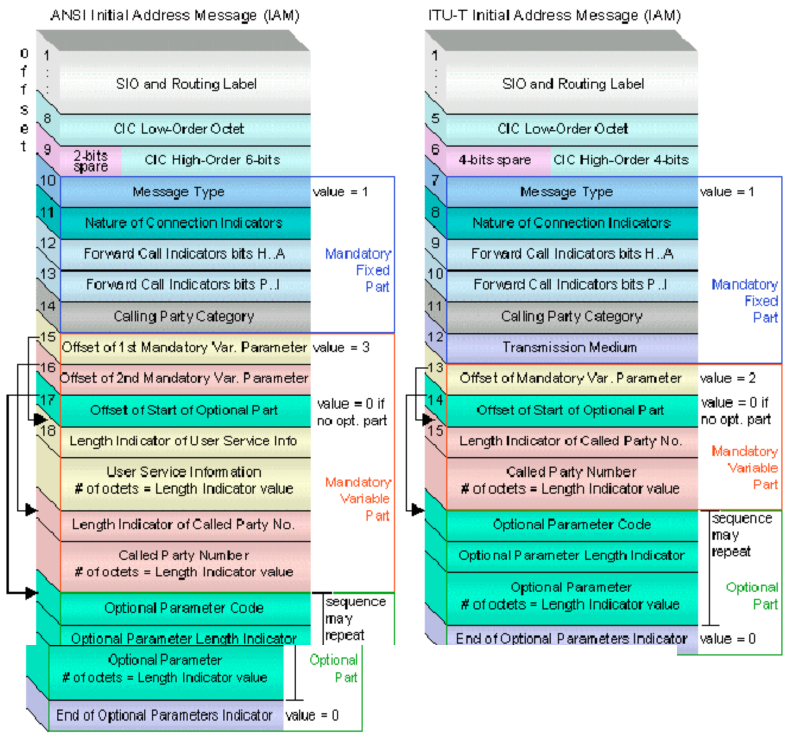
ISUP information is carried in the Signaling Information Field (SIF) of an MSU. The SIF contains the routing label followed by a 14-bit (ANSI) or 12-bit (ITU) circuit identification code (CIC). The CIC indicates the trunk circuit reserved by the originating switch to carry the call. The CIC is followed by the message type field (e.g., IAM, ACM, ANM, REL, RLC), which defines the contents of the remainder of the message.



Each ISUP message contains a mandatory fixed part containing mandatory fixed-length parameters. Sometimes the mandatory fixed part is comprised only of the message type field. The mandatory fixed part may be followed by the mandatory variable part and/or the optional part. The mandatory variable part contains mandatory variable-length parameters. The optional part contains optional parameters, which are identified by a one-octet parameter code followed by a length indicator ("octets to follow") field. Optional parameters may occur in any order. If optional parameters are included, the end of the optional parameters is indicated by an octet containing all zeros.

Initial Address Message

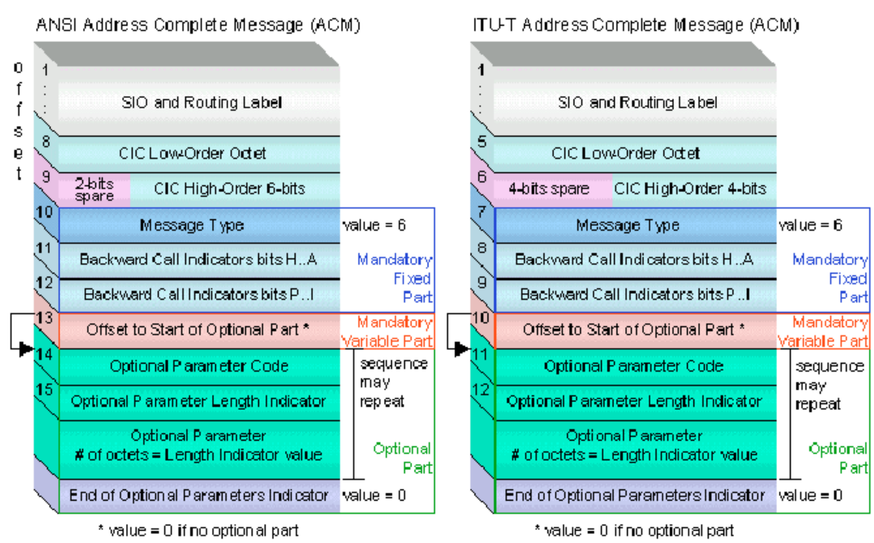
An Initial Address Message (IAM) is sent in the "forward" direction by each switch needed to complete the circuit between the calling party and called party until the circuit connects to the destination switch. An IAM contains the called party number in the mandatory variable part and may contain the calling party name and number in the optional part.



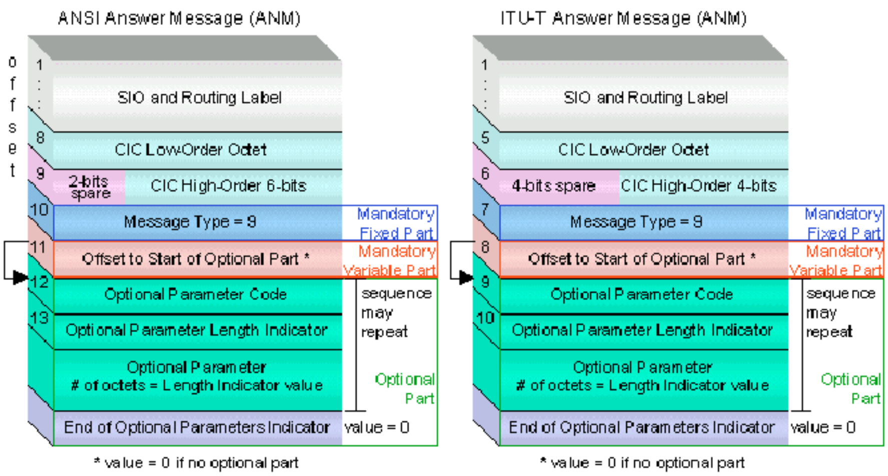
Address Complete Message

An Address Complete Message (ACM) is sent in the "backward" direction to indicate that the remote end of a trunk circuit has been reserved.

The originating switch responds to an ACM message by connecting the calling party's line to the trunk to complete the voice circuit from the calling party to the called party. The originating switch also sends a ringing tone to the calling party's line.

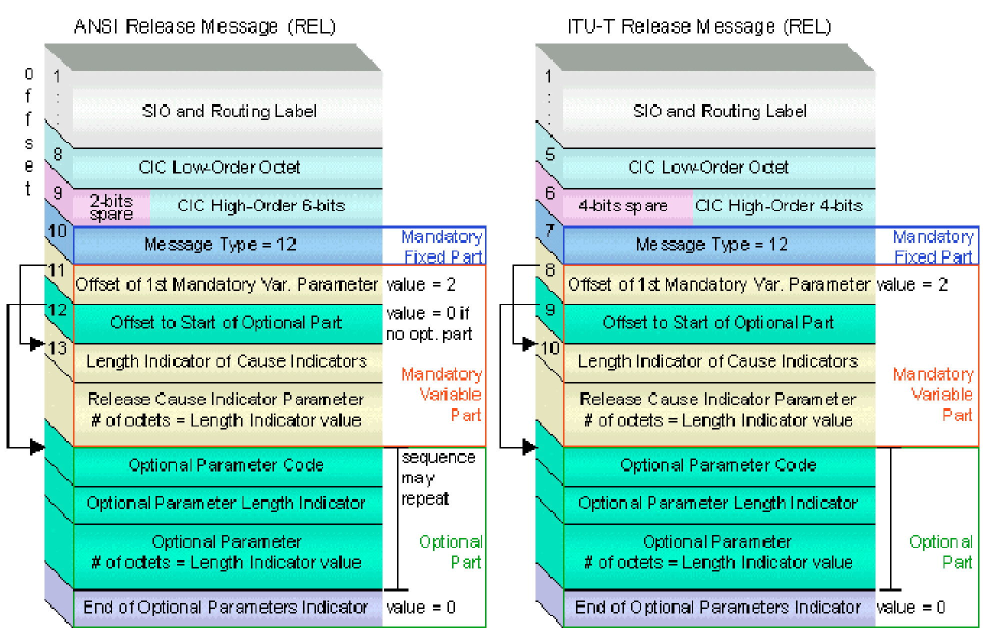


When the called party answers, the destination switch terminates the ringing tone and sends an Answer Message (ANM) to the originating switch. The originating switch initiates billing after verifying that the calling party's line is connected to the reserved trunk.



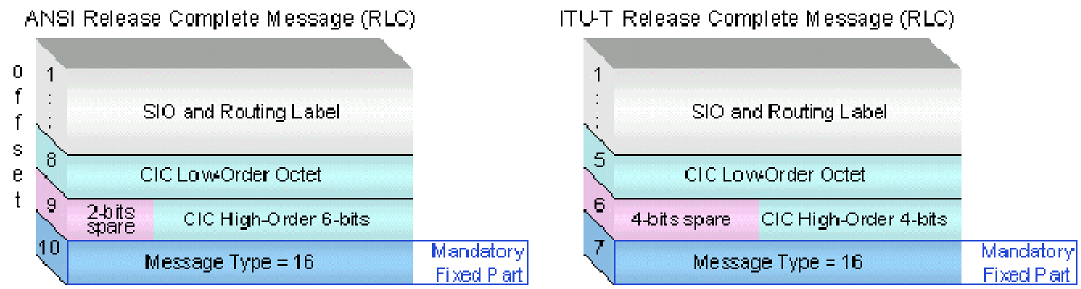
Release Message

A Release Message (REL) is sent in either direction indicating that the circuit is being released due to the **cause indicator** specified. An REL is sent when either the calling or called party "hangs up" the call (cause = 16). An REL is also sent in the backward direction if the called party line is busy (cause = 17).



Release Complete Message

A Release Complete Message (RLC) is sent in the opposite direction of the REL to acknowledge the release of the remote end of a trunk circuit and end the billing cycle as appropriate.



Telephone User Part

In some parts of the world (e.g., China), the Telephone User Part (TUP) supports basic call processing. TUP handles analog circuits only - digital circuits and data transmission capabilities are provided by the Data User Part.

Signaling Connection Control Part

SCCP provides connectionless and connection-oriented network services above MTP Level 3. While MTP Level 3 provides point codes to allow messages to be addressed to specific signaling points, SCCP provides subsystem numbers to allow messages to be addressed to specific applications (called subsystems) at these signaling points. SCCP is used as the transport layer for TCAP-based services such as freephone (800/888), calling card, local number portability, wireless roaming and personal communications services (PCS).

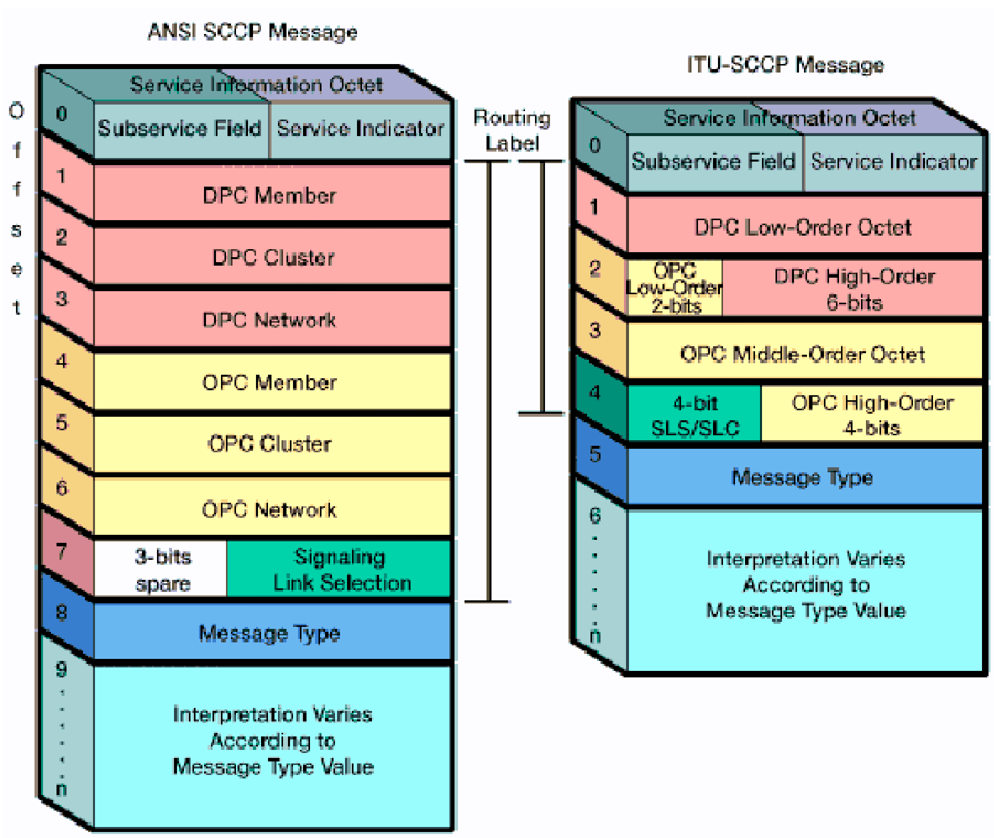
Global Title Translation

SCCP also provides the means by which an STP can perform global title translation (GTT), a procedure by which the destination signaling point and subsystem number (SSN) is determined from digits (i.e., the global title) present in the signaling message.

The global title digits may be any sequence of digits (e.g., the dialed 800/888 number, calling card number or mobile subscriber identification number) pertinent to the service requested. Because an STP provides global title translation, originating signaling points do not need to know the destination point code or subsystem number of the associated service. Only the STPs need to maintain a database of destination point codes and subsystem numbers associated with specific services and possible destinations.

SCCP Message Format

The Service Indicator of the Service Information Octet (SIO) is coded three (binary 0011) for SCCP. SCCP messages are contained within the Signaling Information Field (SIF) of an MSU. The SIF contains the routing label followed by the SCCP message contents. The SCCP message is comprised of a one-octet message type field that defines the contents of the remainder of the message.



Each SCCP message contains a mandatory fixed part (mandatory fixed-length parameters), mandatory variable part (mandatory variable-length parameters) and an optional part that may contain fixed-length and variable-length fields. Each optional part parameter is identified by a one-octet parameter code followed by a length indicator ("octets to follow") field. Optional parameters may occur in any order. If optional parameters are included, the end of the optional parameters is indicated by an octet containing all zeros.

Transaction Capabilities Application Part

TCAP enables the deployment of advanced *intelligent network* services by supporting non-circuit related information exchange between signaling points using the SCCP connectionless service. An SSP uses TCAP to query an SCP to determine the routing number(s) associated with a dialed 800, 888 or 900 number. The SCP uses TCAP to return a response containing the routing number(s) (or an error or reject component) back to the SSP. Calling card calls are also validated using TCAP query and response messages. When a mobile subscriber roams into a new mobile switching center (MSC) area, the integrated visitor location register requests service profile information from the subscriber's home location register (HLR) using mobile application part (MAP) information carried within TCAP messages. TCAP messages are contained within the SCCP portion of an MSU. A TCAP message is comprised of a *transaction portion* and a *component portion*.

Transaction Portion

The transaction portion contains the package type identifier. There are seven package types:

Unidirectional: Transfers component(s) in one direction only (no reply expected).

Query with Permission: Initiates a TCAP transaction (e.g., a 1-800 query). The estination

node may end the transaction.

Query without Permission: Initiates a TCAP transaction. The destination node may *not* end the transaction.

Response: Ends the TCAP transaction. A response to a 1-800 query with permission may

contain the routing number(s) associated with the 800 number.

Conversation with Permission: Continues a TCAP transaction. The destination node may end the transaction.

Conversation without Permission: Continues a TCAP transaction. The destination node may *not* end the transaction.

Abort: Terminates a transaction due to an abnormal situation.

The transaction portion also contains the Originating Transaction ID and Responding Transaction ID fields, which associate the TCAP transaction with a specific application at the originating and destination signaling points respectively.

Component Portion

The component portion contains *components*. There are six kinds of components:

- **Invoke (Last)**: Invokes an operation. For example, a Query with Permission transaction may include an Invoke (Last) component to request SCP translation of a dialed 800 number. The component is the "last" component in the query.
- **Invoke (Not Last)**: Similar to the Invoke (Last) component except that the component is followed by one or more components.
- **Return Result (Last)**: Returns the result of an invoked operation. The component is the "last" component in the response.
- **Return Result (Not Last)**: Similar to the Return Result (Last) component except that the
- component is followed by one or more components.
- **Return Error**: Reports the unsuccessful completion of an invoked operation.
- **Reject**: Indicates that an incorrect package type or component was received.
- Components include parameters**,** which contain application-specific data carried unexamined by TCAP.

Chapter 6 - Mobile Communication

Introduction

Mobile Communication plays a vital role in the present society. It passes several generations and now it is reaching third generation. It was commercially used in 1980's. The summary of history is as follows.

1st Generation Analog Mobile Systems

Year	***System***	***Frequency Band***	***Countries Used***
1981	Advanced Mobile Phone Service [Amps]	850 MHz	USA Russia
1985	Total Access Communication	900 MHz	Europe UK

2nd Generation Digital Mobile Systems

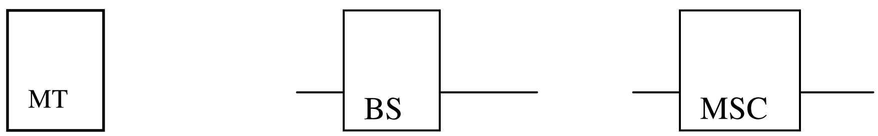
Year	***System***	***Frequency Band***	***Countries Used***
1991	Global System for Mobile Communication [GSM]	900 MHz	Worldwide
1992	Personal Communication Network or DCS 1800	1800 MHz	German UK
1991	Personal Communication Service [PCS]	1900 MHz	USA

Fundamental of Mobile Communication

The following aspects to be considered on mobile communication.

The basic components of mobile communication are,

- Mobile phone [Mobile Station - MS]
- Base Station [BS]
- Mobile Switch [Mobile Switching Centre - MSC]



Mobile Terminal [MT] Base Station [BS] Mobile Switching Centre [MSC]

The MS connects to BS. Since MS is a movable object, the connectivity is done by radio. This is called the Access Network part. Since there is a limited frequency band can be used, for mobile communication the given frequency band for a particular operator should be used effectively.

For example the GSM frequency band is 890 – 960 MHz.
The allocated bandwidth = 960 – 890

= 70 MHz

= 70,000 KHz

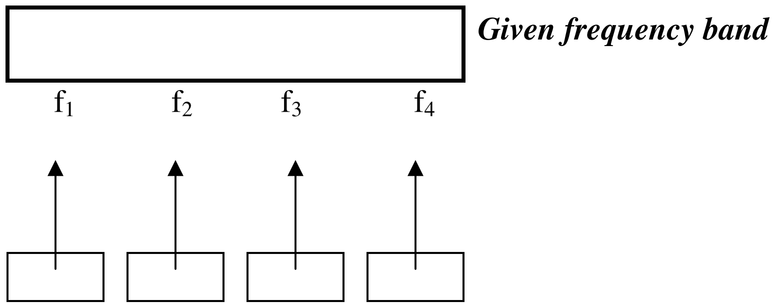
If one channel needs 200 KHz, the number of telephone channel can be used is

= 70,000/200

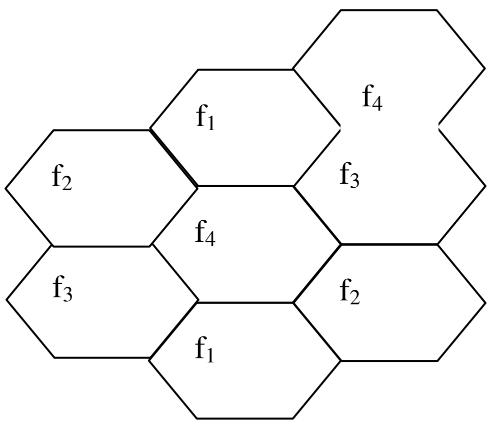
= 350

That means, only 350 people can use mobile pones simultaneously. But we know that there are millions of mobile users in Sri Lanka. There are techniques to re-use the frequencies.

One method is the mobile areas are divided several parts. Each part is called a Cell. Each cell has a Base Station. Each mobile operator is given a particular frequency band by the Telecom Regulatory Commission [TIC]. They divide the given frequency bands to sub bands. Suppose the center frequencies of sub bands are f_1, f_2, f_3 and f_4.



By limiting the transmit power, the same frequency can be used in different areas calls cells.



Each cell has a Base Station.
This method is called the special separation.

[Space wise separation]

If same frequency band is used for 100 cells, the effective number of channels can be increased by 100 times.

Access Technologies in Cells

Suppose a cell frequency band is 10 MHz. The Mobile users [MS] access to the BS by using a radio signal. For this purpose different access technologies can be used. The widely used access technologies are,

Frequency Division Multiple Access	[FDMA]
Time Division Multiple Access	[TDMA]
Code Division Multiple Access	[CDMA]

FDMA

The allocated bandwidth is separated into several small bandwidths.

E.g.	Cell Band width is 10 MHz	=	10,000 KHz
	Per user bandwidth	=	200 KHz
	The number of Users	=	10,000/200
		=	50

Note : All the carriers are continuously transmission carriers.

TDMA

All users use the whole bandwidth. The carriers are transmitted intermittently, not continuously. Each user is given a time slot to transmit his carrier.

Suppose each user is given a 1ms time slot.

The number of possible users is,

1 ms	-----	1 use
1 s	-----	1/1000
	=	1000 users

CDMA

All users use the whole bandwidth. [Same as TDMA]

The carriers are transmitted continuously. [Same as FDMA]

Each users data is mixed with a code before modulation. Each user is given a unique code. At the receiver, after demodulator using the same code can separate the data.

Suppose 1500 codes are used.

The number of users = 1500

Hand over

A major aspect to be considered in Mobile Communication is the continuation of conversation while changing over from one cell to another. This is done by MSC with the coordination of Base Station. The MSC instruct the new BS to which channel to be allocated to the Mobile Terminal.

Control and Voice Channels

There are two types of channels used in Mobile System.

- Voice channels are used to data.
- Control channels are used to send different control signals.

Call Set Up

Calls can be originated from PSTN, another mobile network or same mobile network.

Same mobile network it can be in the MSC area or a different MSC.

A MSC connected to PSTN or another mobile network is called a Gateway MSC [GMSC]

GSM Network Architecture

In GSM the Mobile Station is divided into two parts. Subscriber Identify Module [SIM] and Mobile Station [MS]. SIM has subscriber identify number and subscriber authentication key.

Base Station consists of two parts. Transmit/Receive part [Base Transceiver Station [BTS] and Base Station Controller [BSC]. The BSC is connected to several BTS . The cell handover from one BTS to and BTS is done by BSC.

A number of BSCs are controlled by MSC. MSC is doing the authentication of user and switch to another MSC or PSTN.

Location Updating

Each BTS broadcasts a location code using a dedicated control channels. The Mobile Terminals update the location in the SIM. Always it compares the received location and location in the SIM. If there is a difference the Mobile Terminal transmit a location updated to MSC using another control channel.

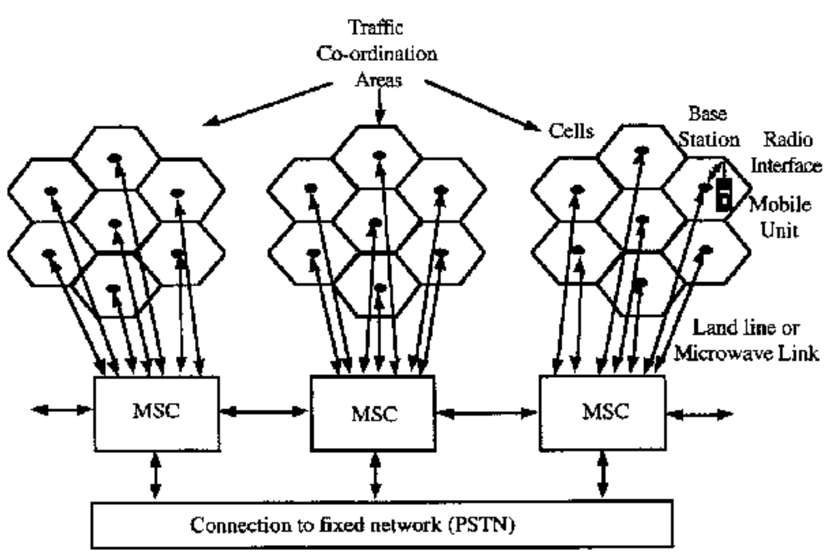
Incoming Call Processing

Consider a call originated from PSTN or any other mobile network. It will reach the Gateway MSC [GMSC]. The GMSC find out the location information and direct the call to relevant MSC. The MSC delivered call to relevant Mobile Terminal via relevant BSC and BTS.

Outgoing Call

The Mobile Station requests a channel through a control channel. The MSC allocates a cannel. The MS dialed the number and MSC will switch the call to required destination.

Analogue Network Archicture



GSM Network Architecture

Introduction

Figure 5-2 shows a schematic of the GSM digital network architecture. Although many elements are essentially the same as for analogue networks, there are key differences which add to the complexity and require a higher degree of distributed intelligence across the network. The key elements are discussed below.

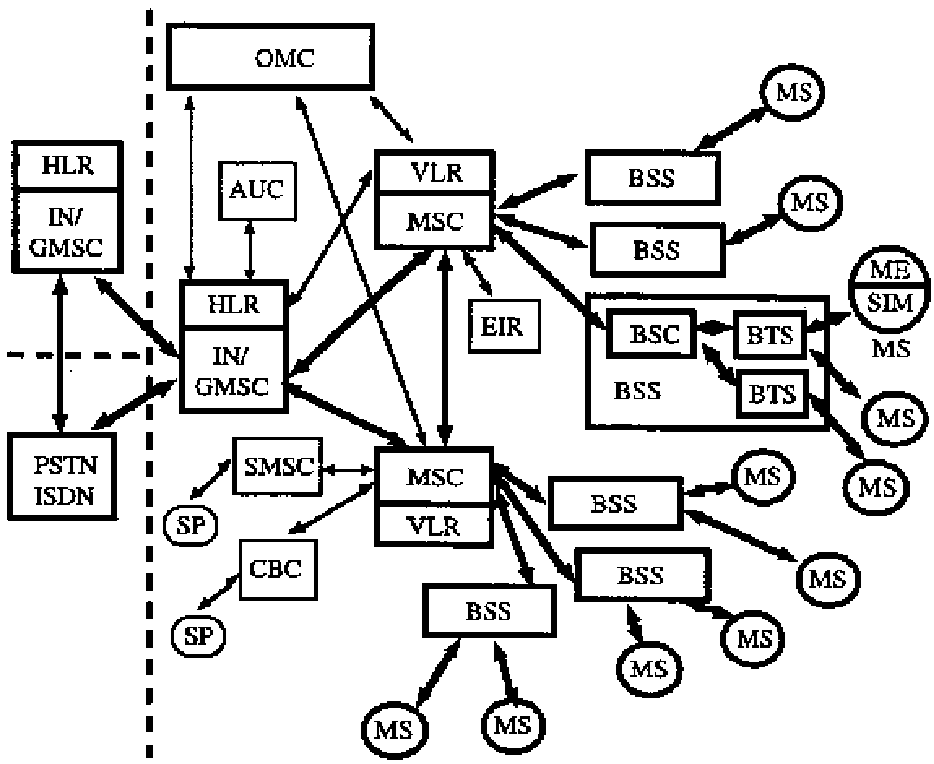


Figure 5-2 GSM network architecture schematic
(Dashed lines denote the boundaries of different networks)

MS	- Mobile Station	ME	- Mobile Equipment
SIM	- Subscriber Identity Module	BSS	- Base Station Subsystem
BSC	- Base Station Controller	BTS	- Base Transceiver Station
MSC	- Mobile Switching Centre	VLR	- Visitor Location Register
HLR	- Home Location Register	AUC	- Authentication Centre
IN/GMSC	- Interrogating Node/Gateway MSC	EIR	- Equipment Identity Register
OMC	- Operations & Management Centre	SP	- Service Provider
SMSC	- Short Message Service Centre	CBC	- Cell Broadcast Centre
PSTN	- Public Switched Telephone Network	ISDN	- Integrated Services Digital Network

Mobile System

The mobile station (MS) is split into the subscriber identity module (SIM) and the mobile equipment (ME). The ME has its own pre-programmed international mobile equipment identity number (IMEI), which can be checked against the equipment identity register (EIR) database to detect stolen, non-type approved or equipment logged as faulty.

The SIM carries the International Mobile Subscriber Identity (IMSI) number which identifies the HLR of the subscriber. It also stores the 'subscriber authentication key' which is used to provide a 'signed response' to an authentication request from the network. This occurs at each registration, location update, network access or additional service request.

Base station subsystem

In GSM and PCN, network intelligence is added at the base station level. The base station subsystem (BSS) is made up of two elements, the base transceiver station and the base station controller. The base transceiver station (BTS), which effectively replaces the analogue base station, comprises the radio transceiving equipment, antennae and associated signal processing to perform the function of a sophisticated radio modem. Hence, it is designed to provide radio coverage for one cell. Many of the functions performed by the MSC in an analogue network in order to maintain reliable radio links, e.g. local handover and power control, are devolved to the base station controller (BSC). The BSC acts as a local traffic concentrator, providing local switching to achieve handover between several BTSs under its control. The transcoding and rate adaptation unit (TRAU) may be located at the BSC or, more commonly, at the MSC. This is responsible for converting between the data rates used over the network and over the radio air interface e.g. speech is carried over the network at 64kbps but is transcoded to 13kbps for radio transmission.

Mobile Service Switching Center

A number of BSCs are controlled by an MSC which is essentially an ISDN switch with enhanced functionality. Its function is to provide call handling for all MSs within its traffic area and the generation of charging records for subsequent forwarding to the billing computer at the operations and management centre (OMC). The MSC is central to the execution of many mobility procedures and supplementary services as detailed below.

The initial target of calls bound for GSM users within a particular network is called the interrogating node (IN). This is usually combined with an MSC to form a gateway MSC (GMSC), responsible for determining the present location of a called subscriber and routing the call to them.

The MSC provides an interface to other networks and thus has interworking functions (IWF) to satisfy particular network and services interconnection requirements. There are separate IWFs for interworking with, for example, the PSTN via modems to support voice and data communication, ISDN and other data networks via rate adaptors, fax transmission via group 3 fax adaptors. Protocol conversion is also carried out and, to compensate for codec delays, echo cancellers can be switched into speech circuits. 90ms speech coding delays are introduced in each direction by the BSS. Echoes generated at the PSTN/ISDN switch interface would be heard by the MS user as an irritating 200ms delayed version of their own voice. The MSC has an echo canceller unit to prevent echoes being transmitted to the MS. (The MS is also equipped with an echo canceller to prevent delayed retransmission of received signals to callers).

The MSC is linked to all other MSCs, the OMC and several intelligent databases (AUC, HLR, VLR, EIR). It is also linked to a short message service centre (SMSC) and may be linked to a cell broadcast centre (CBC) to route short messages from other MSs (or SMS application service providers) or broadcast text messages from service providers (SP) to individual users or cells, respectively.

Location Registers

The **HLR**, as in analogue networks, manages mobile subscribers' records. It performs database and intelligent service control functions, storing system identities (IMSI numbers) and diallable mobile subscriber ISDN numbers (MSISDN) and service subscription details, along with updated location information plus authentication and encryption keys. The HLR for any given MS is usually associated (and co-located) with the interrogating node of its home network, which may be in another country. VLRs are thus essential to localise the mobile/network access (especially for outgoing calls) and prevent unnecessary, potentially long-distance, repeated communication with the HLR. The HLR is interrogated at the first registration with the local MSC and transfers the required information to its VLR for temporary use while the MS remains within its traffic area.

Authentication and Encryption

In order to ensure service provision occurs for authorised users only, another intelligent database, the authentication centre (AUC) is provided within the network, usually co-located with the HLR. The principle for secure access is shown schematically in figure 5-3.

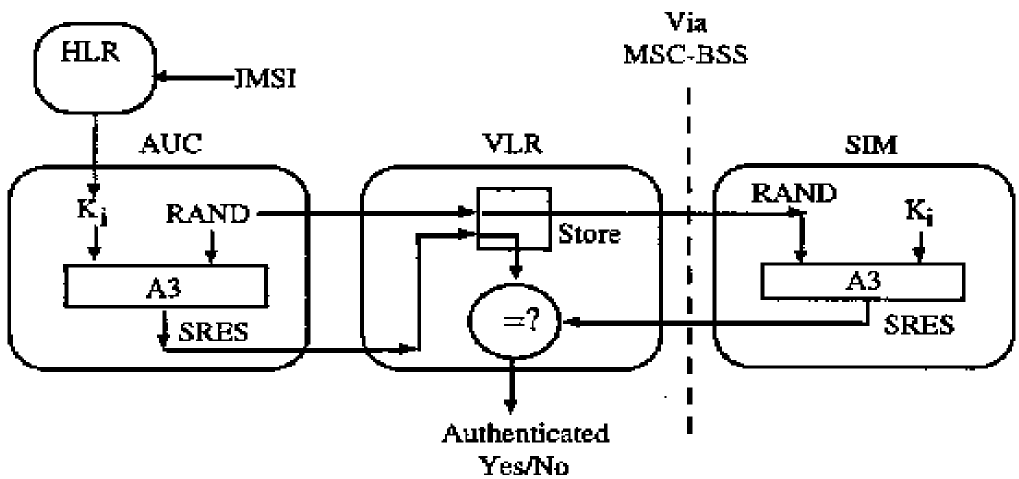


Figure 5-3 Authentication Schematic

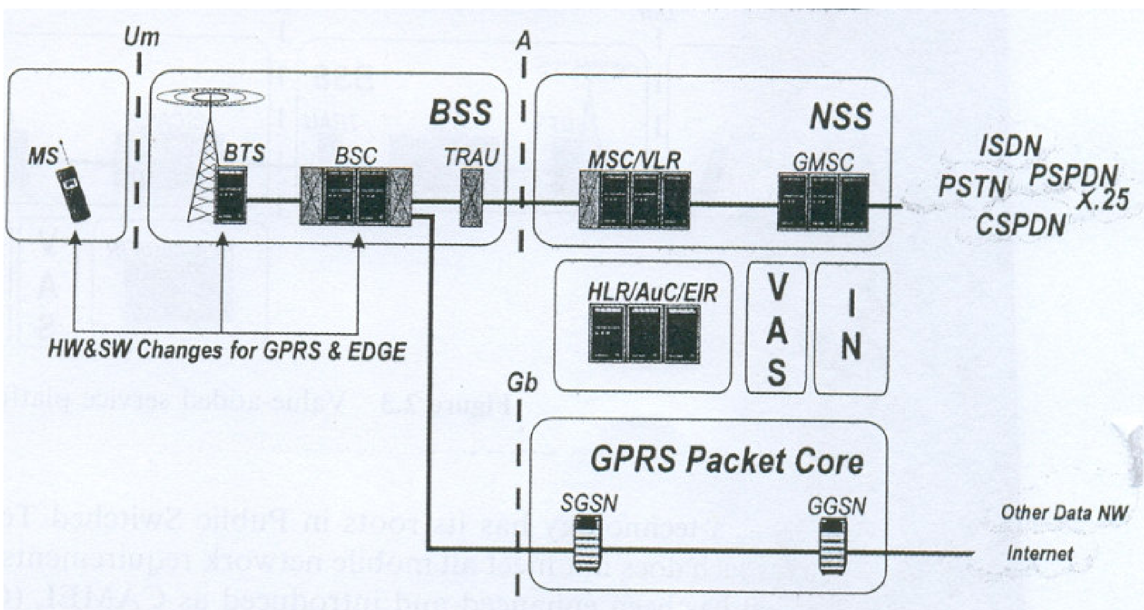
Every subscriber's unique authorisation key (K_i) is stored on the SIM and is available at the AUC (via the HLR). This key is used with a known algorithm (A3) to generate a signed response (SRES) to a challenge random number (RAND) generated by the AUC (based on the TDMA frame number). Only the authentic SIM will be able to generate the correct signed response and thus gain access to the network. The VLR also forwards a ciphering key sequence number (CKSN) to the authenticated MS for subsequent use in ciphering.

Once communication is established, security can be enhanced by transmission encryption. A unique encryption key, K_c, is supplied to the BTS by the VLR and can be generated by the MS from RAND and K_i using another algorithm.(A8). Speech over the air interface can then be encrypted using K_c and a third algorithm (A5). The MS sends the latest CKSN at the beginning of each encrypted transaction. This can be compared to the CKSN stored in the VLR as a simple check to remove the necessity to perform full authentication each time ciphering is used.

Network Management Function

Operations and management functionality for the whole network is provided at the OMC, by management and administrative entities concerned with network operations and maintenance and subscriber management. The latter includes service subscription and billing, while the former encompasses network performance monitoring and optimisation (via modifications to network configuration), plus fault logging and repair scheduling.

GSM Reference Archicture



Mobile Communication

Three generations of Mobile Communication were implemented.

1G - Analogue or Semi Analogue

Analogue Radio + Digital Switch

Established in mid 1980s.

E.g. Nordic Mobile Telephone [NMT]

American Mobile Phone System [AMPS]

Speech Services only.

Developed with National Scope.

2G - Digital

Established in 1990s.

E.g. Global System for Mobile Communication [GSM]

Developed with Regional Scope.

3G - Based on GSM Technology.

Some requirements for 3G were short listed as follows.

1. Specifications should be valid worldwide.
2. Major interfaces should be standardized and open.
3. Multimedia and all of its components must be supported throughout the system.
4. Radio access must provide wideband capacity.
5. The services for end users must be independent of radio access technology.
6. End to end IP based system.

4G - Development specification has been already started.

XG Technology

1G -
- Basic Mobility
- Basic Services
- Incompatibility
- 1980.

2G -
- Advanced Mobility [Roaming]
- More Services [Data Presence]
- Towards Global Solution

3G -
- Seamless Roaming
- Service Concepts & Models
- Global Radio Access
- Global Solution

4G -
- IP based mobility
- Very high data rates
- Complete telecom/datacom convergence

Specification Process for 3G

3G naming policy.

Natural term – Third Generation

Europe - Universal Mobile Telecommunication System [UMTS]

Japan US - IMT 2000 [ITU Standard]

IMT – International Mobile Telephones

US - CDMA 2000 is also an aspect of 3G.

3GPP - Mainly use UMTS Standards.

3G Partnership Project [3GPP]

The 3GPP Organization is a umbrella organization aiming to form compromised standards.

It consists of following standard bodies.
ETSI [European Telecommunication Standard Institute]/Europe.
ARIB [Association of Radio Industries and Business]/Japan
CWTS [China Wireless Telecommunication Standard group]/China
T1 [Standardization Committee T1 – Telecommunications]/US.
TTA [Telecommunication Technology Association]/Korea
TTC [Telecommunication Technology Committee]/Japan

Later this was formed as
3GPP 2 - To accommodate American viewpoint.

Then 3GPP different standard were release as

3GPP Release "Number"

ie. 3GPP R4 and 3GPP R5.

3GPP99, 3GPP00

3G Variants

Variant	**Radio access**	**2G basis**
3G [US]	WCDMA, EDGE, CDMA2000	IS-95, GSM1900, TDMA
3G [Europe]	WCDMA, GSM, EDGE	GSM900/1800
3G [Japan]	WCDMA	PDC

Evoluation from GSM to UMTS

Basic GSM Network

The main idea behind the GSM was to define several open interfaces.

The network is decided to from separate subsystem,

Network Subsystem [NSS]
Base Station Subsystem [BSS]
Network Management Subsystem [NMS]
Mobile Station [MS]

BSS	-	Responsible for radio control path.
NSS	-	Call Control Function.
NMS	-	Operation and Maintenance [O&M] Part of the network.

Mobile Station [MS]

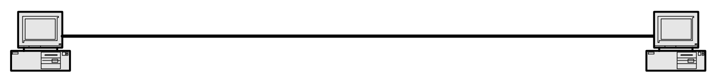
MS	=	ME + SIM
ME	-	Mobile Equipment Radio Part
SIM	-	Service Identify Module Subscriber data.

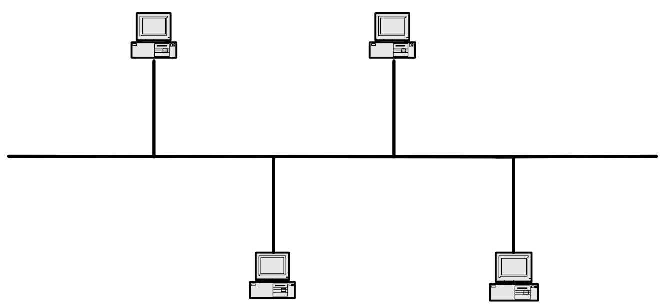
Base Station System [BSS]

BSS	-	BTS + BSC + TRAU
BSC	-	Base Station Controller
TRAU	-	Transcoding and Rate Adoptation Unit

Chapter 7 Data Communication

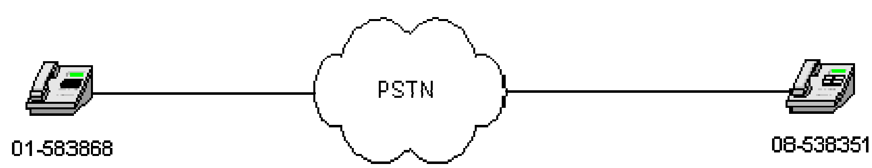
How can we connect two or more computers?
If the computers are located in close proximity, cables can be used.



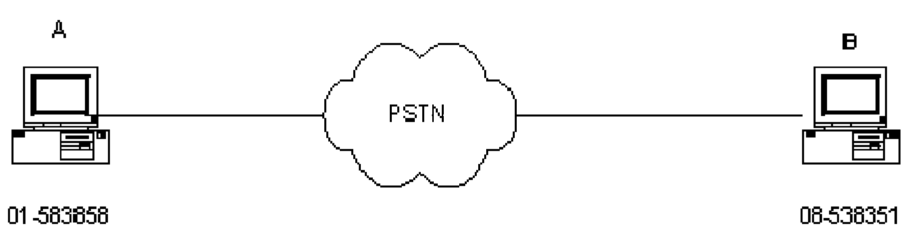


This type of network is called a Local Area Network. [LAN]
If the computers are located very far away a Public Network can be used.

A well-known Public Network is the Telephone Network. Normally this is called a Public Switched Telephone Network. [PSTN]



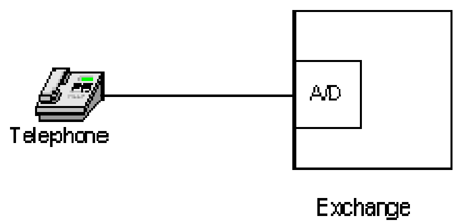
We can remove two telephones and connect two computers instead.



If computer **A** dials the number 08-538351, it can connect to the computer **B.**

This type of connection is called a dial up connection. Also this type of a network is called a Wide Area Network. [WAN]

However, there is a problem in connecting a computer directly to PSTN.

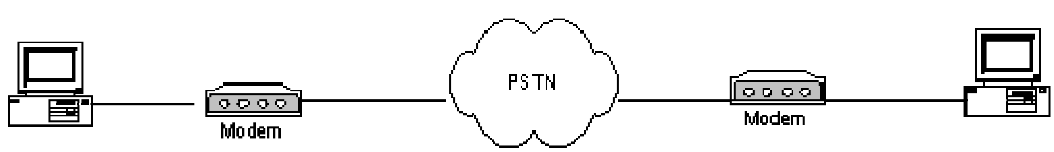


The telephone output signal is an analog signal with a bandwidth of 0.3 – 3.4 KHz.

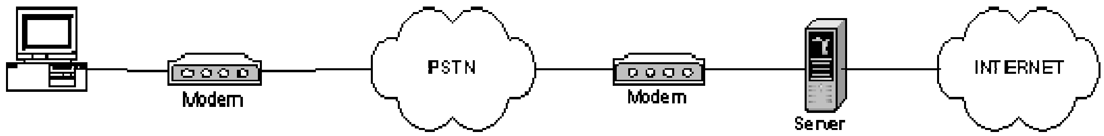
At the telephone exchange this analog signal is converted to a digital signal. [Analog to Digital conversion or A/D Conversion]. For this purpose Pulse Code Modulation [PCM] is used.

Since the exchange expects an analog signal, the computer output digital signal should be converted to an analog signal. For this purpose a Modulator is required. Similarly for the received side of computer, the analog signal sent by the exchange has to be converted to a digital signal. For this purpose a Demodulator is required.

Normally the Modulator and Demodulator comes as one unit and it is called a MODEM.



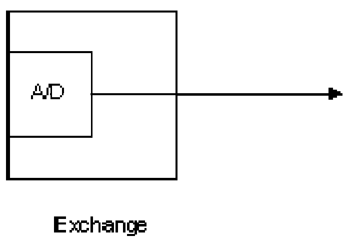
How to connect Internet through PSTN



To connect to the SLT Internet Server, you have to dial the number 150.

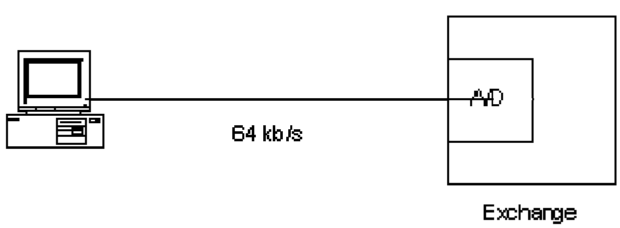
After connecting a Modem to a computer, it should be configured using the appropriate software which comes in a diskette along with the Modem.

The bit rates supported by a normal modem is 4.8 kb/s, 9.6 kb/s, 14.4 kb/s, 19.2 kb/s, 28.8 kb/s, 56 kb/s.



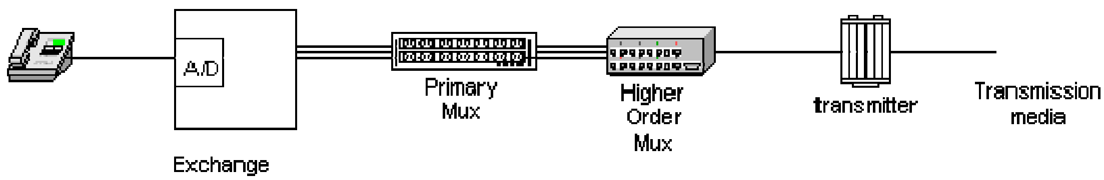
The A/D conversion process at the exchange is PCM.

The PCM output is 64 kb/s.

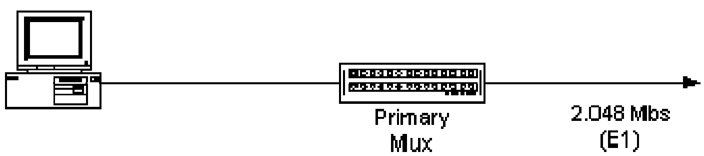


If we bypass the A/D conversion part of exchange, we can send a 64 kb/s digital signal from the computer to the exchange.

Normal telecommunication system is as follows.



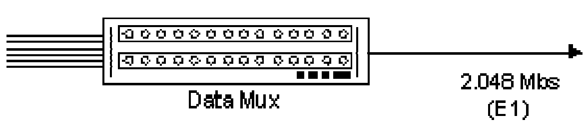
If we by-pass the exchange the computer can be directly connect to the Primary MUX.



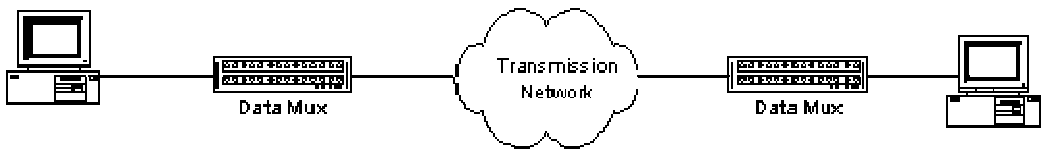
Each input channel of the Primary MUX is 64 kb/s. Therefore the computer can send 64 kb/s digital signal.

However, the Primary MUX does not have the ability to provide service for bit rates of multiples of 64 kb/s. That is 128 kb/s, 192 kb/s, 256 kb/s, 512 kb/s etc.

Therefore for data transmission purposes the Data MUX is used.



The input ports can be configured for different bit rates. That is 128 kb/s, 192 kb/s, 256 kb/s, 512 kb/s etc.

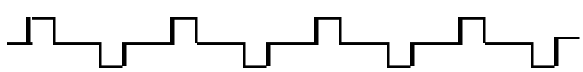


This type of circuit is called a leased circuit.

The computer output signal is a unipolar binary signal.

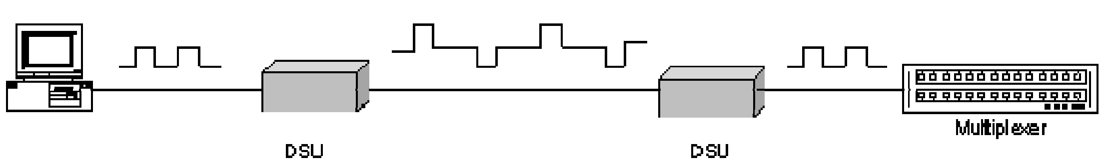
When such a signal travels more than about 10 m it can get attenuated.

In order to send a Computer signal to a long distance, it should be converted to some pattern of a bipolar signal.



This is called a line code.

By using a Digital Service Unit [DSU], line coding can be achieved.



For leased digital circuits we have to use a DSU at the computer and DSU at the Data Mux.

For dial up connections we have to use a Modem at the computer.

General Packet Radio Service

General Packet Radio Service (GPRS) is a Mobile Data Service available to users of Global System for Mobile Communications (GSM) and IS-136 mobile phones. GPRS data transfer is typically charged per megabyte of transferred data, while data communication via traditional circuit switching is billed per minute of connection time, independent of whether the user has actually transferred data or has been in an idle state. GPRS can be used for services such as Wireless Application Protocol (WAP) access, Short Message Service (SMS), Multimedia Messaging Service (MMS), and for Internet communication services such as email and World Wide Web access.

2G cellular systems combined with GPRS is often described as "2.5G", that is, a technology between the second (2G) and third (3G) generations of mobile telephony. It provides moderate speed data transfer, by using unused Time division multiple access (TDMA) channels in for example the GSM system. Originally there was some thought to extend GPRS to cover other standards, but instead those networks are being converted to use the GSM standard, so that GSM is the only kind of network where GPRS is in use. GPRS is integrated into GSM Release 97 and newer releases. It was originally standardized by European Telecommunications Standards Institute (ETSI), but now by the 3rd Generation Partnership Project (3GPP).

Basics

GPRS is packet-switched, which means that multiple users share the same transmission channel, only transmitting when they have data to send. Thus the total available bandwidth can be immediately dedicated to those users who are actually sending at any given moment, providing higher use where users only send or receive data intermittently. Web browsing, receiving e-mails as they arrive and instant messaging are examples of uses that require intermittent data transfers, which benefit from sharing the available bandwidth. By contrast, in the older Circuit Switched Data (CSD) standard included in GSM standards, a connection establishes a circuit, and reserves the full bandwidth of that circuit during the lifetime of the connection.

Usually, GPRS data are billed per kilobyte of information transceived, while circuit-switched data connections are billed per second. The latter is because even when no data are being transferred, the bandwidth is unavailable to other potential users.

The multiple access methods used in GSM with GPRS are based on frequency division duplex (FDD) and FDMA. During a session, a user is assigned to one pair of up-link and down-link frequency channels. This is combined with time domain statistical multiplexing, i.e. packet mode communication, which makes it possible for several users to share the same frequency channel. The packets have constant length, corresponding to a GSM time slot. The down-link uses first-come first-served packet scheduling, while the up-link uses a scheme very similar to reservation ALOHA. This means that slotted Aloha (S-ALOHA) is used for reservation inquiries during a

contention phase, and then the actual data is transferred using dynamic TDMA with first-come first-served scheduling.
GPRS originally supported (in theory) Internet Protocol (IP), Point-to-Point Protocol (PPP) and X.25 connections. The last has been typically used for applications like wireless payment terminals, although it has been removed from the standard. X.25 can still be supported over PPP, or even over IP, but doing this requires either a router to perform encapsulation or intelligence built in to the end-device/terminal e.g. UE(User Equipment). In practice, when the mobile built-in browser is used, IPv4 is being utilized. In this mode PPP is often not supported by the mobile phone operator, while IPv6 is not yet popular. But if the mobile is used as a modem to the connected computer, PPP is used to tunnel IP to the phone. This allows DHCP to assign an IP Address and then the use of IPv4 since IP addresses used by mobile equipment tend to be dynamic.

Capability classes

Class A
Can be connected to GPRS service and GSM service (voice, SMS), using both at the same time. Such devices are known to be available today.

Class B
Can be connected to GPRS service and GSM service (voice, SMS), but using only one or the other at a given time. During GSM service (voice call or SMS), GPRS service is suspended, and then resumed automatically after the GSM service (voice call or SMS) has concluded. Most GPRS mobile devices are Class B.

Class C
Are connected to either GPRS service or GSM service (voice, SMS). Must be switched manually between one or the other service.

A true Class A device may be required to transmit on two different frequencies at the same time, and thus will need two radios. To get around this expensive requirement, a GPRS mobile may implement the dual transfer mode (DTM) feature. A DTM-capable mobile may use simultaneous voice and packet data, with the network coordinating to ensure that it is not required to transmit on two different frequencies at the same time. Such mobiles are considered pseudo-Class A. Some networks are expected to support DTM in 2007.

Multislot classes

GPRS speed is a direct function of the number of TDMA time slots assigned, which is the lesser of (a) what the particular cell supports and (b) the maximum capability of the mobile device expressed as a **GPRS Multislot Class**.

Multislot Class	Downlink Slots	Uplink Slots	Active Slots
1	1	1	2
2	2	1	3
3	2	2	3
4	3	1	4
5	2	2	4
6	3	2	4
7	3	3	4
8	4	1	5
9	3	2	5
10	4	2	5
11	4	3	5
12	4	4	5
32	5	3	6

The most common GPRS multislot classes are:

Class 2

Minimal GPRS implementation

Class 4

Modest GPRS implementation, 50% faster download than Class 2

Class 6

Modest implementation, but with better uploading than Class 4

Class 8

Better implementation, 33% faster download than Classes 4 & 6

Class 10

Better implementation, and with better uploading than Class 8, seen in better cell phones and PC Cards

Class 12

Best implementation, with maximum upload performance, typically seen only in high-end PC Cards

Coding scheme

Transfer speed depends also on the channel encoding used. The least robust, but fastest, coding scheme (CS-4) is available near a base transceiver station (BTS), while the most robust coding scheme (CS-1) is used when the mobile station (MS) is further away from a BTS.

Coding scheme	Speed (kbit/s)
CS-1	8.0
CS-2	12.0
CS-3	14.4
CS-4	20.0

Using the CS-4 it is possible to achieve a user speed of 20.0 kbit/s per time slot. However, using this scheme the cell coverage is 25% of normal. CS-1 can achieve a user speed of only 8.0 kbit/s per time slot, but has 98% of normal coverage. Newer network equipment can adapt the transfer speed automatically depending on the mobile location.
Like CSD, HSCSD establishes a circuit and is usually billed per minute. For an application such as downloading, HSCSD may be preferred, since circuit-switched data are usually given priority over packet-switched data on a mobile network, and there are relatively few seconds when no data are being transferred.

Technology	Download (kbit/s)	Upload (kbit/s)	Configuration
CSD	9.6	9.6	1+1
HSCSD	28.8	14.4	2+1
HSCSD	43.2	14.4	3+1
GPRS	80.0	20.0 (Class 8 & 10 and CS-4)	4+1
GPRS	60.0	40.0 (Class 10 and CS-4)	3+2
EGPRS (EDGE)	236.8	59.2 (Class 8, 10 and MCS-9)	4+1

EGPRS (EDGE)	177.6	118.4 (Class 10 and MCS-9)	3+2

GPRS is packet based. When TCP/IP is used, each phone can have one or more IP addresses allocated. GPRS will store and forward the IP packets to the phone during cell handover (when you move from one cell to another). A radio noise induced pause can be interpreted by TCP as packet loss, and cause a temporary throttling in transmission speed.

Services and hardware

GPRS upgrades GSM data services providing:

- Multimedia Messaging Service (MMS)
- Push to talk over Cellular PoC / PTT
- Instant Messaging and Presence -- Wireless Village
- Internet Applications for Smart Devices through Wireless Application Protocol (WAP)
- Point-to-point (PTP) service: internetworking with the Internet (IP protocols)
- Short Message Service (SMS)
- Future enhancements: flexible to add new functions, such as more capacity, more users, new accesses, new protocols, new radio networks.

USB GPRS modem

USB GPRS modems use a terminal-like interface USB 2.0 and later, data formats V.42bis, and RFC 1144 and external antennas. Modems can be add in cards (for laptop) or external USB devices which are similar in shape and size to a computer mouse.

Enhanced Data Rates for GSM Evolution

Enhanced Data rates for GSM Evolution (**EDGE**) or **Enhanced GPRS** (**EGPRS**), is a digital mobile phone technology that allows it to increase data transmission rate and improve data transmission reliability. Althougth technically a 3G network technology it is generally classified as the unofficial standard 2.75G, due to its slower network speed. EDGE has been introduced into GSM networks around the world since 2003, initially in North America.
It can be used for any packet switched application such as an Internet connection. High-speed data applications such as video services and other multimedia benefit from EGPRS' increased data capacity. EDGE Circuit Switched is a possible future development.
EDGE Evolution continues in Release 7 of the 3GPP standard providing doubled performance e.g. to complement High-Speed Packet Access (HSPA).

Technology

EDGE/EGPRS is implemented as a bolt-on enhancement to 2G and 2.5G GSM and GPRS networks, making it easier for existing GSM carriers to upgrade to it. EDGE/EGPRS is a superset to GPRS and can function on any network with GPRS deployed on it, provided the carrier implements the necessary upgrade.
Although EDGE requires no hardware or software changes to be made in GSM core networks, base stations must be modified. EDGE compatible transceiver units must be installed and the base station subsystem (BSS) needs to be upgraded to support EDGE. New mobile terminal hardware and software is also required to decode/encode the new modulation and coding schemes and carry the higher user data rates to implement new services.

Transmission techniques

In addition to Gaussian minimum-shift keying (GMSK), EDGE uses 8 phase shift keying (8PSK) for the upper five of its nine modulation and coding schemes. EDGE produces a 3-bit word for every change in carrier phase. This effectively triples the gross data rate offered by GSM. EDGE, like GPRS, uses a rate adaptation algorithm that adapts the modulation and coding scheme (MCS) according to the quality of the radio channel, and thus the bit rate and robustness of data transmission. It introduces a new technology not found in GPRS, Incremental Redundancy, which, instead of retransmitting disturbed packets, sends more redundancy information to be combined in the receiver. This increases the probability of correct decoding.
EDGE can carry data speeds up to 236.8 kbit/s for 4 timeslots (theoretical maximum is 473.6 kbit/s for 8 timeslots) in packet mode and will therefore meet the International Telecommunications Union's requirement for a 3G network, and has been accepted by the ITU as part of the IMT-2000 family of 3G standards. It also enhances the circuit data mode called HSCSD, increasing the data rate of this service.

EGPRS modulation and coding scheme (MCS)

Coding and modulation scheme (MCS)	Speed (kbit/s/slot)	Modulation
MCS-1	8.8	GMSK
MCS-2	11.2	GMSK
MCS-3	14.8	GMSK
MCS-4	17.6	GMSK
MCS-5	22.4	8-PSK

MCS-6	29.6	8-PSK
MCS-7	44.8	8-PSK
MCS-8	54.4	8-PSK
MCS-9	59.2	8-PSK

Classification

Whether EDGE is 2G or 3G depends on implementation. While Class 3 and below EDGE devices clearly are not 3G, class 4 and above devices perform at a higher bandwidth than other technologies conventionally considered as 2G (such as 1xRTT). Because of the variability, EDGE is generally classified as 2.75G network technology.

EDGE Evolution

EDGE Evolution improves on EDGE in a number of ways. Latencies are reduced by lowering the Transmission Time Interval by half (from 20 ms to 10 ms). Bit rates are increased up to 1 MBit/s peak speed and latencies down to 100 ms using dual carriers, higher symbol rate and higher-order modulation (32QAM and 16QAM instead of 8-PSK), and turbo codes to improve error correction. And finally signal quality is improved using dual antennas. An EDGE Evolution terminal or network can support some of these improvements, or roll them out in stages.

Networks

EDGE is actively supported by GSM operators in North America. Some GSM operators elsewhere view UMTS as the ultimate upgrade path and either plan to skip EDGE altogether or use it outside the UMTS coverage area. However, the high cost and slow uptake of UMTS have resulted in fairly common support for EDGE in the global GSM/GPRS market.

Evolution-Data Optimized

Evolution-Data Optimized or Evolution-Data only, abbreviated as **EV-DO** or **EVDO** and often **EV**, is a telecommunications standard for the wireless transmission of data through radio signals, typically for broadband Internet access. It is classified as a broadband technology, because it uses a broad band of radio frequencies. It employs multiplexing techniques such as CDMA (Code division multiple access) as well as Frequency division duplex (FDD) to maximize the amount of data transmitted. It is standardized by 3rd Generation Partnership

Project 2 (3GPP2) as part of the CDMA2000 family of standards and has been adopted by many mobile phone service providers around the world – particularly those previously employing CDMA networks (as opposed to GSM networks).
The EV-DO feature of CDMA2000 networks is significantly faster than the Enhanced Data Rates for GSM Evolution (EDGE) used by GSM networks. It provides access to mobile devices with air interface speeds of up to 2.4 Mbit/s with Rev. 0 and up to 3.1 Mbit/s with Rev. A. High-Speed Downlink Packet Access (HSDPA), a competing technology for Wideband Code Division Multiple Access (W-CDMA), along with the new Qualcomm Rev A modems have the ability to maintain both circuit switched voice and packet data calls from the same radio; this functionality is not available in Qualcomm Rev. 0 chipsets. It provides an IP based network.
There have been several revisions of the standard, named alphabetically starting with the first as **Rev. A** ("revision A") while the first standard is referred to simply as **Rev. 0**.

Rev 0

The initial design of EV-DO was developed by Qualcomm in 1999 to meet IMT-2000 requirements for a greater-than-2-Mbit/s downlink for stationary communications (as opposed to mobile communication such as a moving cellular phone). Initially, the standard was called High Data Rate (HDR), but was renamed to 1xEV-DO after it was ratified by the International Telecommunication Union (ITU); it was given the numerical designation **IS-856**. Originally, 1xEV-DO stood for "1x Evolution-Data Only", referring to its being a direct evolution of the 1x (1xRTT) air interface standard, with its channels carrying only data traffic. (The title of the 1xEV-DO standard document is "cdma2000 High Rate Packet Data Air Interface Specification", as cdma2000 (lowercase) is another name for the 1x standard, numerically designated as IS-2000.)
Later, likely due to the possible negative connotations of the word "only", the "DO" part of the standard's name 1xEV-DO was changed to stand for "Data Optimized". So EV-DO now stands for "Evolution-Data Optimized", the 1x prefix has been dropped by the many major carriers, and is marketed simply as EV-DO. [5] This provides a more marketing-friendly emphasis that the technology was optimized for data transfers.

Rev A

Rev. A offers fast packet establishment on both the forward and reverse links along with air interface enhancements that reduce latency and improve data rates. In addition to the increase in the maximum burst downlink rate from 2.45 Mbit/s to 3.1 Mbit/s, Rev. A has a significant improvement in the maximum uplink data rate, from 153 kbit/s to a maximum uplink burst rate of 1.8 Mbit/s. This improvement assumes early acknowledgement of the first subpacket, typical data rates therefore average below 1 Mbit/s.

EV-DO Rev A has been commercially deployed in New Zealand by Telecom New Zealand, is currently being commercially deployed in Japan by KDDI, in the United States by Sprint Nextel and Verizon Wireless and in Canada by Bell Mobility and Telus Mobility.

Rev B

EV-DO Rev B is the progressive evolution of the Rev A specification. It maintains the apabilities of EVDO Rev A, and provides the following enhancements: Higher rates per carrier (up to 4.9 Mbit/s on the downlink per carrier). Typical deployments are expected to include 3 carriers for a peak rate of 14.7 Mbit/s Higher rates by bundling multiple channels together enhances user experience and enables new services such as high definition video streaming. Use statistical multiplexing across channels to further reduce latency, enhancing the experience for latency-sensitive services such as gaming, video telephony, remote console sessions and web browsing.
Increased talk-time and standby time Hybrid frequency re-use which reduces the interference from the adjacent sectors and improves the rates that can be offered, especially to users at the edge of the cell.

Efficient support for services that have asymmetric download and upload requirements (i.e. different data rates required in each direction) such as file transfers, web browsing, and broadband multimedia content delivery.

Potential competing standards

Motorola proposed a new system called 1Xtreme as an evolution of CDMA2000, but it was rejected by 3GPP2 standardization body. Later, a competing standard called EV-DV developed by Qualcomm, Lucent, Nokia, Motorola, etc. in 3GPP2 was proposed as an alternate evolution of CDMA. EV-DV stands for Evolution-Data and Voice, since the channel structure was backwards compatible with IS-95 and IS-2000 (1xRTT), allowing an in-band network deployment. (EV-DO requires an overlay network when deployed in mixed mode.)
At the time, there was much debate as to the favorability of DV and DO. Traditional operators with an existing voice network preferred deploying DV, since it does not require an overlay. Other design engineers, and newer operators without a 1x voice network, preferred EV-DO because it did not have to be backward compatible, and so could explore different pilot structures, reverse link silence periods, improved control channels, etc. And, the network cost was lower, since EV-DO uses an IP network and does not require a SS7 network and complex network switches such as a mobile switching center (MSC). Also, equipment was not available for EV-DV in time to meet market demands whereas the EV-DO equipment and mobile application-specific integrated circuits (ASIC) were available and tested by the time the EV-DV standard was completed. As a result, the EV-DV standard was less attractive to operators, and has not been implemented. Verizon Wireless, then Sprint Nextel in 2004 and smaller operators in 2005 announced their plans to deploy EV-DO. So in March 2005, Qualcomm suspended development of EV-DV chipsets, and focused on improving the EV-DO product line.
Several network operators are transitioning their customers to HSDPA networks. In Australia, Telstra has announced the closure of its EV-DO network and is moving customers to its faster HSDPA network. In South Korea, KTF and SK Telecom have stopped investing in their

CDMA2000 networks and in early 2007 began transitioning customers to their new HSDPA networks.Producers interest in CDMA is decreasing too. Nokia stated its decision to pull out of CDMA R&D, with the intention to continue CDMA business in selected markets

High-Speed Downlink Packet Access

High-Speed Downlink Packet Access (**HSDPA**, also known as **High-Speed Downlink Protocol Access**) is a 3G (third generation) mobile telephony communications protocol in the High-Speed Packet Access (HSPA) family, which allows networks based on Universal Mobile Telecommunications System (UMTS) to have higher data transfer speeds and capacity. Current HSDPA deployments support down-link speeds of 1.8, 3.6, 7.2 and 14.4 Mbit/s, and can provide each customer with 30 gigabytes of data per month.[1] Further speed increases are planned for the near future. The networks are then to be upgraded to Evolved HSPA, which provides speeds of 42 Mbit downlink in its first release

Technology

The High-Speed Downlink Shared Channel (HS-DSCH) lacks two basic features of other W-CDMA channels — variable spreading factor and fast power control. Instead, it delivers the improved downlink performance using adaptive modulation and coding (AMC), fast packet scheduling at the base station, and fast retransmissions from the base station, known as hybrid automatic repeat-request (HARQ).

Hybrid automatic repeat-request (HARQ)

HARQ uses incremental redundancy, where user data is transmitted multiple times using different codings. When a corrupted packet is received, the user device saves it and later combines it with the retransmissions, to recover the error-free packet as efficiently as possible. Even if the retransmitted packets are corrupted, their combination can yield an error-free packet.

Fast packet scheduling

The HS-DSCH downlink channel is shared between users using channel-dependent scheduling to make the best use of available radio conditions. Each user device periodically transmits an indication of the downlink signal quality, as often as 500 times per second. Using this information from all devices, the base station decides which users will be sent data on the next 2 ms frame and how much data should be sent for each user. More data can be sent to users which report high downlink signal quality.
The amount of the channelisation code tree, and thus network bandwidth, allocated to HSDPA users is determined by the network. The allocation is "semi-static" in that it can be modified while the network is operating, but not on a frame-by-frame basis. This allocation represents a trade-off between bandwidth allocated for HSDPA users, versus that for voice and non-HSDPA

data users. The allocation is in units of channelisation codes for Spreading Factor 16, of which 16 exist and up to 15 can be allocated to HSDPA.
When the base station decides which users will receive data on the next frame, it also decides which channelisation codes will be used for each user. This information is sent to the user devices over one or more HSDPA "scheduling channels"; these channels are not part of the HSDPA allocation previously mentioned, but are allocated separately. Thus, for a given 2 ms frame, data may be sent to a number of users simultaneously, using different channelisation codes. The maximum number of users to receive data on a given 2 ms frame is determined by the number of allocated channelisation codes. By contrast, in CDMA2000 1xEV-DO, data is sent to only one user at a time.

Adaptive modulation and coding

The modulation scheme and coding is changed on a per-user basis depending on signal quality and cell usage. The initial scheme is Quadrature phase-shift keying (QPSK), but in good radio conditions 16QAM modulation almost doubles data throughput rates. With 5 Code allocation, QPSK typically offers up to 1.8 Mbit/s peak data rates, while 16QAM up to 3.6. Additional codes (e.g. 10, 15) can also be used to improve these data rates or extend the network capacity throughput significantly. Theoretically, HSDPA can give throughput up to 10.8 Mbit/s.

Other improvements

HSDPA is part of the UMTS standards since release 5, which also accompanies an improvement on the uplink providing a new bearer of 384 kbit/s. The previous maximum bearer was 128 kbit/s.
As well as improving data rates, HSDPA also reduces latency and so the round trip time for applications.
Along with the HS-DSCH channel, three new physical channels are also introduced: HS-SCCH, HS-DPCCH and HS-PDSCH. The High Speed-Shared Control Channel (HS-SCCH) informs the user that data will be sent on the HS-DSCH 2 slots ahead. The Uplink High Speed-Dedicated Physical Control Channel (HS-DPCCH) carries acknowledgment information and current channel quality indicator (CQI) of the user. This value is then used by the base station to calculate how much data to send to the user devices on the next transmission. The High Speed-Physical Downlink Shared Channel (HS-PDSCH) is the channel mapped to the above HS-DSCH transport channel that carries actual user data.

HSDPA UE categories

HSDPA comprises various versions with different data speeds.

Category	Max. number of	Modulation	Max. data rate

	HS-DSCH codes		**[Mbit/s]**
1	5	QPSK and 16-QAM	1.2
2	5	QPSK and 16-QAM	1.2
3	5	QPSK and 16-QAM	1.8
4	5	QPSK and 16-QAM	1.8
5	5	QPSK and 16-QAM	3.6
6	5	QPSK and 16-QAM	3.6
7	10	QPSK and 16-QAM	7.3
8	10	QPSK and 16-QAM	7.3
9	15	QPSK and 16-QAM	10.2
10	15	QPSK and 16-QAM	14.4
11	5	QPSK only	0.9
12	5	QPSK only	1.8

Roadmap

The first phase of HSDPA has been specified in the 3rd Generation Partnership Project (3GPP) release 5. Phase one introduces new basic functions and is aimed to achieve peak data rates of 14.4 Mbit/s (see above). Newly introduced are the High Speed Downlink Shared Channels (HS-DSCH), the adaptive modulation QPSK and 16QAM and the High Speed Medium Access protocol (MAC-hs) in base station.
The second phase of HSDPA is specified in the upcoming 3GPP release 7 and has been named HSPA Evolved. It can achieve data rates of up to 42 Mbit/s.[1] It will introduce antenna array technologies such as beamforming and Multiple-input multiple-output communications (MIMO). Beam forming focuses the transmitted power of an antenna in a beam towards the user's direction. MIMO uses multiple antennas at the sending and receiving side. Deployments are scheduled to begin in the second half of 2008.
After HSDPA the roadmap leads to HSOPA, a technology under development for specification in 3GPP Release 8. This project is called the Long Term Evolution initiative. It aims to achieve data rates of up 200 Mbit/s for downlink and 100 Mbit/s for uplink using OFDMA modulation.

Adoption

As of May 25 2007, 102 HSDPA networks have commercially launched mobile broadband services in 55 countries. Nearly 40 HSDPA networks support 3.6 Mbit/s peak downlink data throughput. A growing number are delivering 7.2 Mbit/s peak data downlink, leveraging new higher-speed devices coming into the market. One network has been declared as "14.4 Mbit/s (peak) ready" and several others will have this capability by end 2007. The first commercial HSUPA uplink network is launched, with several more set to follow in 2007.
This protocol is a relatively simple upgrade where UMTS is already deployed.

Code division multiple access-Evolution-Data Optimized (CDMA-EVDO) networks had the early lead on performance, and Japanese providers were highly successful benchmarks for it. But lately this seems to be changing in favour of HSDPA as an increasing number of providers worldwide are adopting it. In Australia, Telstra announced that its CDMA-EVDO network would be replaced with a HSDPA network. Rogers Wireless deployed HSDPA system 850/1900 in Canada on April 1, 2007.
So far, 171 device models from 47 suppliers have been launched, comprising: 53 handsets, 35 notebooks, 30 datacards, 19 wireless routers, 15 modems, 11 embedded module, 2 wireless modules, 1 wireless residential gateway, 1 media player, 1 camera, 1 GPS handset, 1 convergence platform & 1 baseband processor. For details, see www.gsmworld.com/HSPA.

High-Speed Uplink Packet Access

High-Speed Uplink Packet Access (HSUPA) is a 3G mobile telephony protocol in the HSPA family with up-link speeds up to 5.76 Mbit/s.

The specifications for HSUPA are included in Universal Mobile Telecommunications System Release 6 standard published by 3GPP.

Technology

3GPP definition of HSUPA - "The technical purpose of the Enhanced Uplink feature is to improve the performance of uplink dedicated transport channels, i.e. to increase capacity and throughput and reduce delay"..
HSUPA is expected to use an uplink *enhanced dedicated channel* (E-DCH) on which it will employ link adaptation methods similar to those employed by HSDPA, namely:

- shorter Transmission Time Interval enabling faster link adaptation;
- HARQ (hybrid ARQ) with incremental redundancy making retransmissions more effective.

Similarly to HSDPA, HSUPA uses a *packet scheduler*, but it operates on a *request-grant* principle where the UEs request a permission to send data and the scheduler decides when and how many UEs will be allowed to do so. A request for transmission contains data about the state of the transmission buffer and the queue at the UE and its available power margin.
In addition to this *scheduled* mode of transmission the standards also allows a self-initiated transmission mode from the UEs, denoted *non-scheduled*. The *non-scheduled* mode can, for example, be used for VoIP services for which even the reduced TTI and the Node-B based scheduler will not be able to provide the very short delay time and constant bandwidth required.
Each MAC-d flow (i.e. QoS flow) is configured to use either *scheduled* or *non-scheduled* modes; the UE adjusts the data rate for *scheduled* and *non-scheduled* flows independently. The maximum data rate of each *non-scheduled* flow is configured at call setup, and typically not changed frequently. The power used by the *scheduled* flows is controlled dynamically by the Node-B through absolute grant (consisting of an actual value) and relative grant (consisting of a single up/down bit) messages.
At Layer 1, HSUPA introduces new physical channels E-AGCH (Access Grant Channel), E-DCH Relative Grant Channel,F-DPCH (Fractional-DPCH),E-HICH (E-DCH Hybrid ARQ Indicator Channel), E-DPCCH (E-DCH Dedicated Physical Control Channel) and E-DPDCH (E-DCH Dedicated Physical Data Channel). E-DPDCH is used to carry the E-DCH Transport Channel. E-DPCCH is used to carry the control information associated with the E-DCH.

Versions

The following table gives uplink speeds for the different categories of HSUPA

HSUPA Category	Max Uplink Speed
Category 1	0.73 Mbit/s

Category 2	1.46 Mbit/s
Category 3	1.46 Mbit/s
Category 4	2.93 Mbit/s
Category 5	2.00 Mbit/s
Category 6	5.76 Mbit/s

Roadmap

After HSUPA the 3GPP is working on further advancing transfer rates. The HSOPA will provide up to 100 Mbit/s for downlink and 50 Mbit/s for uplink.

Chapter – 8 (IP Based Network)

Introduction

Modern digital technology allows different sectors, e.g. telecom, data, radio and television, to be merged together. This occurrence, commonly known as convergence, is happening on a global scale and is drastically changing the way in which both people and devices communicate. At the center of this process, forming the backbone and making convergence possible, are IP-based networks.

Services and integrated consumer devices for purposes such as telephony, entertainment, security or personal computing are constantly being developed, designed and converged towards a communication standard that is independent from the underlying physical connection. The cable network, for instance, first designed for transmitting television to the consumer, can now also be utilized for sending e-mail, surfing the Web or even monitoring a network camera sending live pictures from another continent. Furthermore, these features are also available over other physical networks, e.g. telephone, mobile phone, satellite and computer networks.

This white paper introduces the central components of IP-based network technology, and in doing so it will demonstrate the tremendous benefits this new technology has to offer.

Basics in network communication

The Internet has become the most powerful factor guiding the ongoing convergence process. This is mainly due to the fact that the *Internet protocol suite* has become a shared standard used with almost any service. The Internet protocol suite consists primarily of the *Internet Protocol (IP)* and the *Transport Control Protocol (TCP)*; consequently, the term *TCP/IP* commonly refers to the whole protocol family.

IP-based networks are of great importance in today's information society. At first glance, this technology might appear a bit confusing and overwhelming. Therefore, we'll start by presenting the underlying network components upon which this technology is built.

A network is comprised of two fundamental parts, the nodes and the links. A node is some type of network device, such as a computer. Nodes are able to communicate with other nodes through links, like cables. There are basically two different network techniques for establishing communication between nodes on a network: the *circuit-switched network* and the *packet-*

switched network techniques. The former is used in a traditional telephone system, while the latter is used in IP-based networks.

A circuit-switched network creates a closed circuit between two nodes in the network to establish a connection. The established connection is thus dedicated to the communication between the two nodes. One of the immediate problems with dedicated circuits is wasted capacity, since almost no transmission uses the circuit 100 percent of the time. Also, if a circuit fails in the middle of a transmission, the entire connection must be dropped and a new one established. For illustration purposes, take a look at a telephone connection over a circuit-switched network .

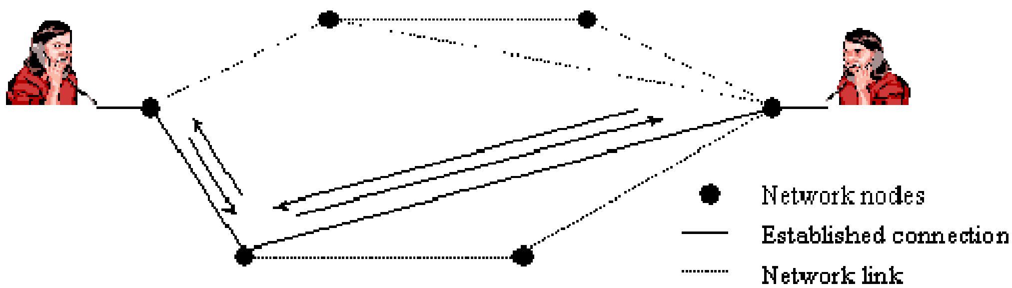


Figure 1: A circuit-switched network utilizes a dedicated closed circuit

IP-based networks on the other hand utilize a packet-switched network technology, which uses available capacity much more efficiently and minimizes the risk of possible problems, such as a disconnection. Messages sent over a packet-switched network are first divided into packets containing the destination address. Then, each packet is sent over the network with every intermediate node and router in the network determining where the packet goes next. A packet does not need to be routed over the same links as previous related packets. Thus, packets sent between two network devices can be transmitted over different routes in the event of a link breakdown or node malfunction .

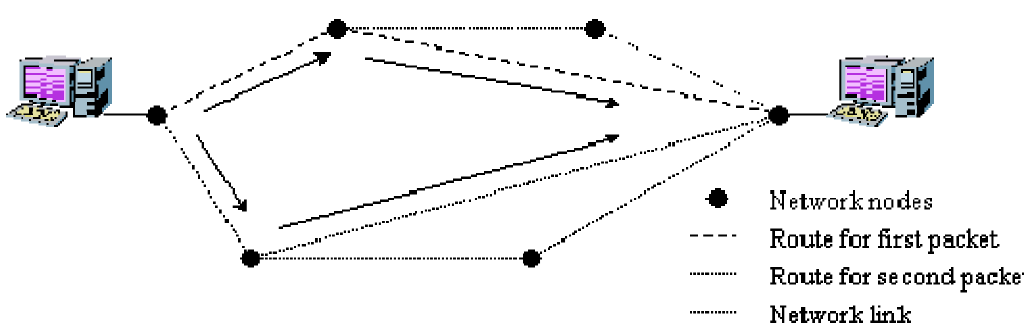


Figure 2: *A packet-switched network routes each packet independently*

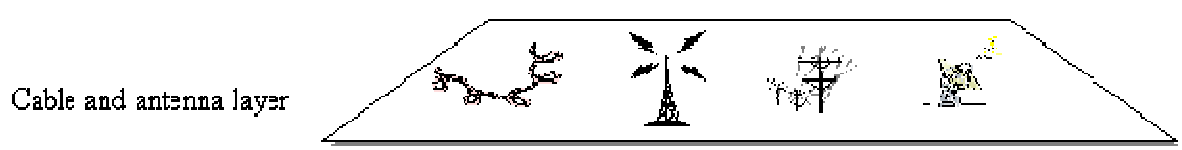
Transmission Fundamentals

IP-based network solutions are both flexible and economical substitutes for solutions that utilize old network technologies. The diverse properties between these technologies result from how information is represented, transmitted and managed. Information is simply structured collections of data, and thus takes its meaning from the interpretation we give it. There are two fundamental types of data, analog and digital, and both possess different behaviors and characteristics.

Analog data is expressed as continuously variable waves and thus takes on continuous values. Examples include voice and video.

Digital data on the other hand is represented as a sequence of bits, or *ones* and *zeros*. This digitization allows any kind of information to be measured and represented as digital data. So, text, sound and pictures can be represented as a sequence of bits. Digital data can also be compressed to allow higher transmission rates and it can be encrypted for secure transmissions. In addition, a digital signal is exact and any related noise can easily be filtered out. Digital data can be transmitted through three general types of media—metal such as copper; optical fiber or radio waves.

The techniques represented below offer the first building block for digital communications, the *cable and antenna layer* (Figure 3). This layer allows us to send and receive digital data over a wide variety of media. However, more building blocks are required for successful digital communication.



The Local Area Network Infrastructure

This section will go one step further by discussing digital *communication*. You might ask, "What is the difference between transmission and communication?" Consider an analogy from human speech. Think about the acoustic waves in the air generated by speaking. These waves are transmitted, but they are a long way from communicating. The words that come out must be organized to make any sense. If they come out to quickly or too slowly, the speaker will not be understood. If many people speak simultaneously no one is understood. If someone speaks a language you don't understand, information is lost. Speaking generates information, but it is not necessarily communicated, or understood.

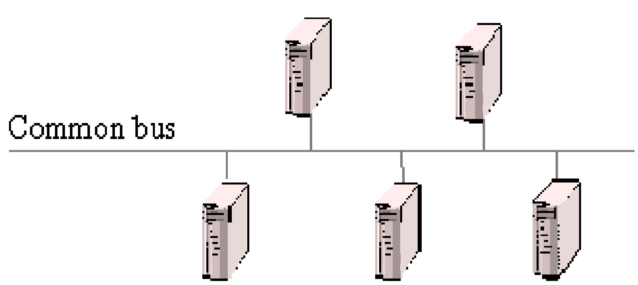
Digital communication has similar problems that need to be overcome. The receiver must know how message bits are organized to understand the message. The receiver must know the rate at which the bits are arriving to interpret the message. Additionally, some rules must specify what will happen if many network devices try to use a shared media simultaneously. The best way to ensure that network devices send and receive in compatible ways is to adhere to standardized *protocols* that define the rules and the manner in which the devices initiate and carry on communication.

We have until now focused on communication between two network devices. However, several different connection strategies and protocols exist that can be used to maintain communication among many network devices.

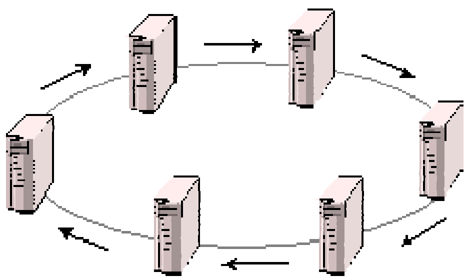
Local Area Networks (LANs) are used for connecting network devices over a relatively short distance. Typically, a LAN operates in a limited space, such as an office building, a school or a home. LANs are usually owned and managed by a single person or organization. They also use certain specific connectivity technologies, often some type of shared media.

An important feature of a LAN is its topology, where the term *topology* refers to the layout of connected network devices on a network. We can think of topology as a network's shape. Network topologies can be categorized into the following basic types:

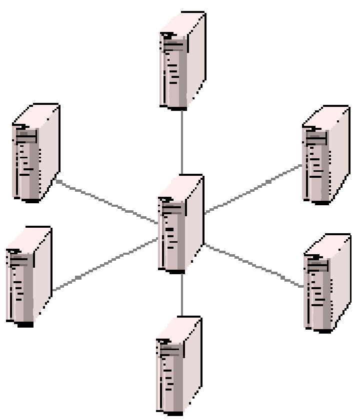
- The **bus topology** uses a shared communication medium, often referred to as a common bus, to connect all network devices (Figure 4). A device that wants to communicate with another device on the network sends the packet onto the bus. All devices that are connected to the bus will receive the sent packet but the intended recipient is the only device that actually accepts and processes the packets.



- The **ring topology** is structured in such a way that every network device on the network has exactly two neighbors for their communication purposes. All packets travel along a ring in the same direction .

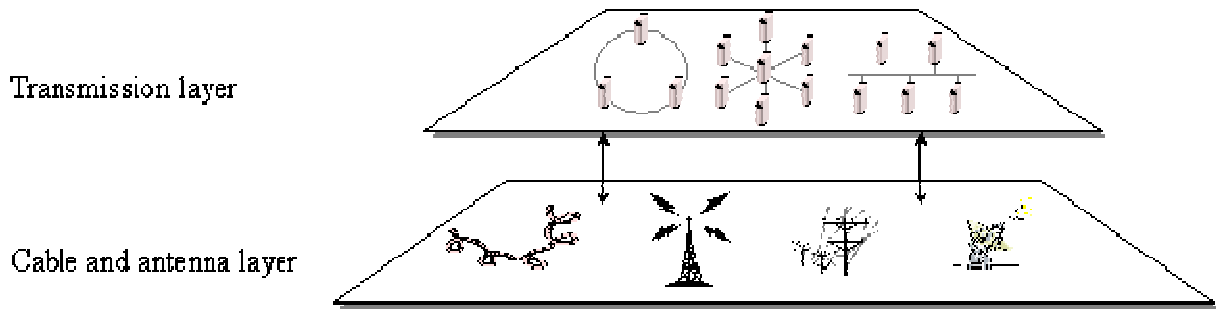


- The **star topology** features a logical communication center to which all network devices are directly connected. Each device requires a separate cable to the central point and consequently all packets will travel through the communication center .



There are several different protocols that can be utilized together with each network topology. Aside from identifying the standards of communications between the network devices, a protocol sets the technical specifications needed to transmit data within a network. To transmit a message to another device in a network, the message is split into *data packets*. These data packets are then transmitted via the communication media and are reassembled again at the receiving end.

The standardized protocols utilize different network topologies together with the cable and antenna layer to build different LAN architectures that are either wired or wireless. These protocols offer the second building block for successful digital communications, the *transmission layer* .



Interconnecting LANs in an IP-based Architecture

So far, we have described how network devices can communicate over different types of LANs. However, different LANs are designed for different goals and needs. Hence, every so often it is necessary to interconnect several LANs to allow communication over the network boundaries. Such a geographically scattered, interconnected collection of LANs is commonly referred to as a *Wide Area Network (WAN).* Probably the most familiar WAN is the Internet, which spans most of the globe.

Shared communication architecture is required for all users, such as private persons, enterprises, public administration offices and other organizations, to be able to exchange digital information with one another over a WAN. This architecture should be an open standard and support different transmission layer protocols, particularly those that can be used over a variety of transmission media. Fortunately, the Internet protocol suite provides a well-designed solution that fits these requirements.

The Internet protocol suite

The Internet protocol suite is a layered protocol family where each layer builds upon the layer below it, adding new functionality. The lowest layer is concerned purely with sending and receiving data utilizing the transmission layer. At the top are protocols designed for specific tasks, such as sending and receiving motion pictures, sound and control information. The protocols in between handle things such as dividing the message data into packets and forwarding them reliably between network devices.

Internet Protocol

The Internet Protocol (IP) is the basis of the Internet protocol suite and is the single most popular network protocol in the world. IP enables data to be transmitted across and between local area networks, hence the name: Inter-net Protocol. Data travels over an IP-based network in the form of *IP packets* (data units). Each IP packet includes both a header and the message data itself, where the header specifies the source, the destination, and other information about the data.

IP is a connectionless protocol where each packet is treated as a separate entity, like a postal service. Any mechanisms for ensuring that sent data arrives in a correct and intact manner are provided by higher-layer protocols in the suite.

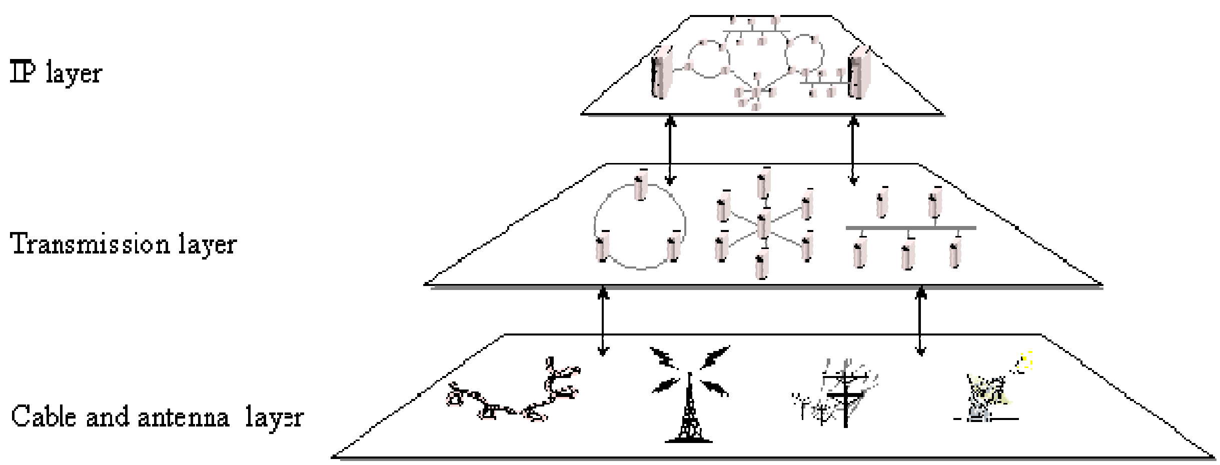
Each network device has at least one IP address that uniquely identifies it from all other devices on the network. In this manner, intermediate nodes can correctly guide a sent packet from the source to the destination.

Transport Protocol

The Transport Control Protocol (TCP) is the most common protocol for assuring that an IP packet arrives in a correct and intact manner. TCP provides reliable transmission of data for upper layer applications and services in an IP environment. TCP offers reliability in the form of a connection-oriented, end-to-end packet delivery through an interconnected network.

An Internet Protocol suite summary

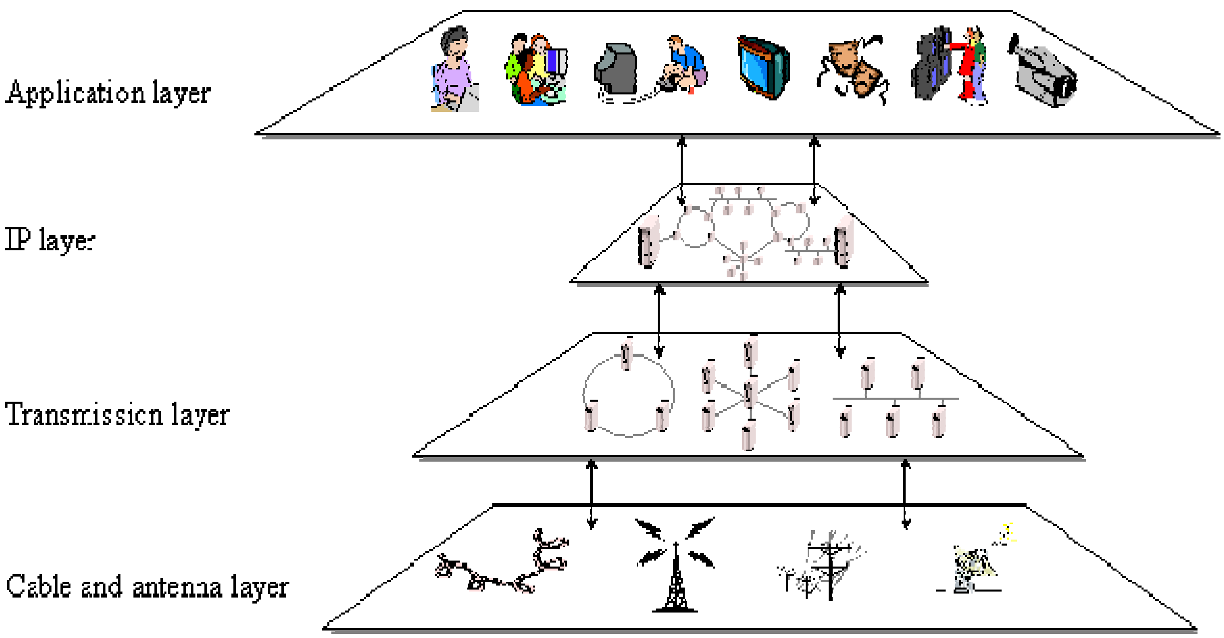
The Internet Protocol suite provides an adaptation to the transmission layer protocols and offers a standardized architecture for communication over an interconnected collection of LANs, i.e. a WAN. This is a tremendous advance, mainly because we're able to connect and communicate over different physical connections in a standardized way. With IP as the basis, the Internet Protocol suite provides the third building block for successful digital communications, the *IP layer* .



Benefit from the IP-based Architecture

The Internet Protocol suite brings together all transmission layer protocols into a single, standardized protocol architecture, which can be utilized by applications for different communication purposes. As a direct result, any application that supports TCP/IP will also be able to communicate over any IP-based network.

It should be easy to see that this standardized architecture has revolutionized network communication. An ever-increasing number of applications that transfer text, sound, live pictures and more utilize IP-based architecture. All these applications and application protocols constitute the *application layer* and provide the fourth, and final, building block for successful digital communications (Figure 9).



Convergence

Modern digital technology allows for convergence where different services, and combinations of these services, can be provided through infrastructures that formerly accommodated only one type of service. There are three major factors that create the conditions for convergence: digital technology, transmission technology and standardized communication protocols. Digital technology allows all information—text, sound and motion pictures, for example—to be represented as bits and transmitted as sequences of ones and zeros. Transmission technology enables better utilization of available capacity in different infrastructures. Consequently, services that require high capacity can be provided by infrastructures previously able to deliver only simpler services.

We have already seen how IP-based technology provides an excellent architecture for the process of ongoing convergence. At the heart of the Internet Protocol suite is the Internet Protocol, which represents the building block that uniformly connects different physical networks with a variety of applications. In addition, presently available IP-based solutions can be fully integrated with other available systems.

Case Study

So far we have discussed the structure of the IP-based architecture, especially in comparison with traditional circuit-switched networks. However, the preceding sections have not contained any real applications that take advantage of this architecture. IP-based architecture creates great opportunities for new application domains. Hence, applications that previously could not be realized can now be successfully implemented. Additionally, application domains built upon older technologies derive increased functionality when utilizing IP-based technology. For illustration, consider an application domain that has clearly taken advantage of IP-based architecture: visual surveillance systems.

In today's society, the demand for visual surveillance systems has been steadily increasing. Different camera solutions are used for monitoring activities in a variety of environments, such as shops, enterprise buildings and prisons. Up until recently, Closed Circuit Television systems (CCTV systems) were the only alternative for such monitoring. These dedicated systems typically require their own communication link between the camera and the monitor. This separate link is expensive to buy, install and maintain. Camera images are transmitted over the dedicated cabling network to time-lapse video recorders or dedicated monitors at a control center.

A modern IP-based visual surveillance system on the other hand is not limited in the same way as a traditional CCTV system. Enterprises can install *network cameras,* IP-based visual surveillance cameras that plug directly into the enterprise network. Such cameras have their own IP address, much like any network device. The main differences between these systems and CCTV systems are that video digitization is performed at the camera level and the Internet Protocol suite is utilized for transferring the pictures onto the network. This is beneficial since IP-based networks are generally available in most buildings, and because TCP/IP can be utilized with almost any existing network, there is probably no need for extra cabling. A network camera system, in comparison with a CCTV system, also saves money by reducing the amount of dedicated equipment needed to manage the security system. For example, no dedicated monitors are required.

An IP-based solution also allows images to be remotely stored and monitored over any interconnected network, such as the Internet. This alone creates huge advantages for enterprises that wish to outsource the monitoring of their offices and facilities to a third party surveillance and monitoring center. This center simply needs a password and the IP-address to access live pictures, via the Internet, from a camera placed anywhere in the world. Moreover, the IP-based architecture creates a new world in which different applications can be completely integrated. For instance, motion pictures can be distributed to other network solutions, such as factory control management systems and access control systems.

Chapter - 9 (VOIP,Softswitch, NGN)

Voice Over IP

Introduction

Although voice over IP (VoIP) has been in existence for many years, it has only recently begun to take off as a viable alternative to traditional public switched telephone networks (PSTN). Interest and acceptance has been driven by the attractive cost efficiencies that organizations can achieve by leveraging a single IP network to support both data and voice. But cost is not enough to complete the evolution; service and feature parity is a main requirement. Customers will not accept voice quality or service that is inferior to what they are used to with a PSTN and, until now, VoIP fell short in delivery.

Voice protocols have evolved to offer a richer set of features, scalability and standardization than what was available only a few years ago. The pace of service integration (convergence) with new and existing networks continues to increase as VoIP products and services develop. Critical to success is the ability to deploy value-added and high-margin services. For example, a service provider can deploy a unified messaging system that synthesizes voice and e-mails over a phone to the subscriber.

This paper will explain the fundamentals of VoIP, focusing on the functions and components that make up a VoIP solution. It will answer the following questions: What does it mean to an organization to deploy VoIP? What makes up a VoIP solution and how can they take advantage of it? Once a general understanding of VoIP is achieved, organizations are better prepared to tackle the more complex issues that go into deploying a secure, reliable and high-performance VoIP network.

Why VoIP?

The cost-effectiveness is initially attractive when looking into VoIP. It is evident that an organization can gain efficiencies by only having to support a single network infrastructure. By using a single packet-switched network, as opposed to having to manage both packet and circuit-switched networks, organizations can realize reduced maintenance and management costs. The same technical personnel are able to operate both voice and data systems instead of requiring resources with different expertise.

This convergence of voice and data networks onto a single IP network also provides some inherent flexibility, in terms of being able to easily add, change or remove nodes (e.g. phones) on the network. As a result, organizations can easily deploy and then redeploy equipment to maximize their investments, without having to do a truck roll or require special expertise on hand.

Finally, VoIP promises to deliver many nice new features, such as advanced call routing, computer integration, unified messaging, integrated information services, long-distance toll bypass, and encryption. Because of the common network infrastructure, it is also possible to integrate other media services, like video or even electronic white boards, to name a few. An example of such features would be the "follow-me" feature where a person is always reachable at the same extension, whether telecommuting from his Lake Tahoe cabin, staying in a hotel abroad, or sitting at his desk in the office. Another feature would be the integration of VoIP with customer relationship management (CRM) software. CallerID or dialed numbers could be linked to a customer's record, which automatically opens on the desktop when the Sales person receives or places a call.

Due to the cost-effectiveness, flexibility and promise that leveraging a single IP network offers, it is no wonder that organizations are looking hard at the VoIP technology and trying to figure out how best to use it to their advantage.

VoIP Functions

Before going into a discussion of the components that make up a VoIP solution, it is important to understand the basic functions of VoIP, particularly as they compare to current PSTNs. As mentioned above, in order to enable organizations to adopt VoIP as a viable solution, its components must be able to perform the same functions as the PSTN network. These are:

- Signaling
- Database services
- Call connect and disconnect (bearer control)
- CODEC operations

Signaling

Signaling is the way that devices communicate within the network, activating and coordinating the various components needed to complete a call.

In a PSTN network, phones communicate with a Class 5 switch (analog) or traditional private branch exchange (PBX) (digital) for call connection and call routing purposes.

In a VoIP network, signaling is accomplished by the exchange of IP datagram messages between the VoIP components. The format of these messages may be dictated by any number of standard protocols, which are covered later in this paper.

Database Services

Database services are a way to locate an endpoint and translate the addressing that two (usually heterogeneous) networks use. A call control database contains these mappings and translations. Another important feature is the generation of transaction reports for billing purposes. You can employ additional logic to provide network security, such as to deny a specific endpoint from making overseas calls. This functionality, coupled with call state control, coordinates the activities of the elements in the network.

A PSTN uses phone numbers to identify endpoints.

A VoIP network uses an IP address (address abstraction could be accomplished with DNS) and port number to identify an endpoint.

Call Connect and Disconnect (Bearer Control)

The connection of a call is made by two endpoints opening a communication session between one another. In the PSTN, the public (or private) switch connects logical (Digital Signal) DS-0 channels through the network to complete the calls.

In a VoIP implementation, this connection is a multimedia stream (audio, video, or both) transported in real time. This connection is the bearer channel and represents the voice or video content being delivered. When a communication is complete, the IP sessions are released and optionally network resources are freed.

CODEC Operations

Traditional voice communication is analog, while data networking is digital, as a result, the network needs a way to be able to convert the voice into a format that it can transport. Since the PSTN is often analog, this is not necessarily a major function, however, for VoIP, it is necessary for “packetiz-ing” the voice. The process of converting analog waveforms to digital information is done with a coder-decoder (CODEC, which is also known as a voice coder-decoder [VOCODER]). There are many ways an analog voice signal can be transformed, all of which are governed by various standards. The process of conversion is complex and beyond the scope of this paper. Suffice it to say that most of the conversions are based on pulse coded modulation (PCM) or variations. Each encoding scheme has its own history and merit, along with its particular bandwidth needs.

The output from the CODECs is a data stream that is put into IP packets and transported across the network to an endpoint. These endpoints must use the standards, as well as a common set of CODEC parameters. If two endpoints use different standards or parameters then the communication will be unintelligible. Table 1 lists some of the more important encoding standards covered by the International Telecommunications Union (ITU). Notice the tradeoff between encoding efficiency, reduced bandwidth consumption, and increased conversion delay.

Table 1: ITU Encoding Standards

ITU Standard	Description	Bandwidth (Kbps)	Conversion Delay (ms)
G.711	PCM	64	< 1.00
G.721	ADPCM	32, 16, 24, 40	< 1.00
G.728	LD-CELP	16	~ 2.50
G.729	CS-ACELP	8	~ 15.00
G.723.1	Multirate CELP	6.3, 5.3	~ 30.00

VoIP Components

The major components of a VoIP network, while different in approach, deliver very similar functionality to that of a PSTN and enable VoIP networks to perform all of the same tasks that the PSTN does. The one additional requirement is that VoIP networks must contain a gateway component that enables VoIP calls to be sent to a PSTN, and visa versa. There are four major components to a VoIP network.

- Call Processing Server/IP PBX
- User End-Devices
- Media/VOIP Gateways
- IP network

Call Processing Server / IP PBX

The call processing server, otherwise known as an IP PBX, is the heart of a VoIP phone system, managing all VoIP control connections. Call processing servers are usually software-based and can be deployed as a single server, cluster of servers, or a server farm with distributed functionality. Call processors may also be based on a router platform or developed as a dedicated appliance.

VoIP communications require a signaling mechanism for call establishment, known as control traffic, and actual voice traffic, known as voice stream or VoIP payload. VoIP control traffic follows the client-server model, with VoIP terminals, including messaging servers that hold voice-mail messages representing the clients that communicate to the call processing servers.

With the exception of routed voice traffic to another call processing server, conferencing functionality and music-on-hold, call processing servers do not handle VoIP payload (which is the RTP stream carrying voice itself) traffic. VoIP payload flows in a peer-to-peer fashion – from every VoIP terminal to every other VoIP terminal. In this case, the VoIP terminals determine traffic flows and the call processing servers negotiate those flows within the control messages. A typical VoIP setup with Call Processing Server is shown in Figure 1.

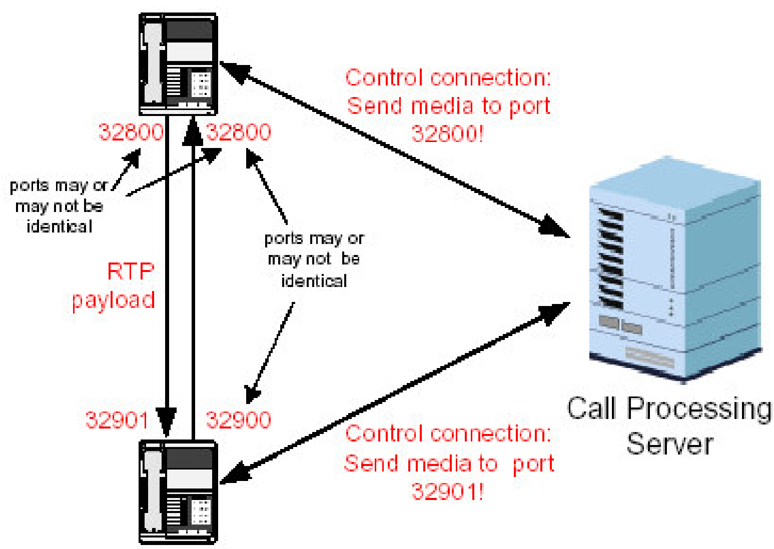


Figure1: Call Processing Server

Figure 2 shows how different signaling protocols can be used by these Call Processing Servers to communicate with IP Phones, Gateways/Gatekeeper, which will be discussed in a following section. Signaling protocols and their functions are also described later, in the Signaling Protocols section.

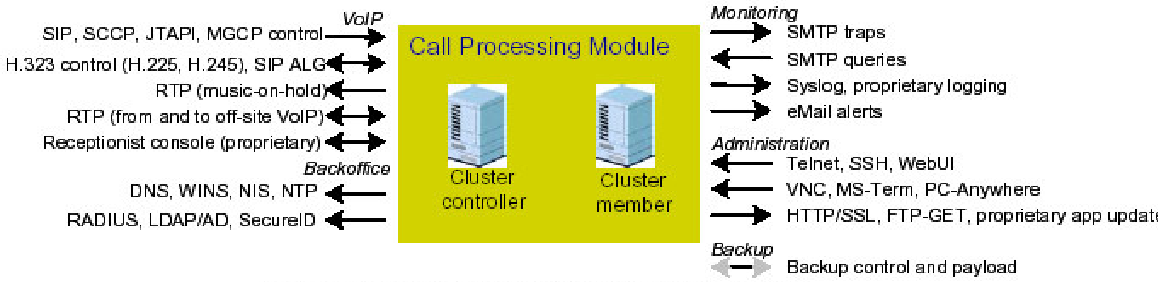


Figure 2: Call Processing Server Signaling

User End-Devices

The user end-devices consist of VoIP phones and desktop-based devices. VoIP phones maybe software based ("softphones") or hardware based ("hard phones" or "handsets", like traditional phones).

- VoIP phones use the TCP/IP stack to communicate with the IP network, as such, they are allocated an IP address for the subnet on which they are installed. VoIP phones may also use additional protocols to support VoIP-enabled features, such as built-in IM applications or directory search functions. Typically, VoIP phones use DHCP to auto-configure themselves, with the DHCP server telling the phone about the location of the configuration server, which most of the time is identical to the call processing server.
- Softphones are software application running on notebook computers, usually targeted towards mobile users. They have the same base features as VoIP phones.
- Consoles, on the other hand, are applications with certain control characteristics. Consoles usually include a Softphone, but may also interact with a legacy phone, via a voice gateway or a VoIP phone. Consoles are special-purpose applications to control call distribution. This includes receptionist consoles with the ability to connect calls, executive consoles with the ability to see call states of special groups of phones, and customer relations consoles with the ability to support call distribution. The distinction among the different types of consoles is not too clear. All VoIP consoles have in common the use of proprietary protocol extensions. Proprietary protocol extensions can be problematic for all stateful firewalls, unless the firewall can understand the non-standard signaling. Consoles should be installed on dedicated desktop computers, with no access to the Internet and only controlled access to data network services, in order not to expose the voice network. Consoles are usually static and should be confined to their own network within the module.

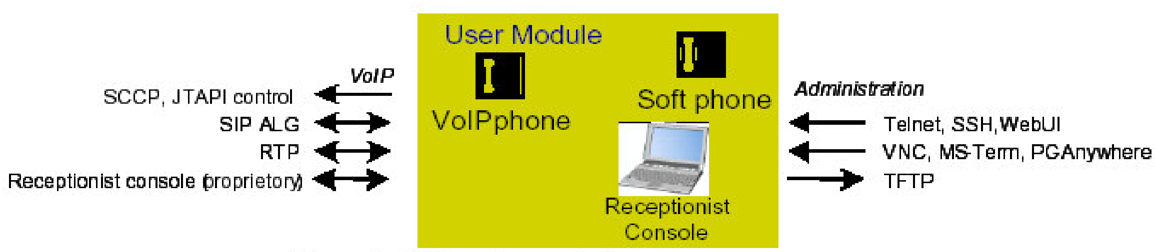


Figure 3: VoIP end user devices

Media/VOIP Gateways/Gatekeepers

The terms gateway and gatekeeper are sometimes used interchangeably. Traditionally gatekeepers have been mainly used for Call Admission and control and bandwidth management. But this has changed recently, as technology has allowed this functionality to co-exist within traditional gateways (described below).

The major function of media gateways is analog-to-digital conversion of voice and creation of voice IP packets (CODEC functions). In addition, media gateways have optional features, such as voice (analog and/or digital) compression, echo cancellation, silence suppression, and statistics gathering.

The media gateway forms the interface that the voice content uses so it can be transported over the IP network. Media gateways are the sources of bearer traffic. Typically, each conversation (call) is a single IP session transported by a Real-time Transport Protocol (RTP) that runs over UDP or TCP.

Media gateways exist in several forms. For example, media gateways could be a dedicated telecommunication equipment chassis, or even a generic PC running VoIP software. Their features and services can include some or all of the following:

- Trunking gateways that interface between the telephone network and a VoIP network. Such gateways typically manage a large number of digital circuits.
- Residential gateways that provide a traditional analog interface to a VoIP network. Examples of residential gateways include cable modem/cable set-top boxes, xDSL devices and broadband wireless devices.
- Access media gateways that provide a traditional analog or digital PBX interface to a VoIP network. Examples include small-scale (enterprise) VoIP gateways.
- Business media gateways that provide a traditional digital PBX interface or an integrated soft PBX interface to a VoIP network.
- Network access servers that can attach a modem to a telephone circuit and provide data access to the Internet.

Figure 4 shows a VoIP gateway and the signaling protocols it uses to communicate with VoIP Call processing servers and other VOIP devices, such as IP Phones, messaging systems, etc.

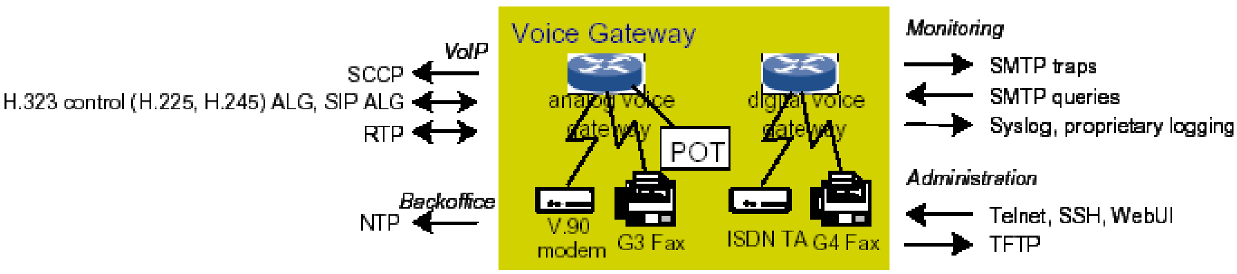


Figure 4: VoIP/Media Gateway

IP Network

You can view the VoIP network as one logical switch. However, this logical switch is a distributed system, rather than that of a single switch entity; the IP backbone provides the connectivity among the distributed elements. Depending on the VoIP protocols used, this system as a whole is sometimes referred to as a *softswitch architecture*.

The IP infrastructure must ensure smooth delivery of the voice and signaling packets to the VoIP elements. Due to their dissimilarities, the IP network must treat voice and data traffic differently. If an IP network is to carry both voice and data traffic, it must be able to prioritize the different traffic types, as VoIP traffic is extremely sensitive to latency.

While there are several similarities between VoIP and circuit-switching components, there are also several differences. One is in the transport of the resulting voice traffic. Circuit-switching telecommunications can be best classified as a TDM network that dedicates channels, reserving bandwidth as it is needed out of the trunk links interconnecting the switches. For example, a phone conversation reserves a single DS-0 channel, and that end-to-end connection is used only for the single conversation. This is not an efficient method of resource utilization.

IP networks are quite different from the circuit-switch infrastructure in that it is a packet-network, and it is based on the idea of statistical availability. Thus network resources are not completely tied up for the duration of the call, unlike in a circuit-switched environment. Class of service (CoS) ensures that packets of a specific application are given priority. This prioritization is required for real-time VoIP applications to ensure that the voice service is unaffected by other traffic flows.

VoIP Signaling Protocols

VoIP signaling protocols are the enablers of the VoIP network. The protocols determine what types of features and functionality are available, as well as how all of the VoIP components interact with one another.

There are a variety of VoIP protocols and implementations, with a wide range of features that are currently deployed. Two major standards bodies govern multimedia delivery (voice being one type) over packet-based networks: International Telecommunications Union (ITU) and Internet Engineering Task Force (IETF). H.323 is the ITU's standard for establishing VOIP connections, while IETF uses Session Initiation Protocol (SIP) as its standard. More implementations tend to be focused on the ITU specifications than those of the IETF, primarily because H.323 is more widely deployed today than SIP. This is expected to change, however. It should be noted, many of the standards in both bodies are based on solving the same problems. The result is some overlap of functionality, as well as differences in approach and nomenclature. To further confuse the issue, some vendors are implementing proprietary schemes that fill apparent gaps in the standards or add functionality that is product dependent. Also, not all of the protocols are used in one specific product group. Instead, the product vendor will code its offerings with what is most applicable for its scope, services, and market.

Each of the voice protocols has its own strengths and weaknesses, and each takes a different approach to service delivery. Each of these protocols is successful in different products having a specific market focus. The protocols listed in this section are the most prevalent, but do not constitute and exhaustive list; there are a few other protocol options available.

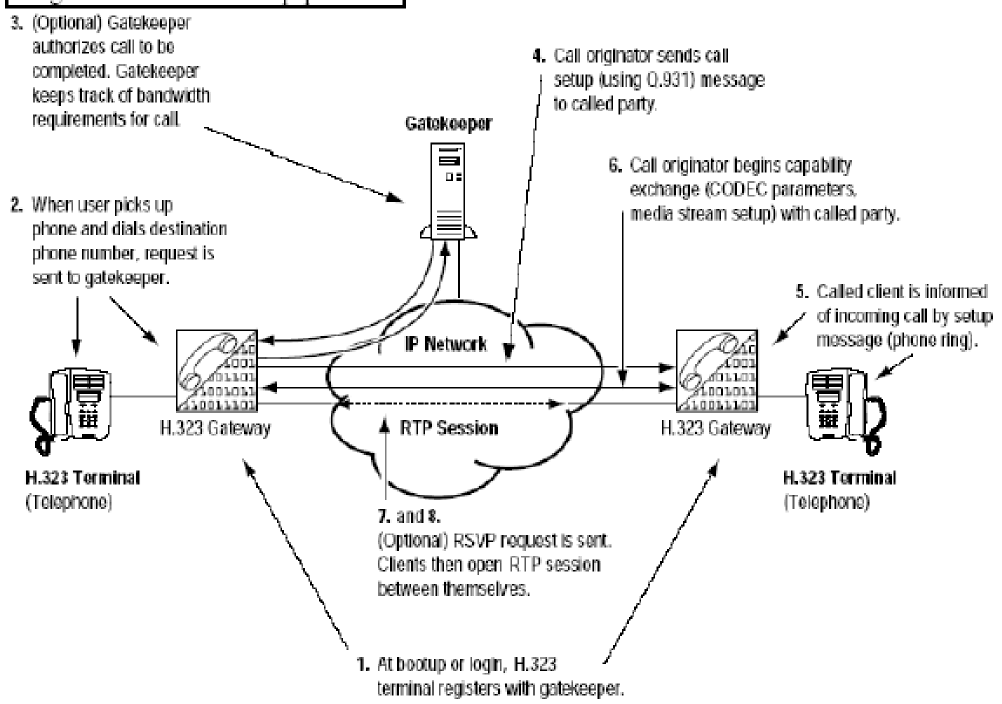
H.323

H.323 is the ITU recommendation. It is a packet-based multimedia communication system that is a set of specifications. These specifications define various signaling functions, as well as media formats related to “packetized” audio and video services.

H.323 standards were generally the first to classify and solve multimedia delivery issues over LAN technologies. However, as IP networking and the Internet became prevalent, many Internet RFC standard protocols and technologies were developed and based on some of the previous H.323 ideas. Today there is cooperation between the ITU and IETF in solving existing problems, but it is fair to say that the RFC process of furthering the standards has had greater success than the H.323 counterparts.

H.323 networks consist of Call Processing Servers, (media) gateways and gatekeepers. Call Processing Servers provide call routing, and communication to VOIP gateways and end devices. Gateways serve as both the H.323 termination endpoint and interface with non-H.323 networks, such as the PSTN. Gatekeepers function as a central unit for call admission control, bandwidth management and call signaling. Although the gatekeeper is not a required element in H.323, it can help H.323 networks to scale to a larger size, by separating call control and management functions from the gateways.

Figure 5: H323 call setup process



H.323 specifications tend to be heavier (due to chattiness, in terms of control signaling) and with an initial focus in LAN networking. These standards have some shortcomings in scalability, especially in large-scale deployments. Primarily, limitations are due to chattiness or the heavy signaling required to establish H.323 sessions. H.323 is dependant on TCP-based (connection-oriented) signaling. There is a challenge in maintaining large numbers of TCP sessions because of the substantial overhead involved. However, most H.323 scalability limitations are based on the prevalent version two of the specification. Subsequent versions of H.323 have a focus on solving some of these problems.

Let’s look at the main H323 process:

- With each call that is initiated, a TCP session (H.225.0 protocol) is created, using an encapsulation of a subset of Q.931 messages. This TCP connection is maintained for the duration of the call. Complete call setup process is shown in figure 5.
- A second session is established using the H.245 protocol. This TCP-based process is for capabilities exchange, master-slave determination, and the establishment and release of media streams. This group of procedures is in addition to the H.225.0 processes.
- The H.323 quality of service (QoS) delivery mechanism of choice is the Resource Reservation Protocol (RSVP). This protocol is not considered to have good scaling properties due to its focus and management of individual application traffic flows.
- Although H.323 many not be well suited in service provider spaces, it is well positioned to deploy enterprise VoIP applications. As a service provider, it might be necessary to bridge, transport or interface H.323 services and applications to the PSTN.

Real-time Transport Protocol (RTP)

RFC 1889 and RFC 1890 cover the Real-time Transport Protocol (RTP), which provides end-to-end delivery services for data with real-time characteristics, such as interactive audio and video. Services include payload type identification, sequence numbering, time stamping and delivery monitoring. The media gateways that digitize voice use the RTP protocol to deliver the voice (bearer) traffic.

The RTP protocol (Figure 6) provides features for real-time applications, with the ability to reconstruct timing, loss detection, security, content delivery and identification of encoding schemes. For each participant, a particular pair of destination IP addresses defines the session between the two endpoints, which translates into a single RTP session for each phone call in progress. RTP is an application service built on UDP, so it is connectionless, with best-effort delivery. Although RTP is connectionless, it does have a sequencing system that allows for the detection of missing packets.

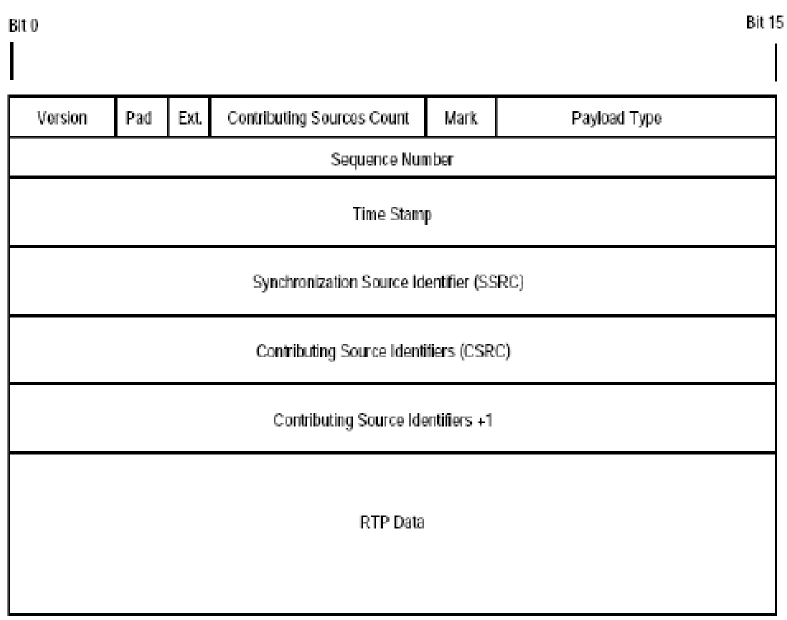


Figure 6: Upper Layer of RTP Protocol

As part of its specification, the RTP Payload Type field includes the encoding scheme that the media gateway uses to digitize the voice content. This field identifies the RTP payload format and determines its interpretation by the CODEC in the media gateway. A profile specifies a default static mapping of payload type codes to payload formats. These mappings represent the ITU G series of encoding schemes.

With the different types of encoding schemes and packet creation rates, RTP packets can vary in size and interval. Administrators must take RTP parameters into account when planning voice services. All the combined parameters of the RTP sessions dictate how much bandwidth is consumed by the voice bearer traffic. RTP traffic that carries voice traffic is the single greatest contributor to the VoIP network load.

Real-time Transport Control Protocol (RTCP)

Real-time Transport Control Protocol (RTCP) is the optional companion protocol to RTP; it is not needed for RTP to work. The primary function of RTCP is to provide feedback on the quality of the data distribution being accomplished by RTP. This function is an integral part of RTP's role as a transport protocol and is related to the flow and congestion control functions of the network. Although the feedback reports from RTCP do not describe where problems are occurring (only that they are), they can be used as a tool to locate problems. With the information generated from different media gateways in the network, RTCP feedback reports enable an administrator to evaluate where network performance might be degrading.

RTCP enables administrators to monitor the quality of a call session by tracking packet loss, latency (delay), jitter, and other key VoIP concerns. This information is provided on a periodic basis to both ends and is processed per call by the media gateways.

Some gateway devices might not employ RTCP because the facility to report such information is not applicable to the end user. For example, a single residential user (with an analog phone) might not have access to the gateway providing the service. Also, the media gateway vendor can use a more scalable approach of tracking call quality statistics. In this case, the storage, transport and presentation of statistical info are device dependent.

If using RTCP (or a vendor-specific implementation) in the network, the organization needs to take into account bandwidth calculations for the protocol. Administrators need to limit the control traffic of RTCP to a small and known fraction of the session bandwidth. It should be small so as not to impair the ability of the transport protocol to carry data. An organization should investigate the amount of bandwidth needed so that they can include the control traffic in the bandwidth specification. RFC specifications recommend that the fraction of the session bandwidth allocated to RTCP be fixed at five percent of RTP traffic.

Media Gateway Control Protocol (MGCP)

The Media Gateway Control Protocol (MGCP, RFC 2705) is along the lines of a softswitch architecture philosophy. It breaks up the role of traditional voice switches into the components of media gateway, media gateway controller and signaling gateway functional units. This facilitates the independent managing of each VoIP gateway as a separate entity.

MGCP is a master-slave control protocol that coordinates the actions of media gateways (Figure 7). The media gateway controller in MGCP nomenclature is sometimes referred to as a call agent. The call agent manages the call-related signaling control intelligence, while the media gateway informs the call agent of service events. The call agent instructs the media gateway to create and tear down connections when the calls are generated. In most cases, the call agent informs the media gateways to start an RTP session between two endpoints.

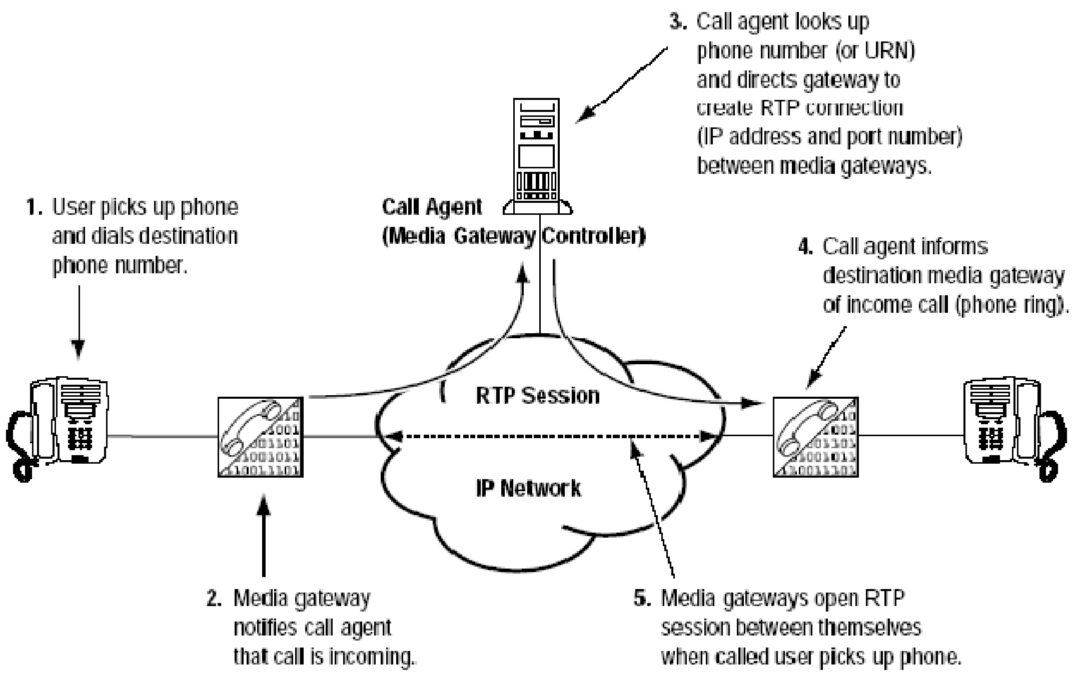


Figure 7: How MGCP Coordinates the Media Gateways

Session Initiation Protocol (SIP)

The Session Initiation Protocol (SIP, RFC 2543) is part of IETF's multimedia data and control protocol framework. SIP is a powerful client-server signaling protocol used in VoIP networks. SIP handles the setup and tear down of multimedia sessions between speakers; these sessions can include multimedia conferences, telephone calls, and multimedia distribution.

SIP is a text-based signaling protocol transported over either TCP or UDP, and is designed to be lightweight. It inherited some design philosophy and architecture from the Hypertext Transfer Protocol (HTTP) and Simple Mail Transfer Protocol (SMTP) to ensure its simplicity, efficiency and extensibility.

SIP uses *invitations* to create Session Description Protocol (SDP) messages to carry out capability exchange and to setup call control channel use. These invitations allow participants to agree on a set of compatible media types.

SIP supports user mobility by proxying and redirecting requests to the user's current location. Users can inform the server of their current location (IP address or URL) by sending a registration message to a *registrar*. This function is powerful and often needed for a highly mobile voice user base. The SIP client-server application has two modes of operation; SIP clients can ether signal through a *proxy* or *redirect* server.

- Using proxy mode (Figure 8), SIP clients send requests to the proxy and the proxy either handles requests or forwards them on to other SIP servers. Proxy servers can insulate and hide SIP users by proxying the signaling messages; to the other users on the VoIP network, the signaling invitations look as if they are coming from the proxy SIP server.

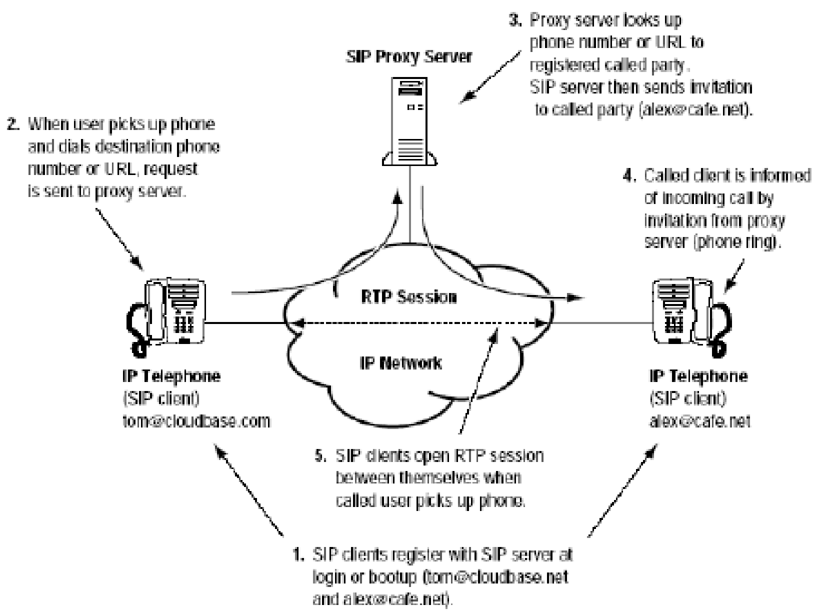


Figure 8: SIP Proxy Operation

- Under redirect operation (Figure 9), the signaling request is sent to a SIP server, which then looks up the destination address. The SIP server returns the destination address to the originator of the call, who then signals the SIP client

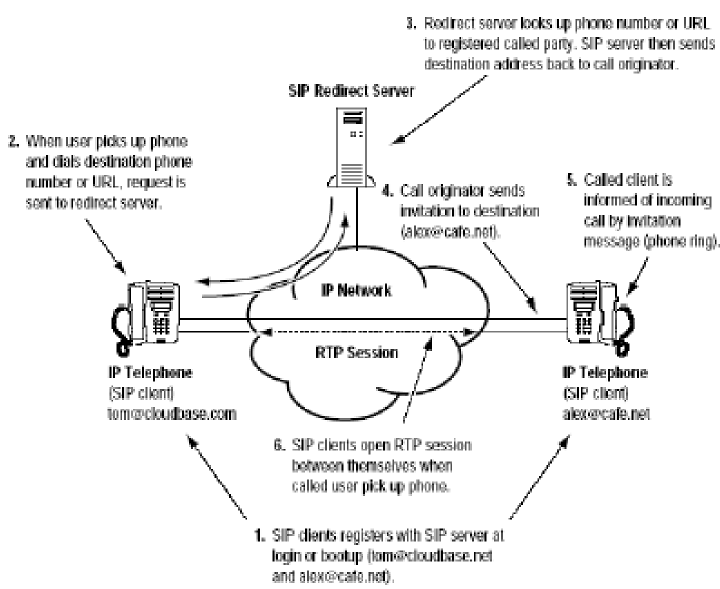


Figure 9: SIP Redirector Server

Megaco/H.248

Megaco/H.248 is a current draft standard and represents a cooperative proposal from the IETF and ITU standards bodies. Megaco has many similarities to MGCP and borrows the same naming conventions for the VoIP elements. The Megaco architecture defines media gateways that provide media conversion and sources of calls, while media gateway controllers provide call control.

Megaco addresses the same requirements as that of MGCP and, as a result, there is some effort to merge the protocols. It defines a series of transactions coordinated by a media gateway controller for the establishment of call sessions. The primary focus of Megaco is the promotion to standardize IP telephony equipment. Some of the design goals are as follows:

- Megaco IP phone meets the basic needs of the business user from day one.
- Provides a path for rapid expansion to support sophisticated business telephony features.
- Allows for a wide range of telephones and similar devices to be defined from very simple to very feature rich.
- Implements a simple, minimal design.
- Allows device cost to be appropriate to capabilities provided. Package and termination types have characteristics that enable reliability.
- IP phone meets the appropriate Megaco/H.248 protocol requirements, as provided in the Megaco requirements document, and are a straightforward application of the Megaco/H.248 protocol.

VoIP Service Considerations

Now that this paper has gone through the functions, components and protocols related to VoIP traffic, let's take a quick look at some of the issues an organization must carefully consider when deploying VoIP solutions, such as traffic parameters and network design. Without such due diligence, an organization could be faced with service that does not function reliably or is severely degraded. The important considerations are as follows:

- Latency
- Jitter
- Bandwidth
- Packet loss
- Reliability
- Security
- Interoperability

Latency

Latency (or delay) is the time that it takes a packet to make its way through a network end-to-end. In telephony terms, latency is the measure of time it takes the talker's voice to reach the listener's ear. Large latency values do not necessarily degrade the sound quality of a phone call, but the result can be a lack of synchronization between the speakers, such that there are hesitations in the speaker' interactions.

Generally, it is accepted that the end-to-end latency should be less than 150 ms for toll quality phone calls. To ensure that the latency budget remains below 150 ms, administrators need to take into account the following primary causes of latency. When designing a multiservice network, the total delay that a signal or packet exhibits is a summation of all the latency contributors.

- One source of latency is the time it takes for the endpoints to create the packets used in voice services. These "packetization" delays are caused by the amount of time it takes to fill a packet with data. Generally, the larger the packet size, the greater the amount of time it takes to fill it. Packetization delay is governed by the CODEC standard being used. This problem also exists on the receiving side because the media gateway must remove and further process the packet data. If the packets are kept small, this amount of delay, in both directions, is usually quite small, depending on the hardware / software implementation of the media gateways. All considerations being equal, nominal operation of any media gateway unit should not exceed 30 ms.
- Another source of latency is the delay it takes to serialize the digital data onto the physical links of the interconnecting equipment. This delay is inversely proportional to the link speed. In other words, the faster the media, the lower the latency. This value is somewhat dependent on the link technology used and its access method. For example, it takes 125 microseconds to place one byte on a 64-Kb circuit. The same byte placed on an OC-3/STM-1 circuit takes 0.05 microseconds. Although this delay is unavoidable (regardless of the bandwidth used), keeping the number of intervening links small and using high bandwidth interfaces reduces the overall latency.
- Propagation delay is the time it takes an electrical (or photonic) signal to traverse the length of a conductor. The speed of these signals is always slower than that of the speed of light. There is always propagation delay; however, it only becomes an issue when the signal (or packet) travels a great distance. The accepted formula for calculating propagation delay is as follows.
- Propagation delay = Circuit km / (299,300 km x .6)
- **Example:** Calculation of one-way propagation delay of a 6,000 km fiber run (discounting any signal repeaters in between)
- 0.0334 sec = 6000 km / (299,300 km x .6)
- By this calculation, the latency contributed by just propagation delay would be 33.4 ms.
- A queuing delay, which is a large source of latency, is the amount of time that a packet remains buffered in a network element while it awaits transmission. Network traffic loads result in variable queuing delays. The amount of buffer that a queue uses is usually a configurable parameter, with a smaller number being better for latency values. However, this delay is also based on the amount of traffic the element is trying to pass through a given link, and therefore it increases with network load. Hence, you need to set aside adequate bandwidth and resources for voice traffic. If the queue used for voice traffic is not serviced fast enough and that queue is allowed to grow too large, the result is greater latency.
- Packet forwarding delay is the time it takes a network device (router, switch, firewall, etc.) to buffer a packet and make the forwarding decision. Included in that decision could be which interface to forward the packet to, whether to drop or forward the packet against an Access Control List (ACL) or security policy, etc. Packet forwarding delay is variable and depends on the function and architecture of the networking device. If a packet must be further buffered as a part of its processing, greater latency is incurred.

Jitter

Jitter is the measure of time between when a packet is expected to arrive to when it actually arrives. In other words, with a constant packet transmission rate of every 20 ms, every packet would be expected to arrive at the destination exactly every 20 ms. This situation is not always the case. For example, Figure 9 shows packet one (P1) and packet three (P3) arriving when expected, but packet two (P2) arriving 12 ms later than expected and packet four (P4) arriving 5 ms late.

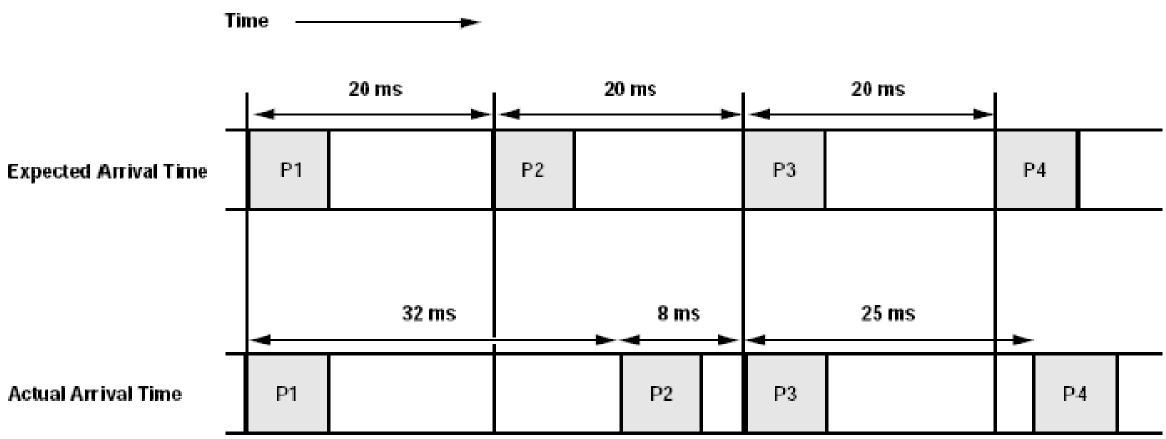


Figure 10: Jitter Example

The greatest culprit of jitter is queuing variations caused by dynamic changes in network traffic loads. Another cause is packets that might sometimes take a different equal-cost link that is not physically (or electrically) the same length as the other links.

Media gateways have *play-out buffers* that buffer a packet stream, so that the reconstructed voice waveform is not affected by packet jitter. Play-out buffers can minimize the effects of jitter, but cannot eliminate severe jitter.

Although some amount of jitter is to be expected, severe jitter can cause voice quality issues because the media gateway might discard packets arriving out of order. In this condition, the media gateway could starve its play-out buffer and cause gaps in the reconstructed waveform.

Bandwidth

An organization can determine how much bandwidth to set aside for voice traffic using simple math. However, in a converged voice and data network, administrators have to make decisions on how much bandwidth to give each service. These decisions are based on careful consideration of the organization's priorities and the available bandwidth that can be afforded. If an administrator allocates too little bandwidth for voice service, there might be unacceptable quality issues. Another consideration is that voice services are less tolerant to bandwidth depletion than that of Internet traffic. Therefore, bandwidth for voice services and associated signaling must take a priority over that of best-effort Internet traffic. If a network were to use the same prevailing encoding (CODEC) scheme as the current PSTN system, bandwidth requirements for VoIP networks would tend to be larger than that of a circuit-switched voice network of similar capacity. The reason is the overhead in the protocols used to deliver the voice service. Typically, an organization would need speeds of OC-12c/STM-4 and higher to support thousands of call sessions. However, VoIP networks that employ compression and silence suppression could actually use less bandwidth than a similar circuit-switched network.

The reason is because of the greater granularity in bandwidth usage that a packet-based network has in comparison to a fixed, channel size TDM network.

Allocations of network bandwidth are based on projected numbers of calls at peak hours. Any over-subscription of voice bandwidth can cause a reduction in voice quality. Also, you must set aside adequate bandwidth for signaling to ensure that calls are complete and to reduce service interruptions.

The formula for calculating total bandwidth needed for voice traffic is relatively straightforward. The formula to calculate RTP bearer voice bandwidth usage for a given number of phone calls is as follows:

bits per sec = packet creation rates per sec x packet size x number of calls x 8 bits per sec

where samples per sec = 1,000 ms / packet creation rate

Example

2,000 full-duplex G.711 encoded voice channels that have a packet creation rate of 20 ms, with a packet size of 200 bytes (40 byte IP header + 160 byte payload)

50 samples per second = 1,000 ms / 20 ms

160 Mbps = 50 x 200 x 2,000 x 8

Note that this number is a raw measure of IP traffic and does not take in account the overhead used by the transporting media (links between the routers) and data-link layer protocols. Add this raw IP value to that of the overhead to determine the link speeds needed to support this number of calls. Note this value represents only the bearer (voice) content.

Signaling bandwidth requirements vary depending on the rate at which the calls are generated and the signaling protocol used. If a large number of calls are initiated in a relatively short period, the peak bandwidth needs for the signaling could be quite high. A general guideline for the maximum bandwidth requirement that an IP signaling protocol needs is roughly three percent of all bearer traffic. Using the previous example, signaling bandwidth requirements, if all 2,000 calls were initiated in one second, would be approximately 4.8 Mbps (3 percent of 160-megabits).

With the calculation of bearer and signaling, the total bandwidth needed to support 2000 G.711 encoded calls would approximately be a maximum of 164.8 MB. This bandwidth requirement is a theoretical maximum for this specific case. If the parameters change, such as call initiation rate, voice encoding method, packet creation rate, employment of compression and silence suppression, the bandwidth requirements would change as well.

With large VoIP implementations requiring sizable bandwidth, it becomes imperative that the IP network delivers the needed service at predictably high performance.

Packet Loss

Packet loss occurs for many reasons, and in some cases, is unavoidable. Often the amount of traffic a network is going to transport is underestimated. During network congestion, routers and switches can overflow their queue buffers and be forced to discard packets. Packet loss for non-real-time applications, such as Web browsers and file transfers, is undesirable, but not critical. The protocols used by non-real-time applications, usually TCP, are tolerant to some amount of packet loss because of their retransmission capabilities.

Real-time applications based on UDP are significantly less tolerant to packet loss. UDP does not have retransmission facilities, however, retransmissions would almost never help. In an RTP session, by the time a media gateway could receive a retransmission, it would no longer be relative to the reconstructed voice waveform; that part of the waveform in the retransmitted packet would arrive too late.

It is important that bearer and signaling packets not be discarded, otherwise, voice quality or service disruptions might occur. In instances where service disruptions may occur, Class of Service (CoS) mechanisms offer a means of controlling packet delivery priority. , which is equivalent to DSCP in IP, but at the Ethernet layer, mechanisms become very important. By configuring CoS parameters, administrators can give packets of greater importance a higher priority in the network, thus ensuring packet delivery for critical applications, even during times of network congestion. Note that CoS is equivalent to DSCP in IP, but at the Ethernet layer.

Although packet loss of any kind is undesirable, some loss can be tolerated. Some amount of packet loss for voice services could be acceptable, as long as the loss is spread out over a large amount of users. As long as the amount of packet loss is less than five percent for the total number of calls, the quality generally is not adversely affected. It is best to drop a packet, versus increasing the latency of all delivered packets by further buffering them.

Reliability

Although network failures are rare, planning for them is essential. Failover strategies are desirable for cases when network devices malfunction or links are broken. An important strategy is to deploy redundant links between network devices and/or to deploy redundant equipment. To ensure continued service, organizations should plan carefully for how media gateways and media gateway controllers can make use of the redundant schemes.

IP networks use routing protocols to exchange routing information. As part of their operation, routing protocols monitor the status of interconnecting links. Routing protocols typically detect and reroute packets around a failure if an alternate path exists. Depending on the interconnecting media used for these links, the time taken to detect and recalculate an alternate path can vary. For example, the loss of signal for a SONET/SDH connection can be detected and subsequently rerouted very quickly. However, a connection through an intervening LAN switch might need to time out the keep-alive protocol before a failure is detected.

Having media gateways and media gateway controllers that can actively detect the status of their next-hop address (default gateway) as part of their failover mechanism decreases the likelihood of a large service disruption. Another possible option is that the media gateway and media gateway controller could be directly connected to the router. In this case, the possibility of a link failure (depending on the nature of the failure) could be immediately detected and the network devices would take appropriate action. Still another option for reducing long-term failure could be to employ a redundancy mechanism such as failover.

Security

Security, especially in a converged voice and data network, is a high priority. Organizations need to protect the voice communication devices from unauthorized access and malicious attack. While organizations can thwart unauthorized access by using security protocols (such as RADIUS and SSH), Denial-of-Service (DoS) attacks can be a real danger to voice services. It is conceivable that such attacks would either cripple or completely disable voice services.

One way to secure VoIP devices is to use private addressing to enumerate the media gateways and call processing servers. Private addressing is not advertised to the public Internet and, therefore, the devices are inaccessible to the outside world.

Additionally, all VoIP processing servers, gateway, and messaging systems should be placed behind firewalls to enforce access control policies and protect them from any DoS attacks. The firewall needs to understand the signaling protocols in use in the network to be able to dynamically open and close ports for the VoIP traffic only for the duration of a call, so that these ports are not left open and cannot be usurped for unauthorized use. These servers are critical to VoIP communication; therefore, firewall policies should be in place to protect communications between these servers and VoIP end-devices. These policies should restrict VoIP communication, based on authorized end-devices or traffic sourced or destined for a particular IP address or interface. Firewalls can be used to segment the VoIP network, separating the voice traffic from other traffic to ensure appropriate priority and policies are applied. Firewalls may also be placed to mitigate DoS attacks and to create logs for forensics. Furthermore, intrusion prevention systems can be deployed to help detect and prevent certain attacks, such as manipulated DHCP messages or flooded FIB tables.

Introduction

The market for information and communications technology is currently undergoing a structural change. The classic telecommunication networks were planned and implemented for the transfer of specific data such as telephone calls or pure data packages. The recent growth in competition, new requirements for the market and technological developments have fundamentally changed the traditional attitudes of the telecommunications industry. The present industry is characterized by the rapid growth of broadband connections, the convergence processes of various network technologies and the emergence of a uniform IP standard for individual and mass communications. Traditional telecommunications operators find themselves confronted with a host of new challenges. In particular, their previously successful fixed-network business is coming increasingly under pressure. New communication possibilities, such as telephoning via the Internet, and also growing market shares in mobile telephony are causing a great deal of concern. To counteract these losses, the network operators are investing more strongly in the growth driver, broadband. The bundling of phone, Internet and television – known in the telecommunications industry as Triple Play Services – has moved into the limelight of these new business models. The traditionally familiar market boundaries between fixed networks, mobile telephony and data networks are disappearing more and more quickly. This gives the customer the advantage that he can call on an extremely wide range of services, regardless of his access technology. This development requires a metainfrastructure beyond the existing, subordinated networks – a core network for all the access networks. This new network is called the Next Generation Network. The Internet Protocol is the most significant integration factor because it is available globally and, at least in principle, it can use almost all the services and applications in all the networks.

Next Generation Network.
Definition

The International Telecommunication Union (ITU) – a global organization dedicated to technical aspects of telecommunications – regards an NGN as a network with an end-to-end service for voice, data and multimedia. The deployed transmission technologies must permit a high quality of service. At the same time, the provided service functions are independent from the basic transmission functions. All the services are based on the Internet Protocol (IP). The advantage of IP networks is their flexibility and the simple integration of new applications.

System Architecture

The basic premise for NGN is an architecture on several independent levels. These include the access area, the core network area, the control level and the service management level (see Fig. 1). The connection of subscribers and terminals to the NGN can be achieved with various access technologies. The information and transmission formats of the various networks must be converted into information that is comprehensible for the NGN. This calls for Gateways for the connection of business and private customers. The core network of the NGN is an IP network. This is a standardized transport platform consisting of various IP routers and switches. The connection control of the individual components is carried out by the control level. Standard and value-added services can then be provided via the service management level.

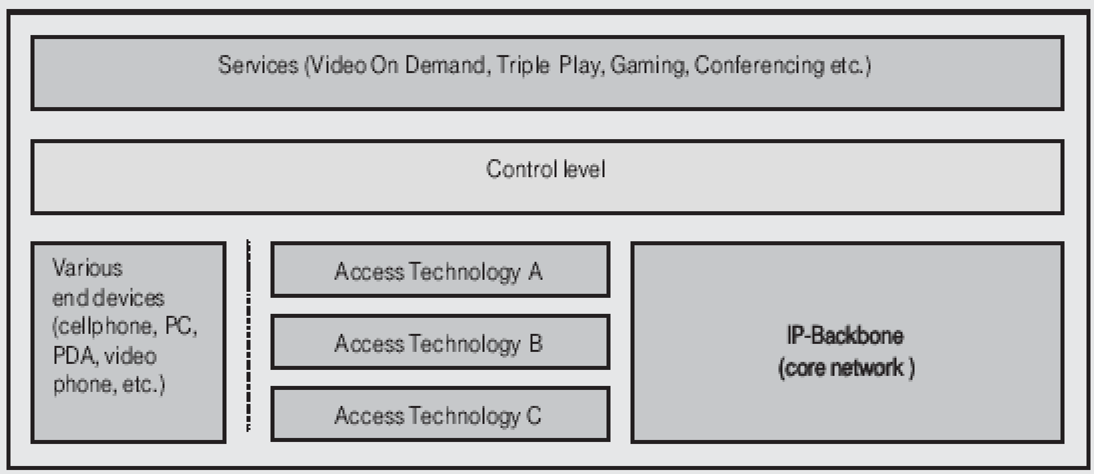


Figure 1: Modular structure of an NGN

The aim of an NGN is to operate the current wide range of access and communications technologies under a common umbrella in the future network on IP. This convergence allows a transition from a vertical to a horizontal service integration. In vertical network structures, services (e.g. phone services, TV services) can only be received with suitable networks and the relevant end devices. With a horizontal approach, on the other hand, users in future will be given the possibility of using the desired services – regardless of the platform and the technology – with a single end device. Figure 2 shows this basic premise.

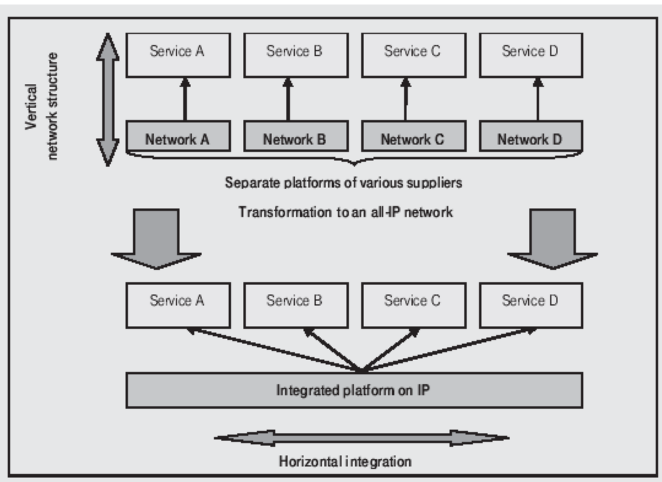


Figure 2: All-IP network

Technologies for Subscriber Access.

The prerequisite for the efficient use of the NGN is a network access with high bandwidth for the subscriber. It is to be expected that access networks in the future will be able to provide bandwidths of up to 100 MBit/s for private subscribers. For business customers, transmission rates in the gigabyte range are conceivable. The possible options for access to the IP backbone would include copper, cable, fiber-optic and wireless connections.

- Copper connections: The bandwidth of DSL via copper lines is restricted to 16 MBit/s for ADSL2 (Asynchronous Digital Subscriber Line) and to 52 MBit/s for VDSL (Very High Speed Digital Subscriber Line). Speeds decrease in relation to the distance to the access Gateway. The Gateway connects the respective network to the backbone.
- Cable connections: The Hybrid Fiber Coaxial Networks (HFC) are networks that consist of a mixture of fiber-optic cables and coaxial cables. As a rule, they are networks for cable television.

- Fiber-optic lines (FTTX): Fiber-optic lines can transport large data volumes at extremely high transmission speeds. In practice, bandwidths up to 155 MBit/s are possible. The different types of fiber-optic connections are referred to generically by the acronym FTTx (Fiber to the x).

- Wireless connections: Radio Access Networks (RAN) are radio-based access networks. They can be implemented with a wide variety of technologies such as 3GPP, 3GPP2 (UMTS), WiFi or WiMAX.

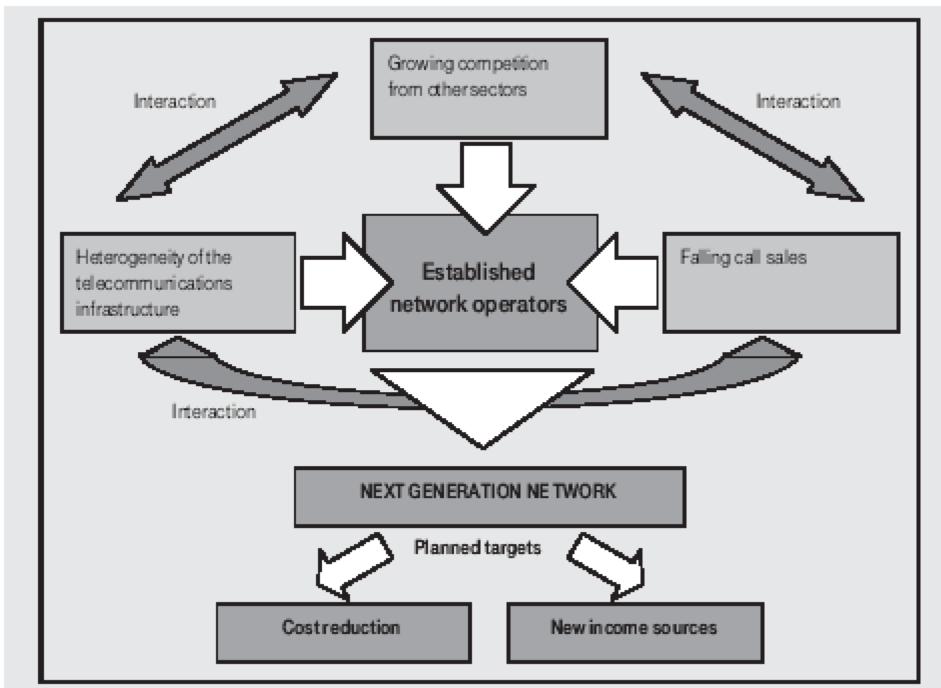
The following different forms of FTTX are available:

- FTTH: Fiber to the Home – the fiber optics end at the broadband connection.
- FTTB: Fiber to the Basement – the fiber optics end at the building (also known as FTTMDU –
  Fiber to the Multi-Dwelling Unit)
- FTTC: Fiber to the Curb – the fiber optics end at the street.
- FTTA: Fiber to the Area – the fiber-optic connection supplies a relatively large area.

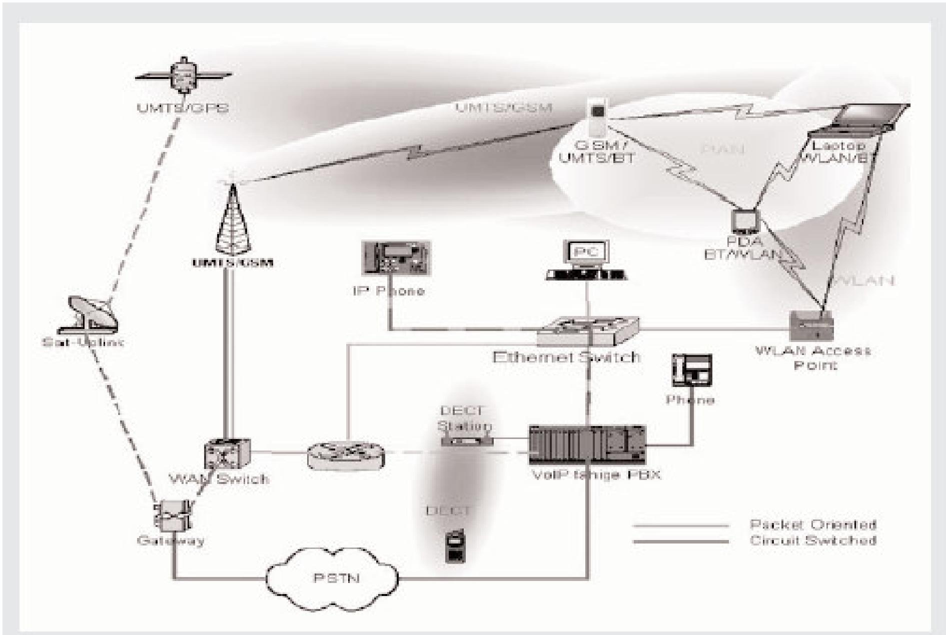
Thanks to all these access technologies, it is possible for the operators to guarantee services of the next generation, given the appropriate investments. The fiber-optic lines offer the highest possible capacity for the transmissions. In spite of the reduced prices for optical transmission technology, the installation of fibre-optic accesses is not yet profitable for the end subscriber. The high costs for the network operators are due to the laying of the cables and the connection technology. At present, the use of this technology is therefore mainly limited to occasional business customer connections. The provision of comprehensive broadband services varies from region to region. In urban environments, it may well be economically viable in the future for network operators to offer customers FTTB or even FTTH. In rural areas with a sparse population, on the other hand, the use of fiber optics is not profitable. In regions like that, the best solution will be to use broadband accesses via radio.

Motivation for NGN.

The heterogeneity of the infrastructure, the growing competition and the falling call sales can be regarded at present as the primary threats to the telecommunications industry. Established etwork operators are finding themselves forced to rethink their business models and to convert their infrastructure to a fully IP-based platform – the Next Generation Network. The overall aim is to reduce costs and to create new sources of income



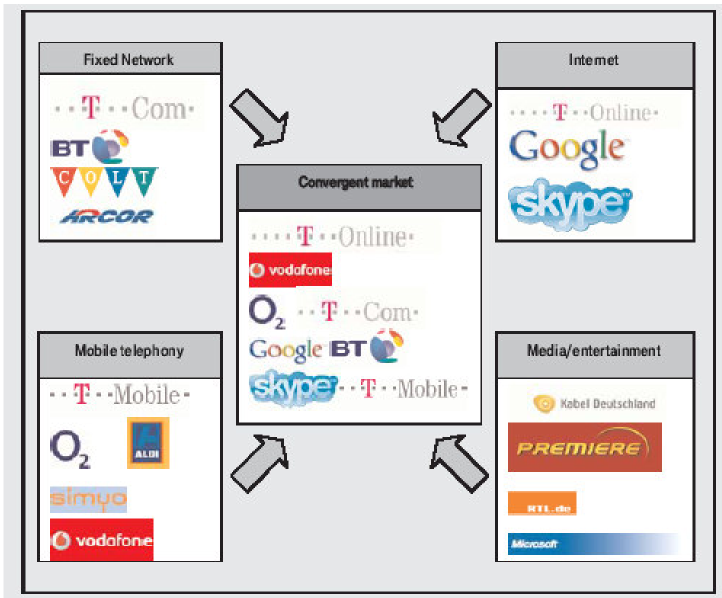
Heterogeneity of the Telecommunications Infrastructure. The modern telecommunications Networks consist of satellite and mobile phone networks such as GSM/UMTS, public phone networks and wireless local traffic networks such as wireless LAN and Bluetooth networks. The latter connect devices in the personal work environments such as PDAs, laptops and cellphones. There are also cabled fixed networks such as Ethernet and also fiber-optic networks



In the traditional network infrastructure, the introduction of new services and applications can be an arduous and expensive process. For instance, a concept for launching innovative services can take between 6 and 18 months. The process requires high staffing costs. Many functionalities in the network have to be configured manually in order to implement new features. Moreover, the variety of networks and the heterogeneous subscriber end devices make the provision of infrastructure-independent services more difficult. As a result, the services can only be used via specific networks and appropriately adjusted end devices such as fixed-network phones, cellphones, televisions, etc. The growing number of services has led to an increase in the platforms needed to provide them, which in turn has increased the complexity of the overall infrastructure. The problems of interoperability between the various systems are becoming more serious, and this growing complexity is also placing greater demands on staff. Maintaining these platforms involves high annual operating costs for the network operators. Established network operators often maintain 15 to 20 different platforms with hundreds of central switches, which inevitably leads to extremely high staffing costs.

Growing Competition from Other Sectors.

As a rule, networks such as mobile telephony, data networks and fixed networks are dominated by different suppliers. Providing services and products in these networks requires an interaction of various, complementary elements. In this sense, it is necessary to differentiate between value-added levels such as hardware, network access, applications and content. The increased use of IP-based networks for the provision of applications and services is allowing the development of new, digital value-added chains. Visions of the gradual convergence of fixed networks, mobile telephony and the Internet are having a crucial influence on the development of this sector. In the future market, the widest possible range of roles will be available for different players. This will particularly threaten the leading position of the established network operators on the Telecommunications market. Apart from the fixed-network and cellphone operators, companies from other sectors will also establish themselves in future on this convergent market. Portal suppliers with strong brand names and powerful financial backing – including Google, MSN, eBay and Yahoo – are planning to penetrate the voice and infrastructure business. They will also be joined by cable network operators and companies that provide media content, such as Microsoft, Kabel Deutschland or Premiere



This convergence is therefore producing virtually inevitable conflicts and incompatibilities. Technologies and market forces are colliding with each other. The market participants are crowding each other out and defending their positions strongly. According to the British media watchdog, Ofcom, a basic change in the familiar competitive structures is to be anticipated in the next 5 to 10 years. In the course of this convergence, the value of the network business will gradually decrease and the service range will make a much larger contribution to end-customer sales. Traditional network operators will have to rethink their business model and also position themselves much more strongly on the upper levels of the value-added chain.

Falling Call Sales.

The increasing competition due to the liberalization of the markets and the arrival of market participants from other sectors are causing great concern to the operators of former state monopolies. The classic telephone business, known as a Public Switched Telephone Network (PSTN), is particularly unsatisfactory. The golden age of the high-margin business with revenue in the billions based on classical phone calls is clearly over. Figure 6 shows the estimated development of the global number of telephone minutes since 1990. In spite of the current fall in fixed-network minutes, a strong growth in the total of telephone minutes is to be expected. Experts see particularly strong potential in the use of the Internet Protocol for phone calls. This so-called Voice over IP (VoIP) is possible with all IP-based networks.

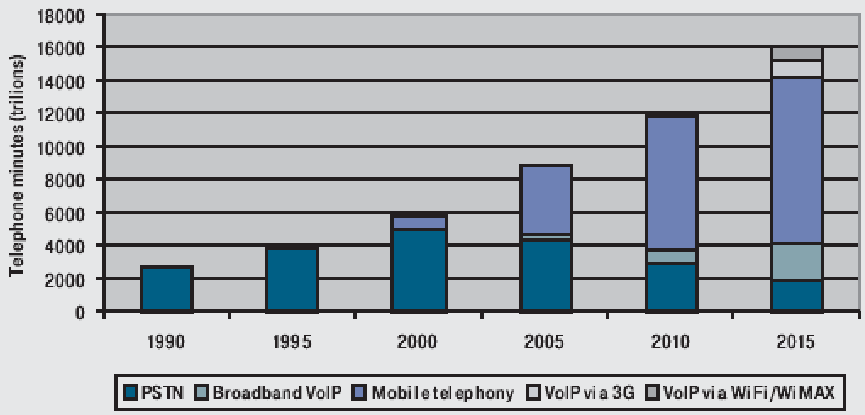


Figure 6: Development of global telephone minutes
Source: With reference to Ovum (2006a), P. 3.

In a current study from 2006, the consultants from Mercer have investigated in great detail the fixed-network market in Western Europe with a total volume of € 114 billion ($ 144 billion). It is particularly striking that the traditional Western European fixed-network operators are increasingly losing market shares in spite of a sustained fall in prices. There has been a decline in sales of approx. 5 percent since 2001 in the telephone business of the former monopoly holders. During the same period, their total market share fell from 70 to 60 percent. That corresponds to a loss of € 4 billion ($ 5 billion) per year. According to Mercer, an even worse assessment can be expected for the coming years with increasing losses due to more intense competition.

While fixed-network calls are stagnating in Germany, mobile telephony is enjoying strong growth. In 2004, the total sales with mobile telephony amounted to over € 20 billion ($ 25 billion) according to the market research analyst Gartner. Market penetration of over 90 percent is expected for Germany in 2006. According to analysts such as Gartner and Ovum, mobile telephony will increasingly replace classical fixed-network phones. Fixed-network operators are afraid of widespread cancellations of fixed-network connections.

Increasing losses on the domestic fixed-network market are therefore forcing the operators to develop new strategies to secure their future and to boost their profitability. No further growth can be expected through call sales alone.

Planned Targets – Cost Reductions and New Sources of Income. Established network operators are pursuing two basic goals with NGN. On the one hand, the optimization of the networks and technology should open up excellent potential for cost savings. On the other hand, they intend to exploit new income sources with the future network. The plan is to create an entirely new form of communication for the customers.

Cost reduction.

With NGN, the established network operators plan to develop a sustainable infrastructure that will remain competitive in a convergent environment. The primary focus will be on the potential for cost savings. These savings will be produced by focusing on a single technology system and through the related reduction in technology sites and technical equipment areas. A single infrastructure is easier to maintain. The simplification of the technology system will therefore promote a reduction in the staffing costs. Moreover, spare parts will only be necessary for a single form of network technology.

Furthermore, the modular structure of the NGN will provide the foundation for the simple and cost-effective development of future services. It will no longer be necessary to carry out the new development and installation of networks for specific services. The open platform will also allow the rapid implementation of customer-specific solutions. For instance, applications from the network operators and other specialists can be inserted more easily in the standardized NGN architecture using Service Creation Environments. Predefined library functions will be used via an Application Programming Interface (API) to activate a Gateway and so ultimately to carry out actions in the network.

The migration to a homogeneous IP platform that supports all services will permit annual cost savings of up to 30 percent, according to Detecon Consulting. According to the specifications of the telecom equipment manufacturer Lucent Technologies, this will represent savings of several billion euros per annum for major, established network operators.
The market research company Ovum has pointed out that it will take some time before the cost-reduction potential
becomes noticeable due to more efficient network management. The procedure will take several years. Apart from anything else, the technical equipment will have to be replaced at all the exchanges in the entire national network. As well as that, the employees will have to be retrained to work on the new network environment. A relatively long period of parallel operation with the already existing, mostly PSTN-based networks will be necessary before they can gradually be replaced by IP. The services provided via traditional networks will have to be provided for a certain period of time through emulation or simulation. Users will be able to continue using their present end devices. Even so, appropriate end devices will have to be developed to use all the functionalities of the forthcoming new services.

New Sources of Income.

Established network operators see the possibility of new income as another motivation for promoting NGN. More and more innovations with new sales opportunities are expected in the field of value-added services. The market development features a range of telecommunications services that have been tried and tested or are still evolving. For instance, these include television, information services, tele-learning and teaching, online games, virtual reality, business-to-business services, business TV, videoconferencing, etc.

However, opinions vary on the level of this income. The emerging price models will have a considerable influence on the generation of new sales. In an all-IP world, there is little correlation between the volumes on offer and the price. This can be seen in the familiar flat-rate tariffs in the broadband sector. In spite of the unlimited transmission volumes, the prices remain relatively stable. Ovum therefore anticipates that the service curve for the NGN environment will flatten off. That means that new sales via NGNs will be fairly restricted in the near future. In contrast, various manufacturers such as Alcatel or Siemens are arguing that only the introduction of innovative services will allow established network operators to increase their profitability. They claim that established network operators will be able to double their average revenue per user (ARPU) and to reduce customer migrations by 40 percent, among other things. As a result, the additional investments in this future technology will pay for themselves in less than five years. In this context, however, we must refer back to the flop with UMTS. Established network operators such as Deutsche Telekom invested billions to acquire the licenses alone, which are not remotely profitable even today.

The Market – Convergence Approaches and Needs of the User.

Initial Convergence Approaches.

The market already features individual examples of a general trend toward the convergence of various technologies, communications channels and media. Particularly remarkable is VoIP, which has developed strongly in the last two years, with its use of the Internet for phone calls (which was not actually designed for this purpose). It is not clear to the user that he is using a different network infrastructure from previously for this voice transmission service. This also allows entirely new service features to be offered, such as e.g. the setting up of phone connections from WWW applications.

The gradual merging of fixed networks and mobile telephony networks (Fixed Mobile Convergence, FMC) is another essential phenomenon of this convergence. The FMC approach caters for availability at any location – either stationary or mobile – using a single phone number. The IP network is used to provide the stationary use. Moreover, the subscriber has just one voice mailbox and receives a single bill. Ovum estimates the current number of users of such an end device worldwide at less than 100,000. The primary advantage for the user is to save expensive cellphone costs as soon as he is within range of a wireless LAN hotspot.

The most topical business model is certainly Triple Play: The customer receives voice, Internet, television and video services in a bundle via a single line. TV cable networks, conventional telephone networks and mobile telephony networks are suitable for this service. High bandwidths with excellent reliability are indispensable for providing this large number of services in parallel. If it is ultimately possible to link offers with attractive contents and prices with technical innovations that traditional television cannot provide, Triple Play will have the potential to be an extremely profitable business model.

At the end of the day, the network convergence will also lead to a convergence of the end devices, depending on the actual needs. Multimedia-compatible computers will be given telephone and video communication functions, data services will be available by telephone and Internet access via the television (browsing using an Internet-compatible setup box) and the cellphone will be common.

Market Needs in Terms of Convergent Services. Beneficial Effects for the Customer.

The interaction of man and technology plays a crucial role in the introduction of previously unknown technologies on the market. The essential prerequisite for the success of innovative information and communications systems is their acceptance by the customers. Characteristics such as the perceived system benefit and the user-friendliness of the technology are extremely important.

One of the desired goals of NGN is the possibility of adapting the services better to the needs of the customer. Due to the future restriction to a single end device – equipped with a wide range of applications and services – the customer will in many ways enjoy improvements on the current situation. At present, customers expect applications for telephony and conferences. This sort of application should be independent of the network type. Customers also want to have more control over their services.

That includes the ability to easily change or add services, regardless of location. Above all, though, the primary focus is on the wish to reduce costs and so there is great interest in package prices. That includes the ability to easily change or add services, regardless of location. Above all, though, the primary focus is on the wish to reduce costs and so there is great interest in package prices.

Control: Current processes require a personal communication with the customer for the activation or deactivation of services. NGNs should give the customer more control over his own service portfolio through online interfaces, such as webpages, for instance. In this way, network operators and service providers will save processing costs and the services will be provided for the customer in real time.

Omnipresent: The term “presence” is frequently used in the mobile world and describes the personalization of services. Personalization characterizes the individual customizing of services to a specific user, in contrast to uniform standard services (e.g. the analog telephone service). Moreover, the services should be provided regardless of the location. The network must detect with which end device the user is currently connected to the net and where he is currently located. His subscribed services are then provided to him regardless of his location.

Flexible billing methods: It will be possible for network operators to charge for scaled services via the NGN. For instance, the customer could be provided with only “best-effort” broadband services for surfing on the Web, but he could also use a much higher bandwidth with QoS parameters on request, to guarantee the required quality. Additional costs may be incurred when downloading a movie, which are automatically integrated in the customer’s bill.

It is therefore to be expected that the perceived benefits – especially because of increasing flexibility, mobility and convenience – will grow as convergent services become more widespread. The increasing personalization of the services will also significantly influence the perceived benefits. The information and services provided will be customized to suit each customer’s personal context. However, it remains to be seen to what extent applications and services can be used with a single end device without any particular technical knowledge. Real growth spurts can be expected especially once a clear, tangible added value is perceptible without any particular complexities and also the majority of the market segments are being addressed. The user-friendliness is a decisive factor particularly for older people. The variety of services must not be too heavily technical, complex or unclear. In the end, the successful interaction between man and technology often proves to be much more difficult than anticipated.

Market Needs in Europe.

The interest of consumers in convergent services is still split at present. About 26,000 consumers in France, Germany, Italy, Holland, Spain, Poland, Sweden and Great Britain were questioned in a survey by Forrester Research. About 35 percent of the participants answered that they would be interested in package offers of voice, video and data services. However, around 44 percent of the consumers said that they were not interested in Triple Play at all. The demand varied across the different European countries

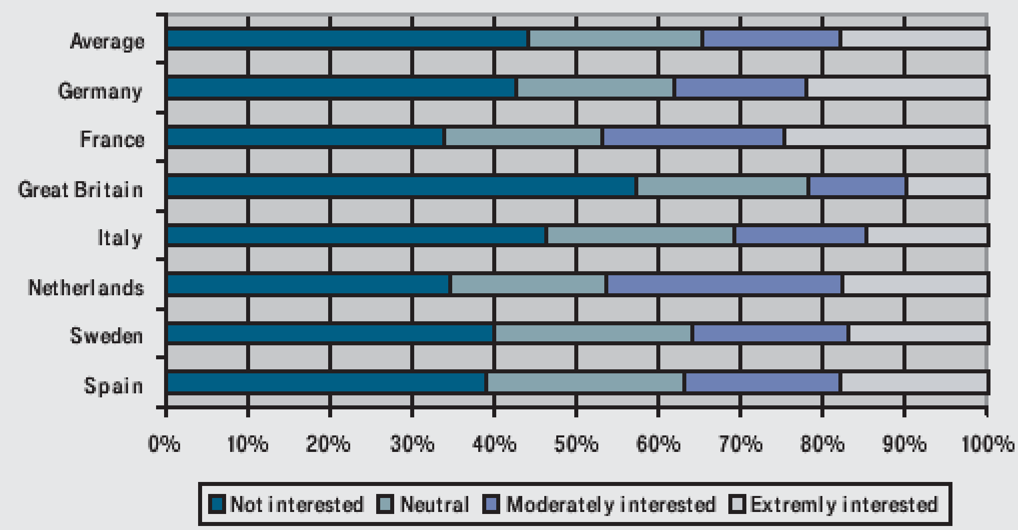


Figure 7: Interest in Triple Play
Source: With reference to Forrester (2006c), P. 3.

Consumers who are interested in Triple Play possess the demographic features shown in Table 1. The average age of 40 years registered here is relatively high. Normally, it would be expected that a younger generation would be particularly interested in Triple Play, because older people in particular tend to have a negative attitude towards change and so prefer to keep the status quo.

	Interest in Triple Play
Average Age	40
Higher education	31%
Higher income	40%
Monthly fixed-network costs	€ 37
Monthly cellphone costs	€ 25
Monthly internet costs	€ 22
Monthly television costs	€ 28
Preferred Triple Play supplier	Fixed-network operator 33%
Already own Pay TV	31%
Use of interactive TV services (sports channels)	17%
Use of interactive TV services (Video On Demand, VOD)	12%

Table 1: Demographic features of prospective Triple Play customers
Source: Cf. Forrester (2006c), P. 5.

This survey indicated that the decisive factors for a customer to choose Triple Play included monthly cost savings, a single bill for all services, higher quality, three months' trial use, a contract partner for the Customer Service and a change of supplier

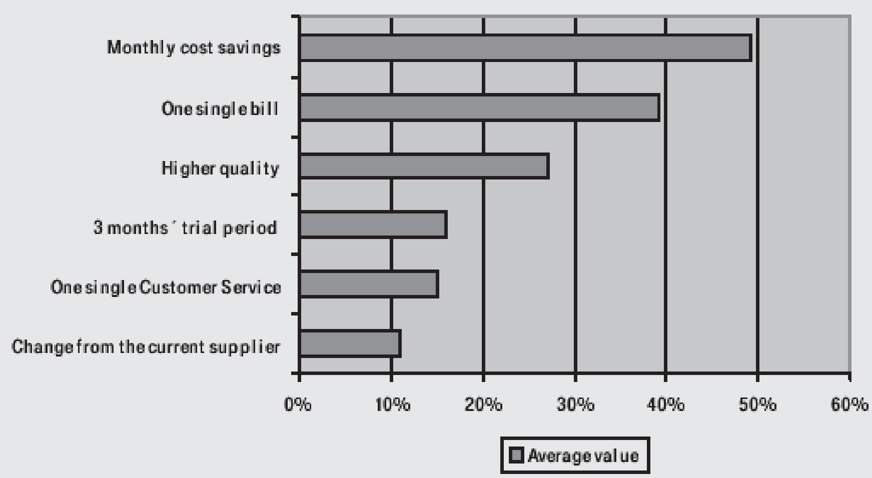


Figure 8: Decisive factors for the choice of Triple Play
Source: With reference to Forrester (2006c), P. 4.

Some analysts doubt the possibility for telecommunications companies to earn profits with Triple Play services. Lars Godell from Forrester Research explained that, although customers find Triple Play offers attractive because they are cheaper overall than the individual connections, they are not necessarily willing to pay more for additional services. It is questionable whether the necessary major investments in infrastructure (e.g. FTTC/FTTH) would really pay off given the intense competition and the current lack of demand. According to the comments of Mr. Godell, network operators would have to invest several hundred euros per month in each individual new customer. At the same time, however, the sale prices of Triple Play packages would be no more than € 50 ($ 63) per month. For example, the net earnings from IPTV through 2016 will level off at approx. € 11 ($ 14) per user per annum. Against all expectations, this total represents just 2.4 percent of the average total earnings for each customer. Access and communications via broadband will continue in future to generate the lion's share with 87 percent of the total earnings.

Softswitch

A **softswitch** is a central device in a telephone network which connects calls from one phone line to another, entirely by means of software running on a computer system. This work was formerly carried out by hardware, with physical switchboards to route the calls.

A softswitch is typically used to control connections at the junction point between circuit and packet networks. A single device containing both the switching logic and the switching fabric can be used for this purpose; however, modern technology has led to a preference for decomposing this device into a Call Agent and a Media Gateway.

The Call Agent takes care of functions like billing, call routing, signalling, call services and so on and is the 'brains' of the outfit. A Call Agent may control several different Media Gateways in geographically dispersed areas over a TCP/IP link.

The Media Gateway connects different types of digital media stream together to create an end-to-end path for the media (voice and data) in the call. It may have interfaces to connect to traditional PSTN networks like DS1 or DS3 ports (E1 or STM1 in the case of non-US networks), it may have interfaces to connect to ATM and IP networks and in the modern system will have Ethernet interfaces to connect VoIP calls. The call agent will instruct the media gateway to connect media streams between these interfaces to connect the call - all transparently to the end-users. The softswitch generally resides in a building owned by the telephone company called a central office. The central office will have telephone trunks to carry calls to other offices owned by the telephone company and to other telephone companies (aka the Public Switched Telephone Network or PSTN). Looking towards the end users from the switch, the Media Gateway may be connected to several access devices. These access devices can range from small Analog Telephone Adaptors (ATA) which provide just one RJ11 telephone jack to an Integrated Access Device (IAD) or PBX which may provide several hundred telephone connections. Typically the larger access devices will be located in a building owned by the telephone company near to the customers they serve. Each end user can be connected to the IAD by a simple pair of copper wires. The medium sized devices and PBXs will typically be used in a business premises and the single line devices would probably be found in residential premises. In more recent times (i.e., the IP Multimedia Subsystem or IMS), the Softswitch element is represented by the Media Gateway Controller (MGC) element, and the term "Softswitch" is rarely used in the IMS context.

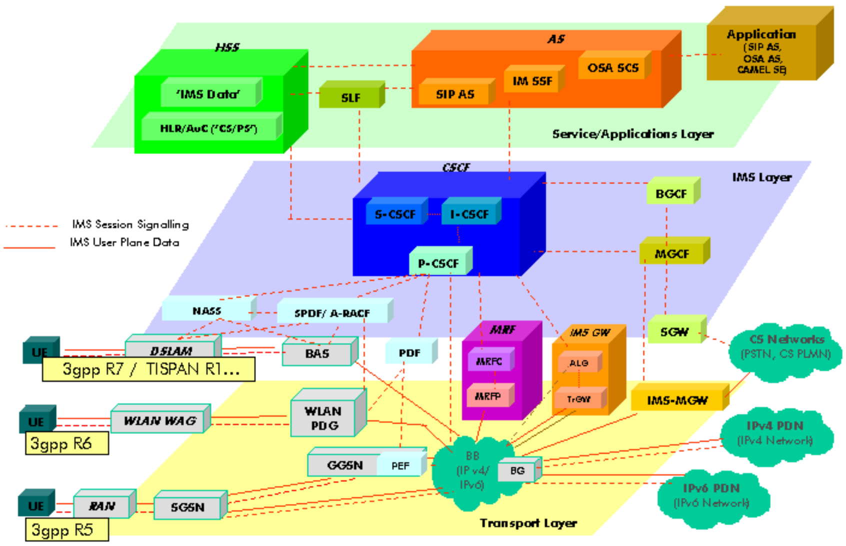
Feature server as a part of softswitch

The feature server, often built into a call agent/softswitch, is the functional component that provides call-related features. Capabilities such as call forwarding, call waiting, and last call return, if implemented in the network, are implemented in the feature server. The feature server works closely with the call agent, and may call upon the media server to provide these services. These features do not require the subscriber to explicitly request them but tend to be triggered within the call handling logic. An example of a feature service is last call return, in which the user picks up the phone, dials *69, and hears, "*The number that last called you was xxx-xxx-xxxx. Press 1 to return this call.*" When the call agent sees the dial string *69, it triggers an invocation of the feature server function. The feature server examines its database, finds the user and the caller identification of the last call, then asks the media server to play the announcement and collect a digit. When the media server returns a "1", the feature server instructs the call agent to establish a call between the user and the party that last called that user.

IP Multimedia Subsystem

The **IP Multimedia Subsystem** (**IMS**) is an architectural framework for delivering internet protocol (IP) multimedia to mobile users. It was originally designed by the wireless standards body 3rd Generation Partnership Project (3GPP), and is part of the vision for evolving mobile networks beyond GSM. Its original formulation (3GPP R5) represented an approach to delivering "Internet services" over GPRS. This vision was later updated by 3GPP, 3GPP2 and TISPAN by requiring support of networks other than GPRS, such as Wireless LAN, CDMA2000 and fixed line. To ease the integration with the Internet, IMS as far as possible uses IETF (i.e. Internet) protocols such as Session Initiation Protocol (SIP). According to the 3GPP[1], IMS is not intended to standardise applications itself but to aid the access of multimedia and voice applications across wireless and wireline terminals, i.e. aid a form of fixed mobile convergence (FMC). This is done by having a horizontal control layer that isolates the access network from the service layer. Services need not have their own control functions, as the control layer is a common horizontal layer. Alternative and overlapping technologies for access and provision of services across wired and wireless networks depend on the actual requirements, and include combinations of Generic Access Network, soft switches and "naked" SIP. This makes the business use of IMS less appealing. It is easier to sell services than to sell the virtues of "integrated services". But, services for IMS have not been prolific. Since IMS was conceived years ago, it is becoming increasingly easier to access content and contacts using mechanisms outside the control of traditional wireless/fixed operators, and so those operators are likely to reconsider their strategies[2]. Although it is expected that eventually IP will be available on all mobile phones and operators, it is not clear how much of the 3GPP/3GPP2/TISPAN IMS as it exists today will be deployed.

Architecture



3GPP / TISPAN IMS Architectural Overview

The IP Multimedia Core Network Subsystem is a collection of different functions, linked by standardized interfaces, which grouped form one IMS administrative network. A function is not a node (hardware box): an implementer is free to combine 2 functions in 1 node, or to split a single function into 2 or more nodes. Each node can also be present multiple times in a single network, for load balancing or organizational issues.

Access network

The user can connect to an IMS network in various ways, all of which use the standard Internet Protocol (IP). Direct IMS terminals (such as mobile phones, personal digital assistants (PDAs) and computers) can register directly on an IMS network, even when they are roaming in another network or country (the visited network). The only requirement is that they can use IPv6 (also IPv4 in early IMS) and run Session Initiation Protocol (SIP) user agents. Fixed access (e.g., Digital Subscriber Line (DSL), cable modems, Ethernet), mobile access (e.g. W-CDMA, CDMA2000, GSM, GPRS) and wireless access (e.g. WLAN, WiMAX) are all supported. Other phone systems like plain old telephone service (POTS -- the old analogue telephones), H.323 and non IMS-compatible VoIP systems, are supported through gateways.

Core network

The Home Subscriber Server (HSS), or User Profile Server Function (UPSF), is a master user database that supports the IMS network entities that actually handle calls. It contains the subscription-related information (user profiles), performs authentication and authorization of the

user, and can provide information about the user's physical location. It is similar to the GSM Home Location Register (HLR) and Authentication Centre (AUC).
An SLF (*Subscriber Location Function*) is needed to map user addresses when multiple HSSs are used. Both the HSS and the SLF communicate through the DIAMETER protocol.

User identities

Normal 3GPP networks use the following identities:

International Mobile Subscriber Identity (IMSI)

Temporary Mobile Subscriber Identity (TMSI)

International Mobile Equipment Identity (IMEI)

Mobile Subscriber ISDN Number (MSISDN)

IMSI is a unique phone identity that is stored in the SIM. To improve privacy, a TMSI is generated per geographical location. While IMSI/TMSI are used for user identification, the IMEI is a unique device identity and is phone specific. The MSISDN is the telephone number of a user. IMS also requires IP Multimedia Private Identity (IMPI) and IP Multimedia Public Identity (IMPU). Both are not phone numbers or other series of digits, but Uniform Resource Identifier (URIs), that can be digits (a tel-uri, like *tel:+1-555-123-4567*) or alphanumeric identifiers (a sip-uri, like *sip:john.doe@example.com*). There can be multiple IMPU per IMPI (often a tel-uri and a sip-uri). The IMPU can also be shared with another phone, so both can be reached with the same identity (for example, a single phone-number for an entire family). The HSS user database contains, the IMPU, IMPI, IMSI, and MSISDN and other information.

Call/session control

Several roles of Session Initiation Protocol (SIP) servers or proxies, collectively called Call Session Control Function (CSCF), are used to process SIP signalling packets in the IMS.

- A Proxy-CSCF (P-CSCF) is a SIP proxy that is the first point of contact for the IMS terminal. It can be located either in the visited network (in full IMS networks) or in the home network (when the visited network isn't IMS compliant yet). Some networks may use a Session Border Controller for this function. The terminal discovers its P-CSCF with either DHCP, or it is assigned in the PDP Context (in General Packet Radio Service (GPRS).

- it is assigned to an IMS terminal during registration, and does not change for the duration of the registration
- it sits on the path of all signalling messages, and can inspect every message

- ➢ it authenticates the user and establishes an IPsec security association with the IMS terminal. This prevents spoofing attacks and replay attacks and protects the privacy of the user. Other nodes trust the P-CSCF, and do not have to authenticate the user again.
- ➢ it can also compress and decompress SIP messages using SigComp, which reduces the round-trip over slow radio links
- ➢ it may include a Policy Decision Function (PDF), which authorizes media plane resources e.g. quality of service (QoS) over the media plane. It's used for policy control, bandwidth management, etc. The PDF can also be a separate function.
- ➢ it also generates charging records

- A Serving-CSCF (S-CSCF) is the central node of the signalling plane. It is a SIP server, but performs session control too. It is always located in the home network. It uses DIAMETER Cx and Dx interfaces to the HSS to download and upload user profiles — it has no local storage of the user. All necessary information is loaded from the HSS.

- ➢ it handles SIP registrations, which allows it to bind the user location (e.g. the IP address of the terminal) and the SIP address
- ➢ it sits on the path of all signaling messages, and can inspect every message
- ➢ it decides to which application server(s) the SIP message will be forwarded, in order to provide their services
- ➢ it provides routing services, typically using Electronic Numbering (ENUM) lookups
- ➢ it enforces the policy of the network operator
- ➢ there can be multiple S-CSCFs in the network for load distribution and high availability reasons. It's the HSS that assigns the S-CSCF to a user, when it's queried by the I-CSCF.

- An I-CSCF (*Interrogating-CSCF*) is another SIP function located at the edge of an administrative domain. Its IP address is published in the Domain Name System (DNS) of the domain (using NAPTR and SRV type of DNS records), so that remote servers can find it, and use it as a forwarding point (e.g. registering) for SIP packets to this domain. The I-CSCF queries the HSS using the DIAMETER Cx interface to retrieve the user location (Dx interface is used from I-CSCF to SLF to locate the needed HSS only), and then routes the SIP request to its assigned S-CSCF. Up to Release 6 it can also be used to hide the internal network from the outside world (encrypting part of the SIP message), in which case it's called a THIG (*Topology Hiding Inter-network Gateway*). From Release 7 onwards this "entry point" function is removed from the I-CSCF and is now part of the

IBCF (*Interconnection Border Control Function*). The IBCF is used as gateway to external networks, and provides NAT and Firewall functions (pinholing).

Application servers

Application servers (AS) host and execute services, and interface with the S-CSCF using Session Initiation Protocol (SIP). An example of an application server that is being developed in 3GPP is the Voice call continuity Function (VCC Server). Depending on the actual service, the AS can operate in SIP proxy mode, SIP UA (user agent) mode or SIP B2BUA (back-to-back user agent) mode. An AS can be located in the home network or in an external third-party network. If located in the home network, it can query the HSS with the DIAMETER Sh interface (for a SIP-AS) or the Mobile Application Part (MAP) interface (for IM-SSF).

- AS: native IMS SIP application server
- IM-SSF: an *IP Multimedia Service Switching Function* interfaces with Customised Applications for Mobile networks Enhanced Logic (CAMEL) Application Servers using Camel Application Part (CAP)

Media Servers

The MRF (*Media Resource Function*) provides media related functions such as media manipulation (e.g. voice stream mixing) and playing of tones and announcements. Each MRF is further divided into a *Media Resource Function Controller* (MRFC) and a *Media Resource Function Processor* (MRFP).

- The MRFC is a signalling plane node that acts as a SIP User Agent to the S-CSCF, and which controls the MRFP with a H.248 interface
- The MRFP is a media plane node that implements all media-related functions.

Breakout Gateway

A BGCF (*Breakout Gateway Control Function*) is a SIP server that includes routing functionality based on telephone numbers. It is only used when calling from the IMS to a phone in a circuit switched network, such as the Public Switched Telephone Network (PSTN) or the Public land mobile network (PLMN).

PSTN Gateways

A PSTN/CS gateway interfaces with PSTN circuit switched (CS) networks. For signalling, CS networks use ISDN User Part (ISUP) (or BICC) over Message Transfer Part (MTP), while IMS uses Session Initiation Protocol (SIP) over IP. For media, CS networks use Pulse-code modulation (PCM), while IMS uses Real-time Transport Protocol (RTP).

- A Signalling Gateway (SGW) interfaces with the signalling plane of the CS. It transforms lower layer protocols as Stream Control Transmission Protocol (SCTP, an Internet Protocol (IP) protocol) into Message Transfer Part (MTP, an Signalling System 7 (SS7) protocol), to pass ISDN User Part (ISUP) from the MGCF to the CS network.
- A Media Gateway Controller Function (MGCF) does call control protocol conversion between SIP and ISUP and interfaces with the SGW over SCTP. It also controls the resources in an MGW with an H.248 interface.
- A Media Gateway (MGW) interfaces with the media plane of the CS network, by converting between RTP and PCM. It can also transcode when the codecs don't match (e.g. IMS might use AMR, PSTN might use G.711).

Media Resources

Media Resources are those components that operate on the media plane and are under the control of IMS Core functions. Specifically, Media Server (MS) and Media gateway (MGW)

Charging

Offline charging is applied to users who pay for their services periodically (e.g., at the end of the month). Online charging, also known as credit-based charging, is used for prepaid services, or real-time credit control of postpaid services. Both may be applied to the same session.

- Offline Charging : All the SIP network entities (P-CSCF, I-CSCF, S-CSCF, BGCF, MRFC, MGCF, AS) involved in the session use the DIAMETER Rf interface to send accounting information to a CCF (*Charging Collector Function*) located in the same domain. The CCF will collect all this information, and build a CDR (*Call Detail Record*), which is sent to the billing system (BS) of the domain. Each session carries an ICID (*IMS Charging Identifier*) as a unique identifier. IOI (*Inter Operator Identifier*) parameters define the originating and terminating networks. Each domain has its own charging network. Billing systems in different domains will also exchange information, so that roaming charges can be applied.

- Online charging : The S-CSCF talks to an SCF (*Session Charging Function*) which looks like a regular SIP application server. The SCF can signal the S-CSCF to terminate the session when the user runs out of credits during a session. The AS and MRFC use the DIAMETER Ro interface towards an ECF (*Event Charging Function*).

  - When IEC (*Immediate Event Charging*) is used, a number of credit units is immediately deducted from the user's account by the ECF and the MRFC or AS is then authorized to provide the service. The service is not authorized when not enough credit units are available.
  - When ECUR (*Event Charging with Unit Reservation*) is used, the ECF first reserves a number of credit units in the user's account and then authorizes the MRFC or the AS. After the service is over, the number of spent credit units is reported and deducted from the account; the reserved credit units are then cleared.

Interfaces description

Interface Name	**IMS entities**	**Description**	**Protocol**
Cr	MRFC, AS	Used by MRFC to fetch documents (scripts and other resources) from an AS	HTTP over dedicated TCP/SCTP channels
Cx	I-CSCF, S-CSCF, HSS	Used to communicate between I-CSCF/S-CSCF and HSS	DIAMETER
Dh	SIP AS, OSA, SCF, IM-SSF, HSS	Used by AS to find a correct HSS in a multi-HSS environment	DIAMETER
Dx	I-CSCF, S-	Used by I-CSCF/S-CSCF to find a correct	DIAMETER

	CSCF, SLF	HSS in a multi-HSS environment	
Gm	UE, P-CSCF	Used to exchange messages between UE and CSCFs	SIP
Go	PDF, GGSN	Allows operators to control QoS in a user plane and exchange charging correlation information between IMS and GPRS network	COPS (Rel5), DIAMETER (Rel6+)
Gq	P-CSCF, PDF	Used to exchange policy decisions-related information between P-CSCF and PDF	DIAMETER
ISC	S-CSCF, I-CSCF, AS	Used to exchange messages between CSCF and AS	SIP
Ma	I-CSCF -> AS	Used to directly forward SIP requests which are destinated to a Public Service Identity hosted by the AS	SIP
Mg	MGCF -> I-CSCF	MGCF converts ISUP signalling to SIP signalling and forwards SIP signalling to I-CSCF	SIP
Mi	S-CSCF -> BGCF	Used to exchange messages between S-CSCF and BGCF	SIP
Mj	BGCF -> MGCF	Used to exchange messages between BGCF and MGCF in the same IMS network	SIP

Mk	BGCF -> BGCF	Used to exchange messages between BGCFs in different IMS networks	SIP
Mm	I-CSCF, S-CSCF, external IP network	Used for exchanging messages between IMS and external IP networks	Not specified
Mn	MGCF, IM-MGW	Allows control of user-plane resources	H.248
Mp	MRFC, MRFP	Used to exchange messages between MRFC and MRFP	H.248
Mr	S-CSCF, MRFC	Used to exchange messages between S-CSCF and MRFC	SIP
Mw	P-CSCF, I-CSCF, S-CSCF	Used to exchange messages between CSCFs	SIP
Sh	SIP AS, OSA SCS, HSS	Used to exchange information between SIP AS/OSA SCS and HSS	DIAMETER
Si	IM-SSF, HSS	Used to exchange information between IM-SSF and HSS	MAP
Sr	MRFC, AS	Used by MRFC to fetch documents (scripts and other resources) from an AS	HTTP

Chapter – 10 (Broadband Access Network)

ADSL

Introduction

- **ADSL** is a form of DSL, a data communications technology that enables faster data transmission over copper telephone lines
- ADSL is capable of providing up to 50 Mbps, and supports voice, video and data.
- ADSL is the #1 Broadband Choice in the World with over 60% market share
- ADSL is now available in every region of the world

What does ADSL mean

- **A**symmetric - The data can flow faster in one direction than the other. Data transmission has faster downstream to the subscriber than upstream
- **D**igital - No type of communication is transferred in an analog method. All data is purely digital, and only at the end, modulated to be carried over the line.

- **S**ubscriber **L**ine - The data is carried over a single twisted pair copper loop to the subscriber premises

ADSL standards :

Standard name	**Common name**	**Downstream rate**	**Upstream rate**
ITU G.992.1	ADSL (G.DMT)	8 Mbit/s	1.0 Mbit/s
ITU G.992.2	ADSL Lite (G.Lite)	1.5 Mbit/s	0.5 Mbit/s
ITU G.992.3/4	ADSL2	12 Mbit/s	1.0 Mbit/s
ITU G.992.3/4 Annex J	ADSL2	12 Mbit/s	3.5 Mbit/s
ITU G.992.3/4 Annex L	RE-ADSL2	5 Mbit/s	0.8 Mbit/s
ITU G.992.5	ADSL2+	24 Mbit/s	1.0 Mbit/s
ITU G.992.5 Annex L	RE-ADSL2+	24 Mbit/s	1.0 Mbit/s
ITU G.992.5 Annex M	ADSL2+	28 Mbit/s	3.5 Mbit/s

ADSL Range

- In general, the maximum range for DSL without a repeater is 5.5 km
- As distance decreases toward the telephone company office, the data rate increases

Data Rate	Wire gauge	Wire size	Distance
1.5 or 2 Mbps	24 AWG	0.5 mm	5.5 km
1.5 or 2 Mbps	26 AWG	0.4 mm	4.6 km
6.1 Mbps	24 AWG	0.5 mm	3.7 km
1.5 or 2 Mbps	26 AWG	0.4 mm	2.7

- For larger distances, you may be able to have DSL if your phone company has extended the local loop with optical fiber cable

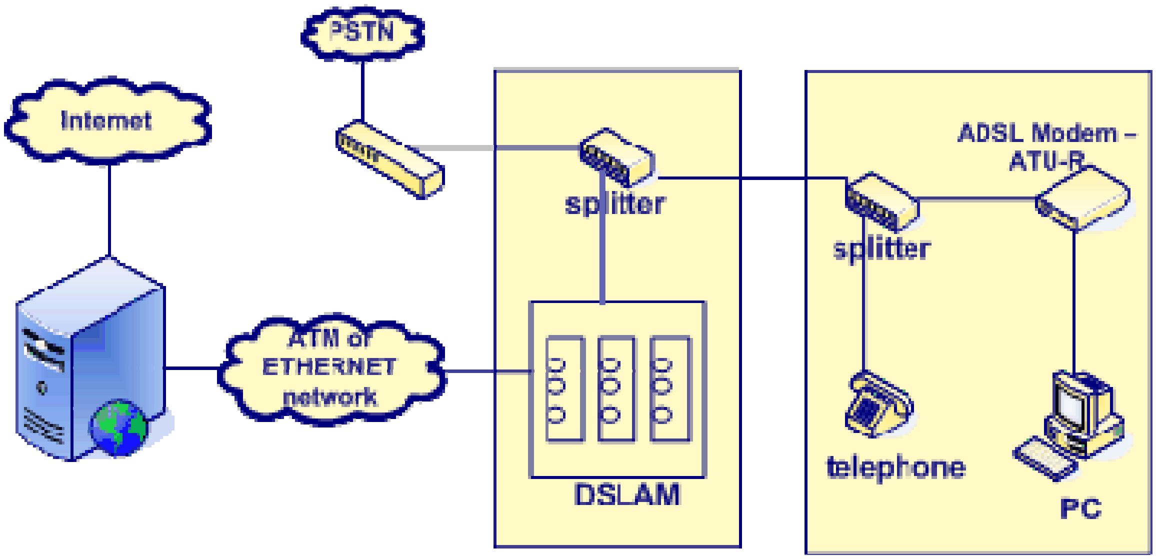
ADSL Speed Factors

- The distance from the local exchange
- The type and thickness of wires used
- The number and type of joins in the wire
- The proximity of the wire to other wires carrying ADSL, ISDN and other non-voice signals
- The proximity of the wires to radio transmitters.

ADSL network components

- The ADSL modem at the customer premises(ATU-R)
- The modem of the central office (ATU-C)
- DSL access multiplexer (DSLAM)
- Broadband Access Server (BAS)
- Splitter - an electronic low pass filter that separates the analogue voice or ISDN signal from ADSL data frequencies DSLAM.

ADSL Loop Architecture

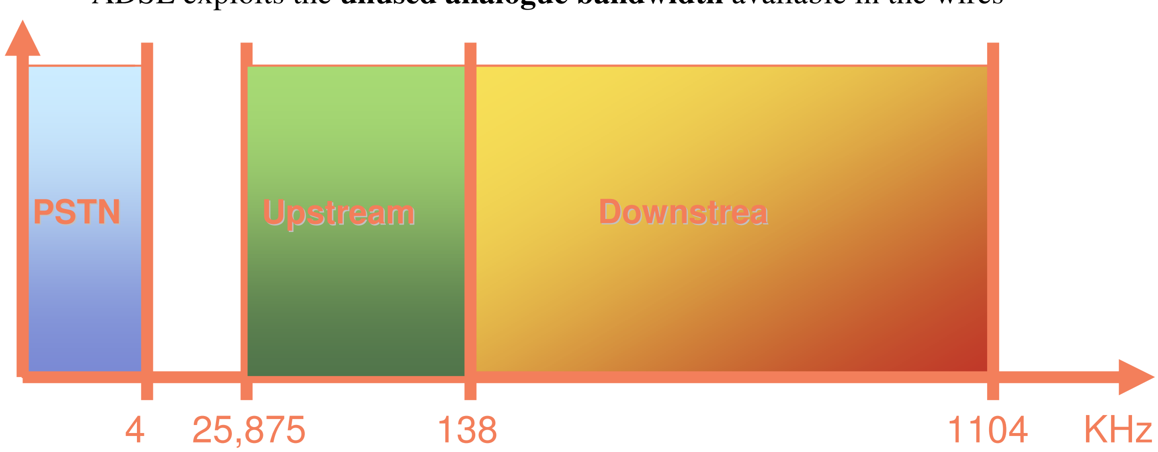


ADSL Requirements

- Phone-line, activated by your phone company for ADSL
- Filter to separate the phone signal from the Internet signal
- ADSL modem
- Subscription with an ISP supporting ADSL

How does ADSL work

- ADSL exploits the **unused analogue bandwidth** available in the wires



- ADSL works by using a frequency splitter device to split a traditional voice telephone line into two frequencies

ADSL Modulation

- Modulation is the overlaying of information (or the signal) onto an electronic or optical carrier waveform
- There are two competing and incompatible standards for modulating the ADSL signal:
  - Carrierless Amplitude Phase (CAP)
  - Discrete Multi-Tone (DMT)

Carrierless Amplitude Phase

- Carrierless Amplitude Phase (CAP) is an encoding method that divides the signals into two distinct bands:

  1. The upstream data channel (to the service provider), which is carried in the band between 25 and 160kHz

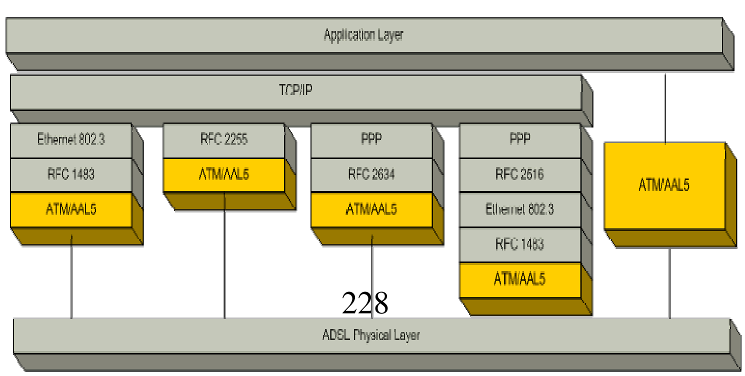
  1. The downstream data channel (to the user), which is carried in the band from 200kHz to 1.1MHz .

- These channels are widely separated in order to minimize the possibility of interference between the channels.

Discrete Multi-tone (DMT)

- Discrete Multi-Tone (DMT) separates the DSL signal so that the usable frequency range is separated into 256 channels of 4.3125kHz each.
- DMT has 224 downstream frequency bins (or carriers) and 32 upstream frequency bins.
- DMT constantly shifts signals between different channels to ensure that the best channels are used for transmission and reception.

ADSL Protocol stacks



Wimax

What is WiMAX?

- Worldwide Interoperability for Microwave Access (WiMAX) is the common name associated to the IEEE 802.16a/REVd/e standards.

- These standards are issued by the IEEE 802.16 subgroup that originally covered the Wireless Local Loop technologies with radio spectrum from 10 to 66 GHz.

IEEE 802.16 -- Introduction

- IEEE 802.16 (2001)
  - Air Interface for Fixed Broadband Wireless Access System MAC and PHY Specifications for 10 – 66 GHZ (LoS)
  - One PHY: Single Carrier
  - Connection-oriented, TDM/TDMA MAC, QoS, Privacy

- IEEE 802.16a (January 2003)
  - Amendment to 802.16, MAC Modifications and Additional PHY Specifications for 2 – 11 GHz (NLoS)
  - Three PHYs: OFDM, OFDMA, Single Carrier
  - Additional MAC functions: OFDM and OFDMA PHY support, Mesh topology support, ARQ

- IEEE 802.16d (July 2004)
  - Combines both IEEE 802.16 and 802.16a
  - Some modifications to the MAC and PHY

- IEEE 802.16e (2005?)
  - Amendment to 802.16-2004
  - MAC Modifications for limited mobility

IEEE 802.16 -- Introduction

Coverage range up to 50km and speeds up to 70Mbps(shared among users).

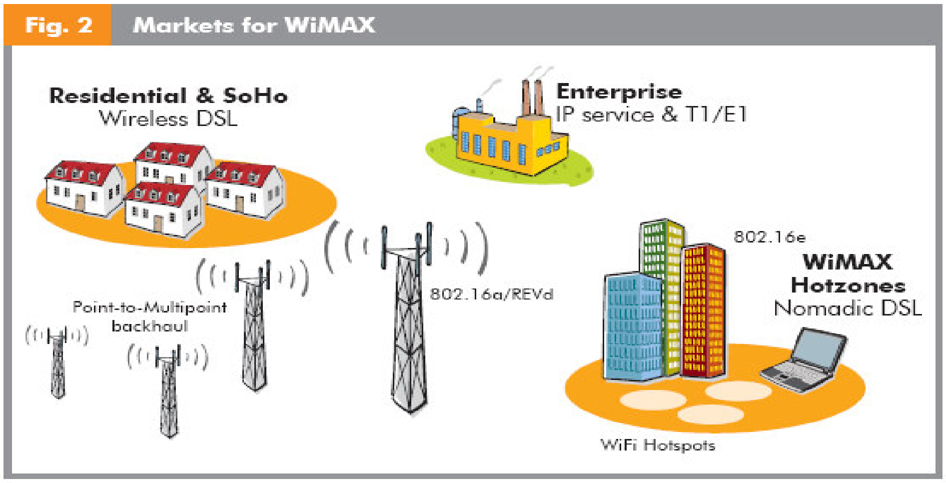


Fig. 2 Markets for WiMAX

IEEE 802.16 -- Introduction

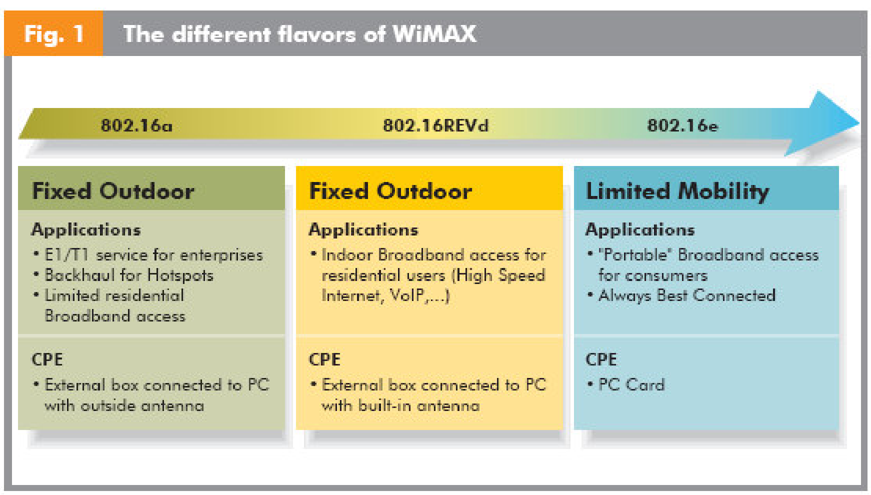
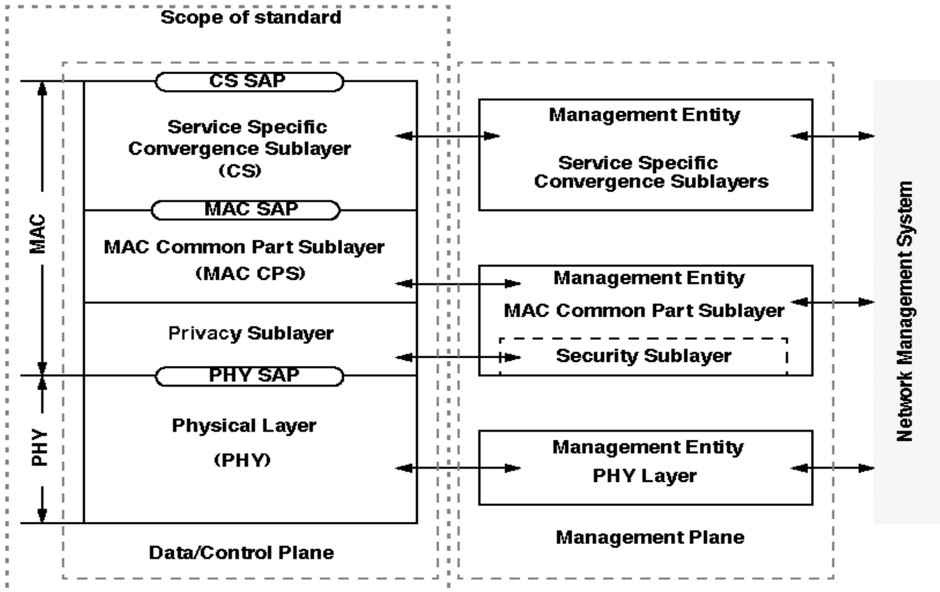
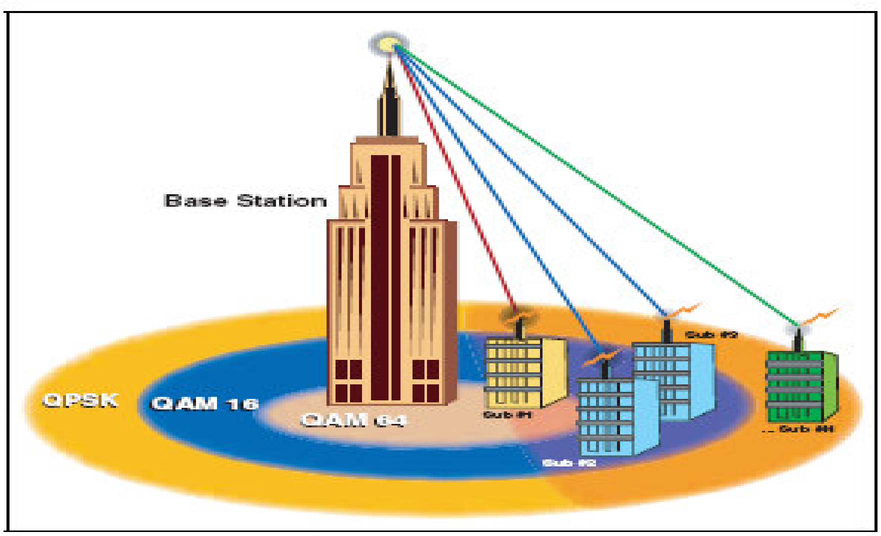


Fig. 1 The different flavors of WiMAX

Reference Model



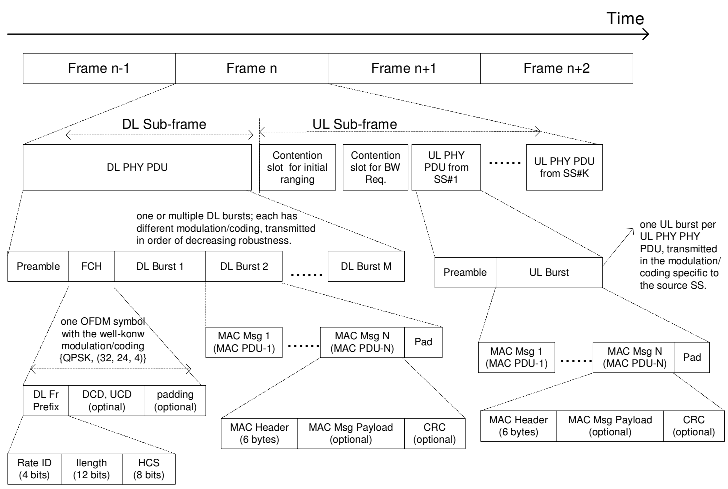
Adaptive PHY



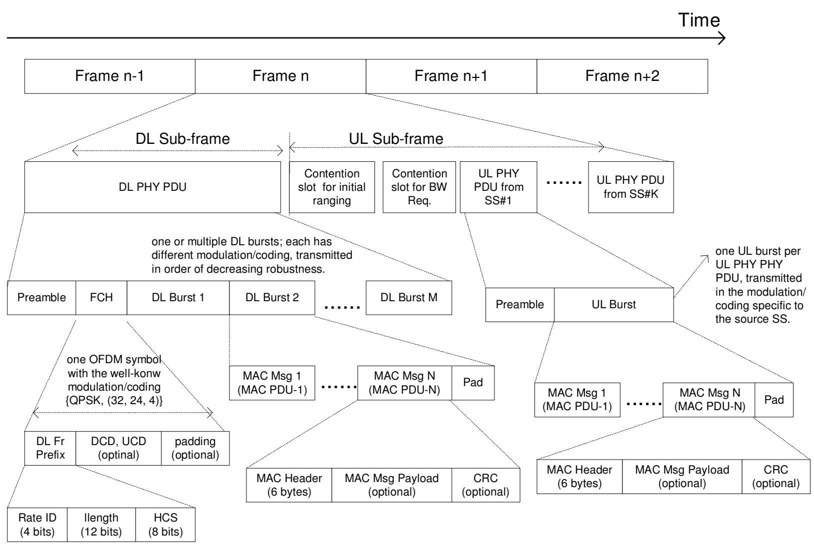
Duplex Scheme Support

- The duplex scheme is Usually specified by regulatory bodies, e.g., FCC
- Time-Division Duplex (TDD)
  - Downlink & Uplink time share the same RF channel
  - Dynamic asymmetry
  - does not transmit & receive simultaneously (low cost)
- Frequency-Division Duplex (FDD)
  - Downlink & Uplink on separate RF channels
  - Full Duplexing (FDX): can Tx and Rx simultaneously;
  - Half-duplexing (HDX) SSs supported (low cost)

IEEE 802.16 MAC – OFDM PHY TDD Frame Structure



IEEE 802.16 MAC – OFDM PHY FDD Frame Structure



IEEE 802.16 MAC addressing and Identifiers

- SS has 48-bit IEEE MAC address
- BS has 48-bit base station ID
  - Not a MAC address
  - 24-bit operator indicator
- 16-bit connection ID (CID)
- 32-bit service flow ID (SFID)
- 16-bit security association ID (SAID)

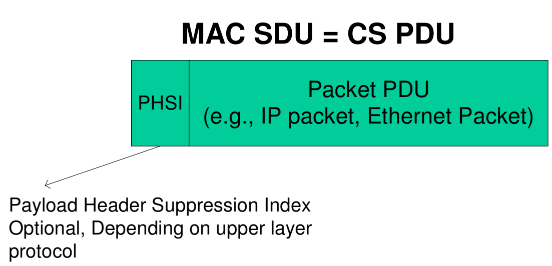
IEEE 802.16 MAC – Convergence Sub-Layer (CS)

- ATM Convergence Sub-Layer:
  - Support for VP/VC switched connections
  - Support for end-to-end signaling of dynamically created connections
  - ATM header suppression
  - Full QoS support

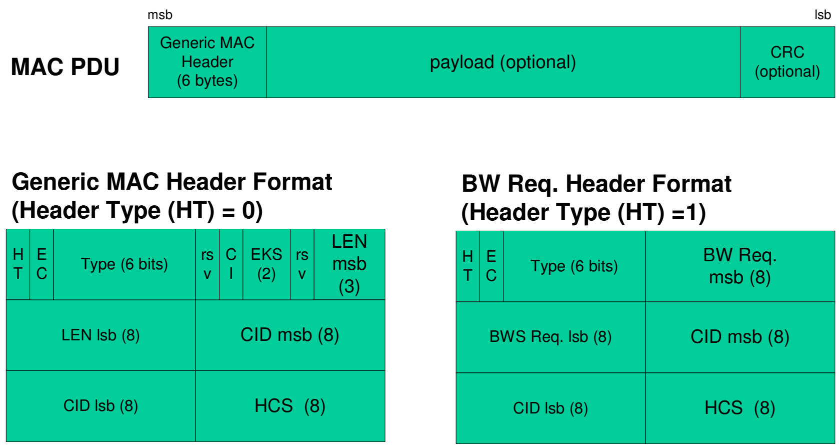
- Packet Convergence Sub-Layer:
  - Initial support for Ethernet, VLAN, IPv4, and IPv6
  - Payload header suppression
  - Full QoS support

IEEE 802.16 MAC – CS – Packet Convergence Sub-Layer

- Functions:
  - Classification: mapping the higher layer PDUs (Protocol Data Units) into appropriate MAC connections
  - Payload header suppression (optional)
  - MAC SDU (Service Data Unit), i.e, CS PDU, formatting



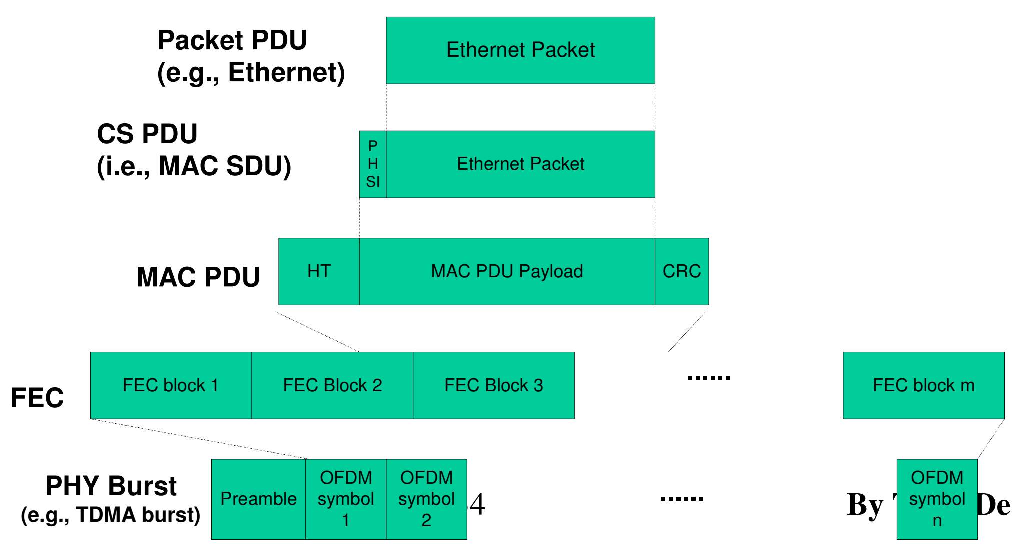
IEEE 802.16 MAC – CPS – MAC PDU Format



IEEE 802.16 MAC -- CPS -- Three Types of MAC PDUs

- Data MAC PDUs
  - HT = 0
  - Payloads are MAC SDUs/segments, i.e., data from upper layer (CS PDUs)
  - Transmitted on data connections
- Management MAC PDUs
  - HT =0
  - Payloads are MAC management messages or IP packets encapsulated in MAC CS PDUs
  - Transmitted on management connections
- BW Req. MAC PDUs
  - HT =1; and no payload, i.e., just a Header

IEEE 802.16 MAC – CPS – Data Packet Encapsulations

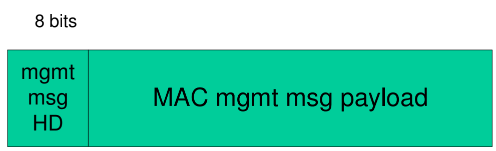


IEEE 802.16 MAC – CPS -- MAC Management Connections

- Each SS has 3 management connections in each direction:
  - **Basic Connection**:
    - short and time-urgent MAC management messages
    - MAC mgmt messages as MAC PDU payloads
  - **Primary Management connection**:
    - longer and more delay tolerant MAC mgmt messages
    - MAC mgmt messages as MAC PDU payloads
  - **Secondary Management Connection:**
    - Standard based mgmt messages, e.g., DHCP, SNMP, …etc
    - IP packets based CS PDU as MAC PDU payload

IEEE 802.16 MAC – CPS – MAC Management Messages

- MAC mgmt message format:



- MAC mgmt msg can be sent on: Basic connections; Primary mgmt connection; Broadcast connection; and initial ranging connections
- 41 MAC mgmt msgs specified in 802.16
- The TLV (type/length/value) encoding scheme is used in MAC mgmt msg, e.g., in UCD msg for UL burst profiles,

  (type=1, length=1, value=1) → QPSK modulation

  (type=1, length=1, value=2) → 16QAM modulation

  (type=1, length=1, value=3) → 64QAM modulation

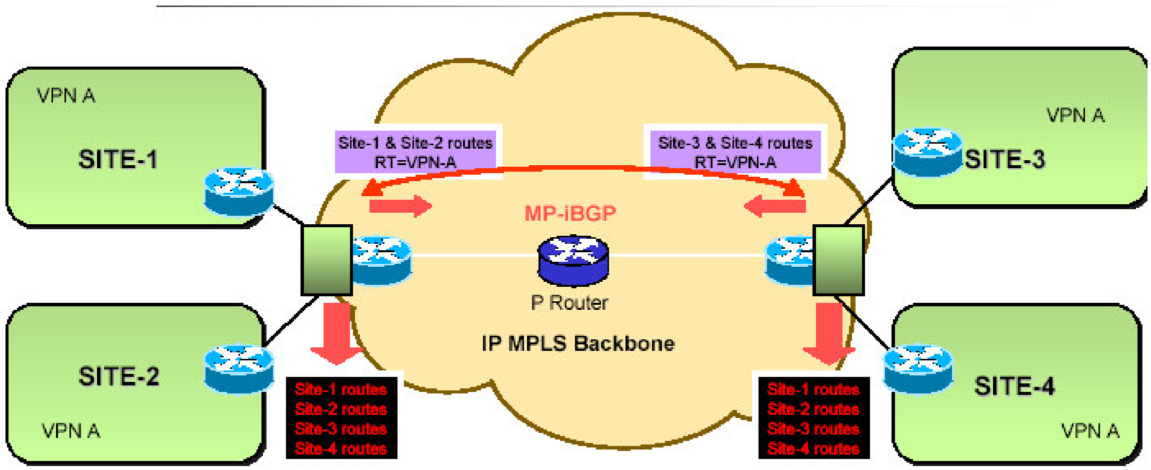
IEEE 802.16 MAC – CPS – MAC PDU Transmission

- MAC PDUs are transmitted in PHY Bursts
- The PHY burst can contain multiple FEC blocks
- MAC PDUs may span FEC block boundaries
- Concatenation
- Packing
- Segmentation
- Sub-headers

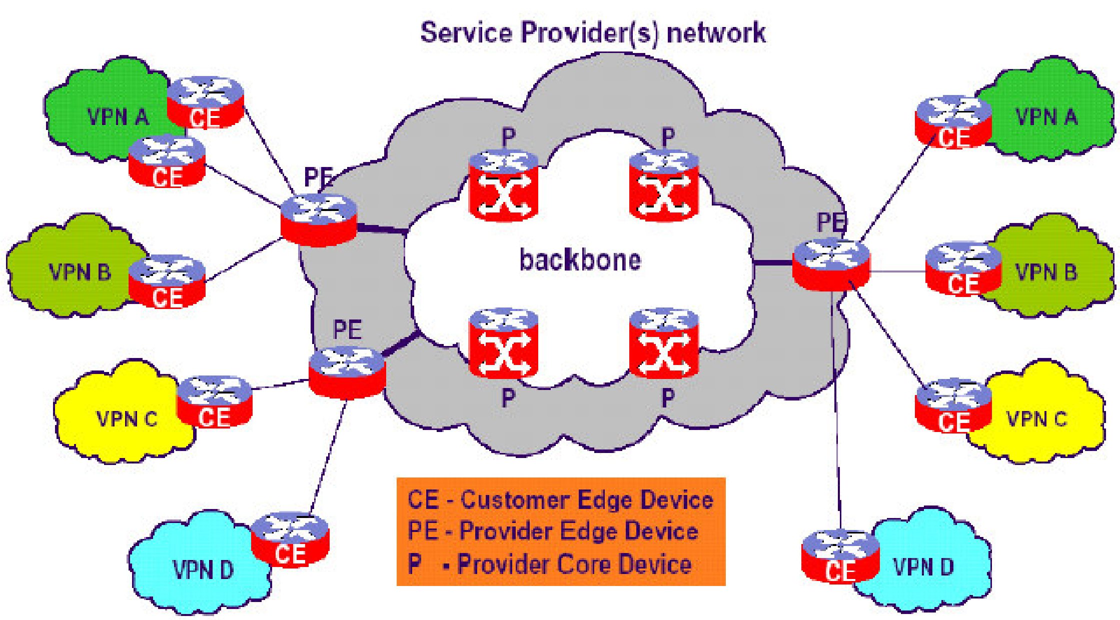
Chapter – 11 (MPLS Core Network)

MPLS

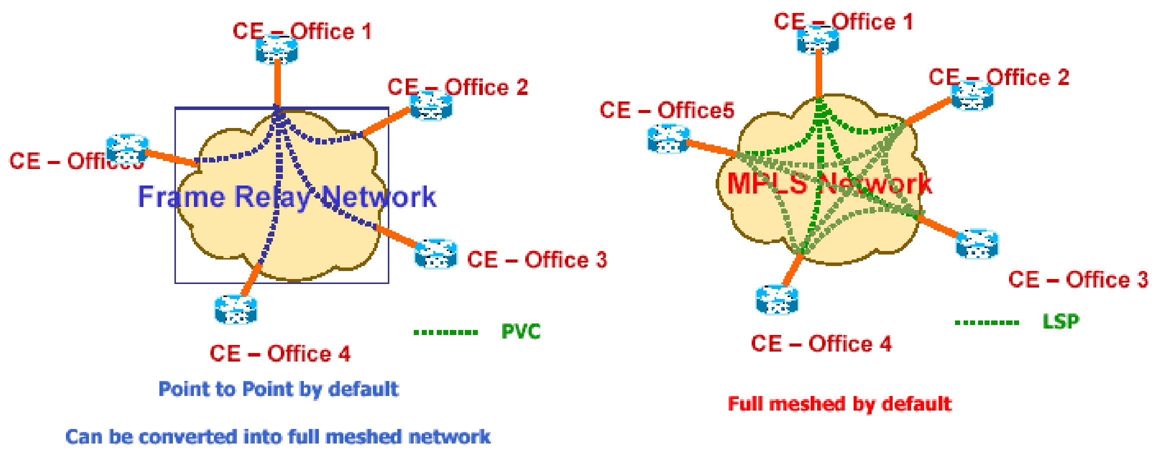
Basic Intranet Model Intranet Model



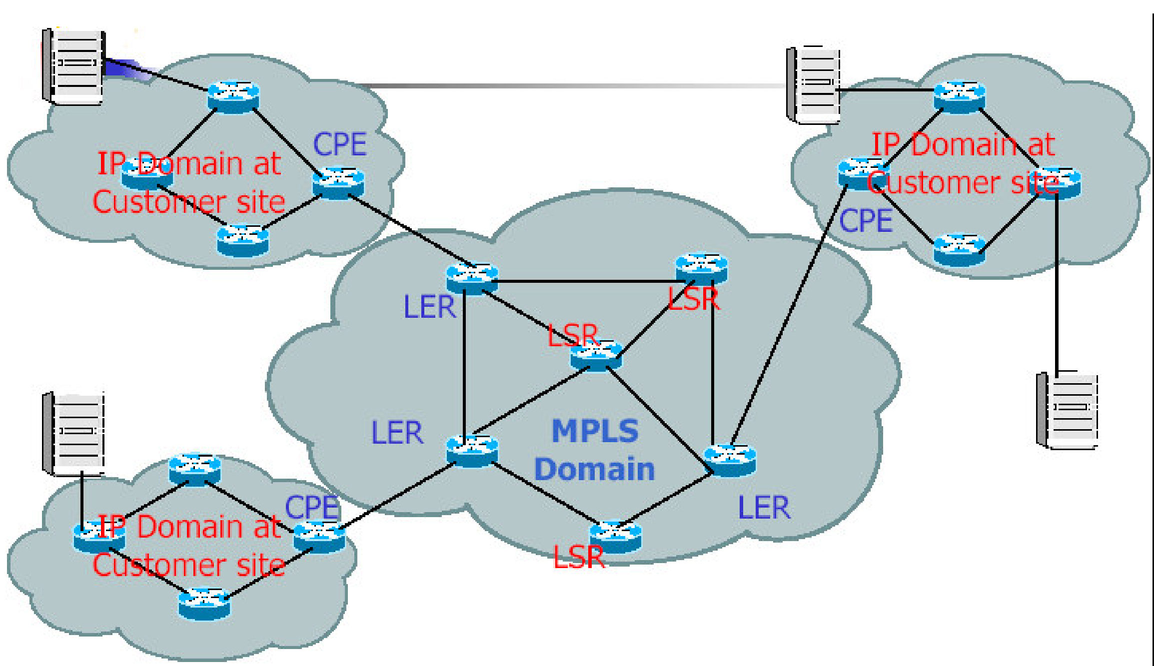
MPLS VPN Network Model



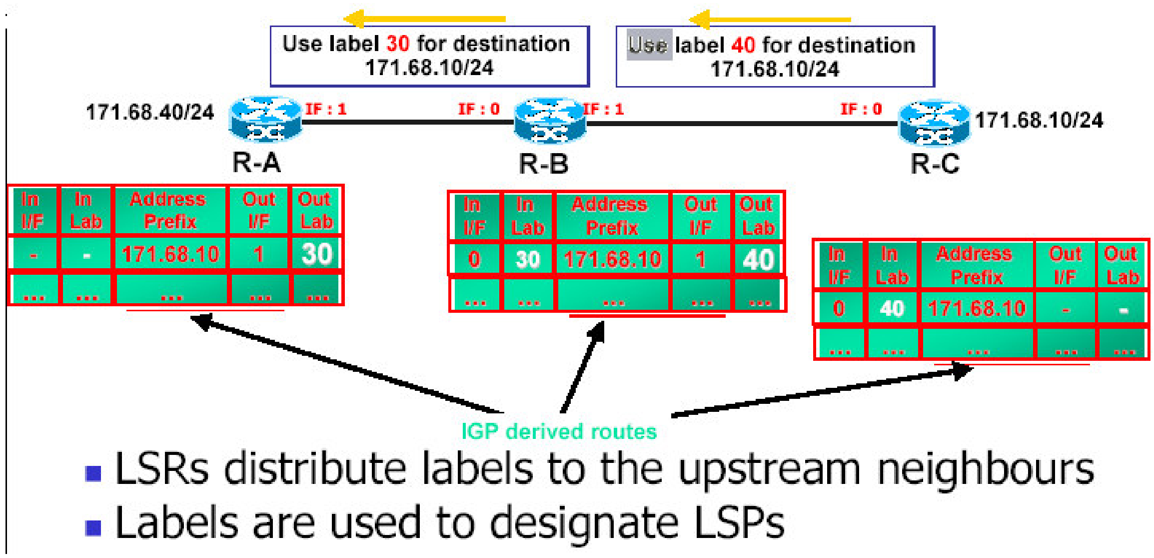
VPN Models



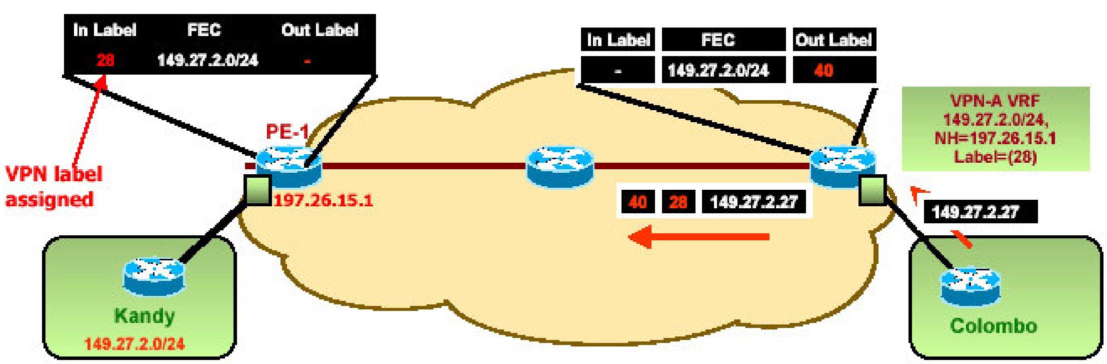
MPLS Domain And Components



Label Assignment and Distribution

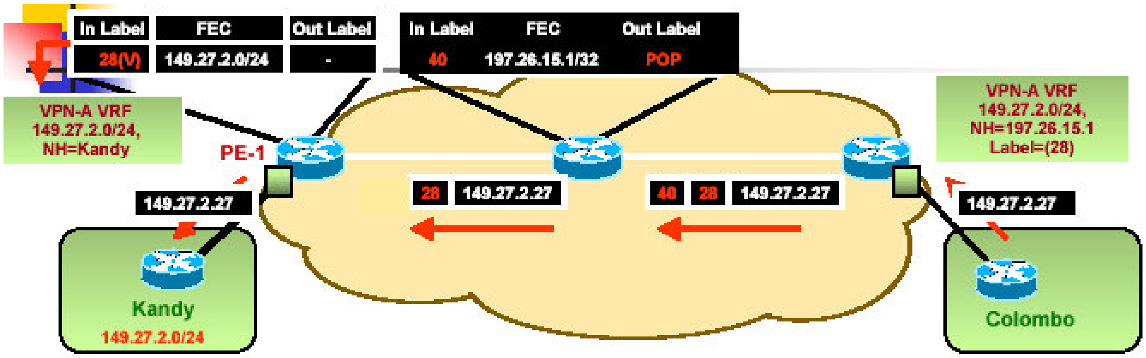


VPN Packet Forwarding VPN Packet Forwarding



• Ingress PE receives IP data packets

• PE router performs IP Best Match from VPN LFIB, finds iBGP next-hop and imposes a stack of labels <IGP, VPN>

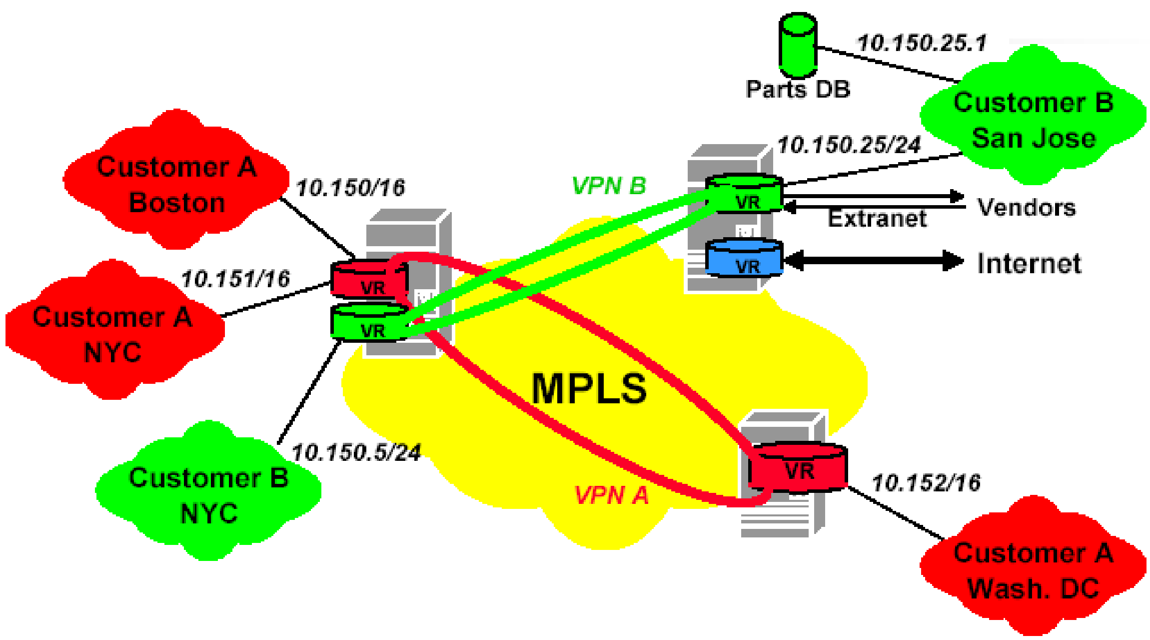


Penultimate PE router removes the IGP label

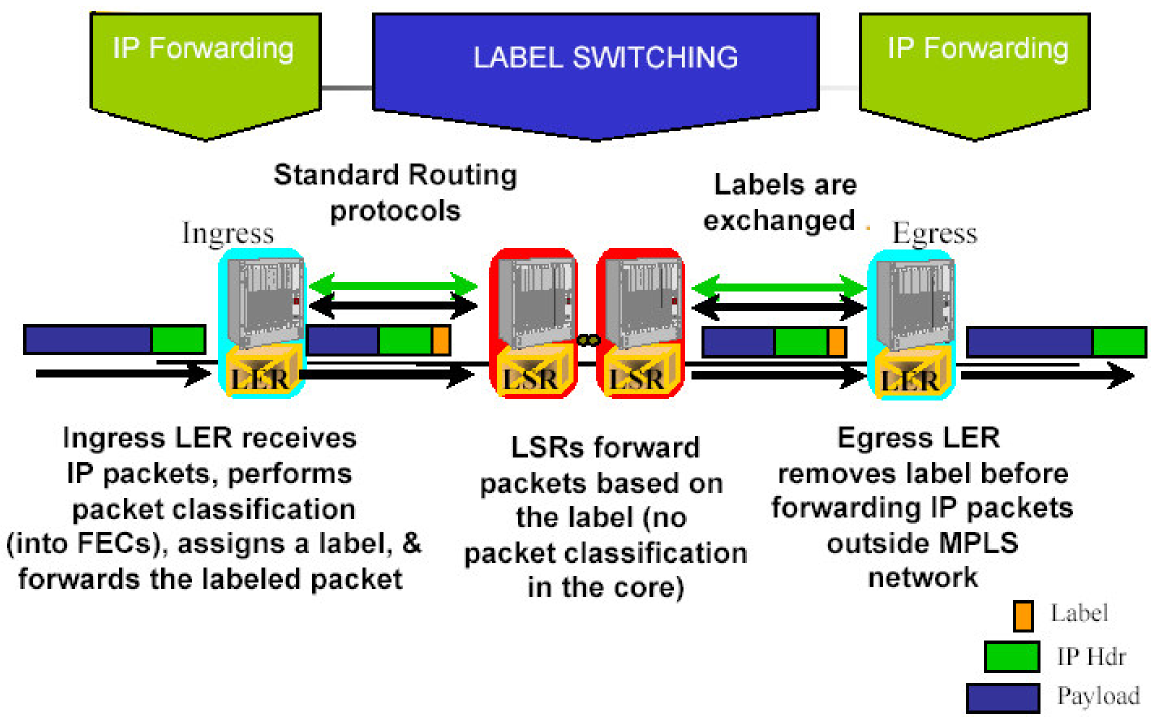
- **Penultimate Hop Popping procedures (implicit-null label)**
- Egress PE router uses the VPN label to select which VPN/CE to forward the packet to
- VPN label is removed and the packet is routed toward the VPN site

Separate Routing - Private

Addressing



MPLS Operation



Terminology

Provider network (P network)
Provider edge router (PE/LER router) – physical connection to CE router and to core of P network
Provider router (P/LSR router) - internal to P network and oblivious to existence of VPNs
Customer edge router (CE router) – physically connected to PE router
Customer router (C router) - internal to C network and invisible to PE router
PE-CE link

Label Edge Routers/Provider

LER Functions

1. Map IP Packets to labels

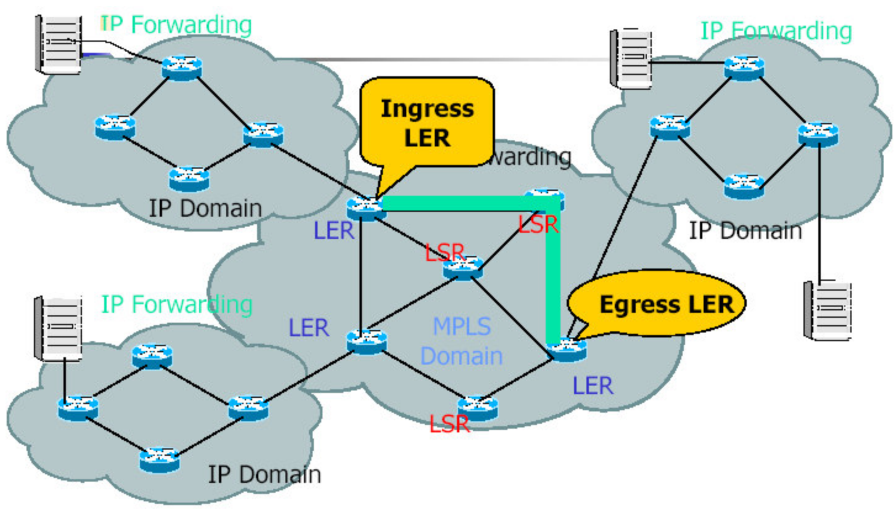
2. Push Labels on IP packets

3. Apply QoS Functions

4. Initiate LSP setup process

5. Traffic Engineering

Ingress and Egress LERs



Label Switched Routers/ Provider

Routers (LSR/P)

LSR Functions

1. Swap Labels
2. Apply QoS Functions
3. Participate in LSP setup process
4. Only knows routes within MPLS Domain

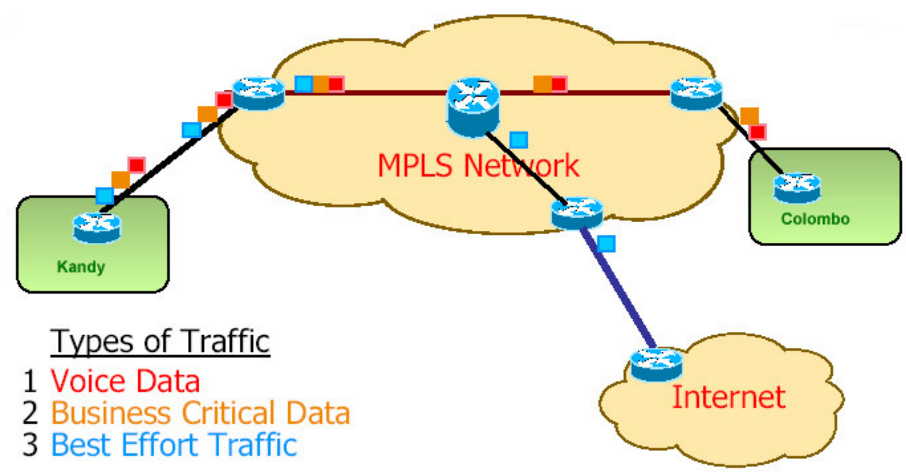
Multi Protocol Label Switching

MPLS is an Internet Engineering Task Force (IETF) specified framework for efficient designing, forwarding, Routing and Switching of traffic flows in a network. (RFC 2547)

MPLS Quality Of Service

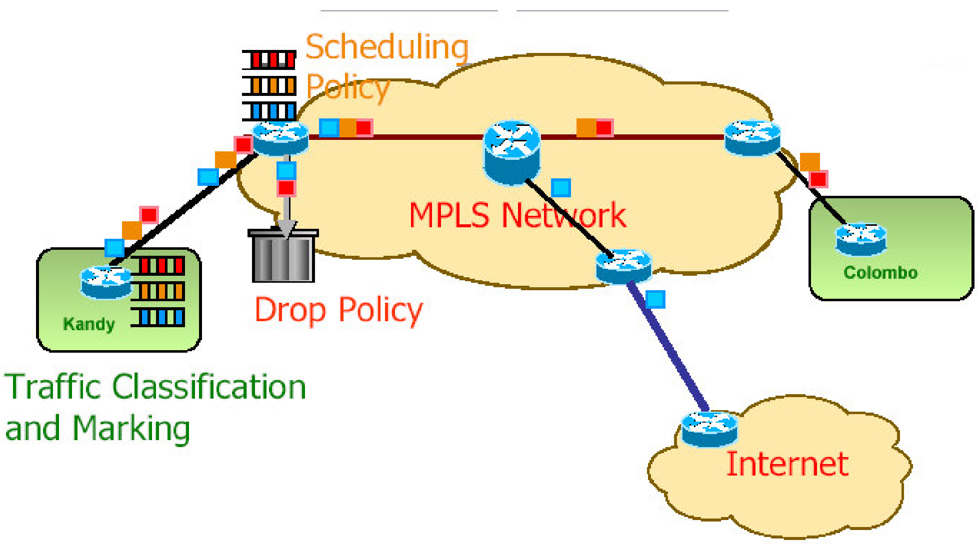
(QOS)

Traffic Classification



Differentiated Model Divide Traffic into Classes

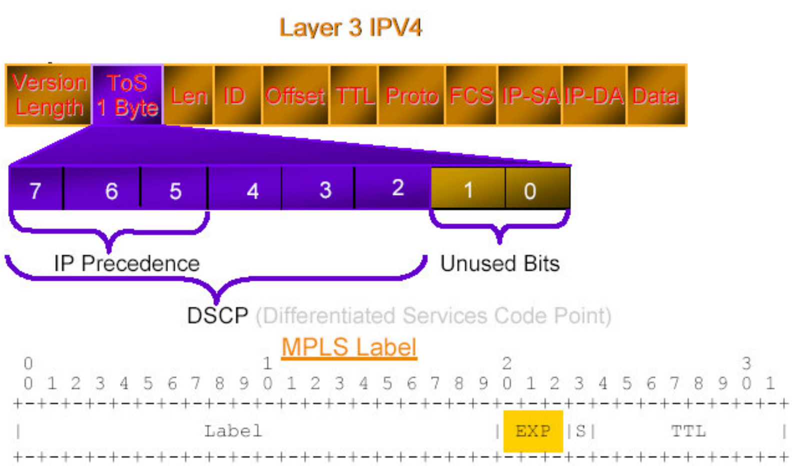
Traffic Policing Traffic Policing



Differential Model Features

Classification
Marking
Policing and Shaping
Congestion Avoidance
Congestion Management

Differentiated Model Features Marking



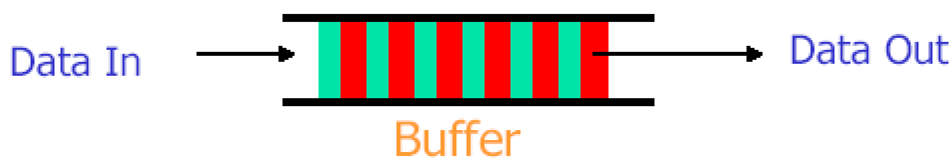
Differentiated Model Features Policing and Shaping

Policing is the QoS component that limits Incoming / Outgoing traffic flow to a defined/assigned bit rate Shaping is the QoS feature component that regulates Outgoing traffic flow to a defined/assigned bit rate

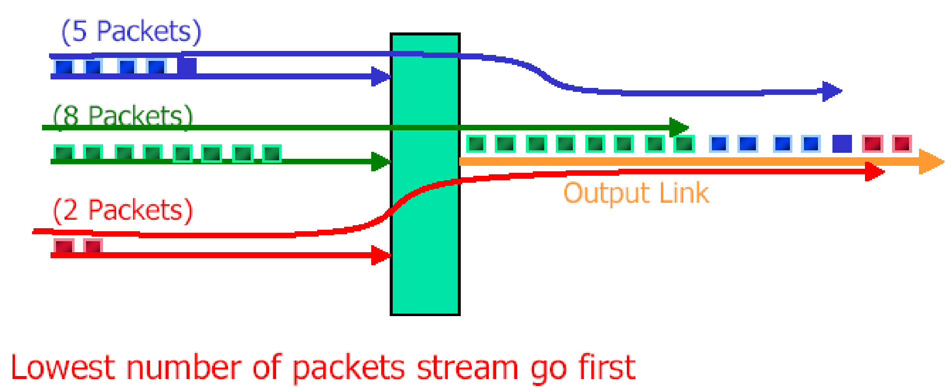
Differentiated Model Features Congestion Management

Scheduling Policy
First In First Out (FIFO)
Weighted Fair Queuing (WFQ)
Class Based Weighted Fair Queuing (CBWFQ)
Priority Queuing

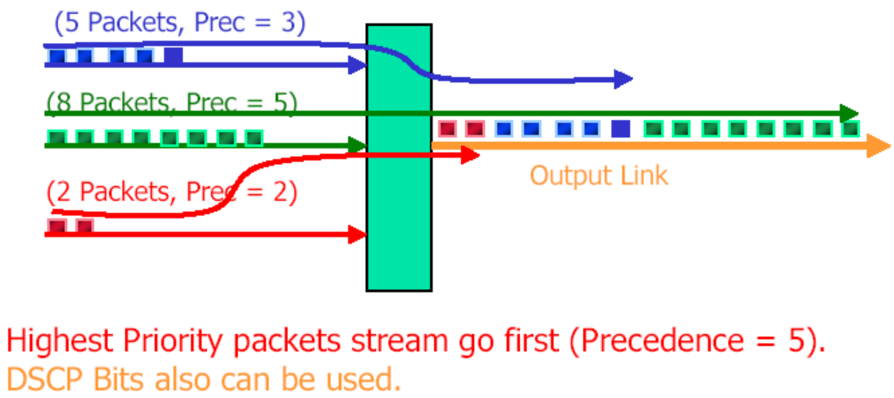
Congestion Management - FIFO



Congestion Management - WFQ



Congestion Management - CBWFQ



MPLS QoS