Chapter 2
Internet Development Versus Networking Modes

Abstract In the 1970s, the ever-increasing application of computers pushed forward the development of computer networks. One famous project is the ARPANET, which was founded by the U.S. Department of Defense and laid the foundation for the Internet. Initially, the objective of developing a computer network was to enable data applications through interconnecting computers located in different sites for data sharing and message exchange. Typical applications available at that time included e-mail, news, File Transfer Protocol (FTP), and Telnet. Later, this technology spread quickly and widely and has now become a worldwide Internet, an essential part of our daily life. The applications supported by the Internet today include not only data applications but also real-time applications such as voice over IP (VoIP), IPTV, networked entertainment, and social networks, with new applications continuously being created. Although there have been many changes in both the number and types of applications and users on the Internet, the networking modes, which collectively refer to the principle and methodology for networking, have remained almost intact. This chapter briefly reviews the major networking modes and discusses the challenges that they may face for the future Internet.

2.1 Networking Modes

As discussed in Chap. 1, visibly a network can be abstracted into a collection of nodes and links. Invisibly, a network is composed of a set of functions which are used to provide network services. These functions must address the networking issues described in Chap. 1 by implementing relevant network protocols and algorithms. The network design must decide what functions to implement and how to distribute them as well as how to make them cooperate as a whole to provide network services. Generally, there are two design principles: the layered model, which governs a vertical distribution of network functions within one network node, and the end-to-end arguments, which mainly guide a horizontal distribution of network functions among different network nodes.

S.M. Jiang, *Future Wireless and Optical Networks*,
Computer Communications and Networks,
DOI 10.1007/978-1-4471-2822-9_2, © Springer-Verlag London Limited 2012

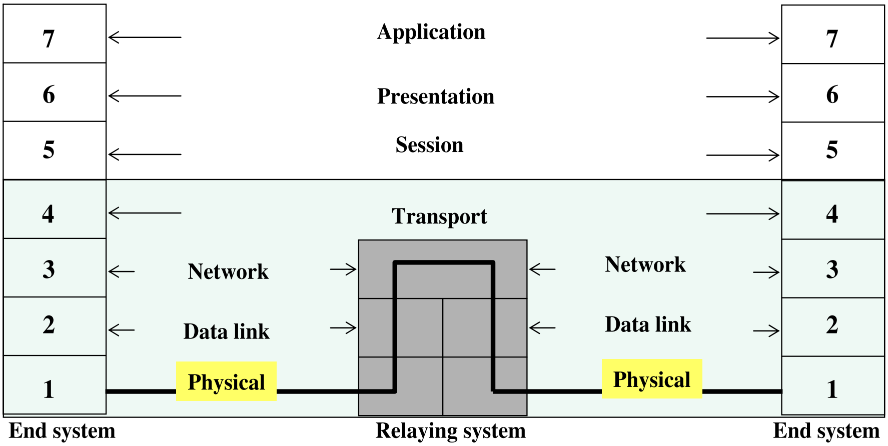


Fig. 2.1 ISO's OSI seven-layer reference model

2.1.1 Layered Models

The major issue to be addressed for interconnecting heterogeneous computers is the methodology to build a network. Obviously, building such a system is a huge and complex engineering task, and it is difficult for just a few companies or institutions to complete the whole system. Therefore, dividing this task into small pieces so that each piece can be solved relatively easily and independently is the basic idea of the layered model. Furthermore, different layers are made independent of each other in terms of layer internal design but with a standardized interface, through which the adjacent layers can interact and communicate with each other. There are three typical layered models: the OSI seven-layer reference model, the five-layer TCP/IP model, and the ATM reference model. However, since ATM is "going downhill," only the former two models are discussed below.

2.1.1.1 OSI Reference Model

In 1978, the ISO started to develop a standard for Open Systems Interconnection (OSI), which was enforced in 1983. This standard allows a system to communicate with any systems anywhere that are designed following the same standard. According to this reference model, an open system is divided into seven layers as illustrated in Fig. 2.1. The primary functions of each layer [1, 2] are briefly described below; some have been newly added following the recent development of networking technologies.

- Physical layer—handles the bit-level communication over media. It provides basic functions for digital communication such as modulation/demodulation, coding/decoding, synchronization, and error control.

- Data link layer—provides the frame-level communication over links. A frame is a formatted bit block. Since the physical layer cannot guarantee error-free communication, the primary function of this layer is error control that provides an error-free communication service for the network layer. For a shared-medium network in which multiple users share a common medium, a MAC protocol is especially needed to coordinate the medium sharing.
- Network layer—provides the packet-level communication across the whole network. A packet is also a formatted bit block. Addressing, routing, congestion control, QoS support, and network security are the primary functions of this layer.
- Transport layer—provides communications between endpoints, relying on the underlying network layer but providing network-independent services to higher layers. Its major functions include flow and congestion control as well as end-to-end transmission reliability control and security.
- Session layer—provides control functions for communications between cooperating applications at endpoints, including exchanging the identifications of endpoints, and establishing, managing, and terminating sessions between them.
- Presentation layer—provides independence for application processes from cooperating applications with different data representation, and services to the application layer by transforming data structures into a format agreed upon by the partners.
- Application layer—provides an interface with an application process requiring communication support, with standard services for transmission between user processes, database access, and running of processes on different computers.

From the above definitions, we find that only layer 1 through layer 4 are related to the networking issues discussed earlier.

Actually, each layer is composed of a set of functions. A function of one layer that can be seen by the layer above is called the service offered to its next higher layer. According to the model, each layer except the highest layer has a set of services that are provided for its next higher layer through the standardized service access points (SAPs). Thus, devices from different manufacturers can work together. Furthermore, only the adjacent layers can communicate with each other through SAPs. A network connection between source and destination must be set up first to provide the network service, particularly for the network layer. Therefore, the network defined by this model is a connection-oriented network.

2.1.1.2 TCP/IP Model

Although the ISO seven-layer reference model was theoretically considered as the ultimate model for worldwide interoperable networking, the TCP/IP model is the most popular and successful implementation in today's Internet. The mapping between the TCP/IP model and the OSI seven-layer model is illustrated in Fig. 2.2, where the function distribution of the TCP/IP model is also depicted. This model loosely follows the OSI reference model with only four layers, and its physical and data link layers are integrated into one link layer.

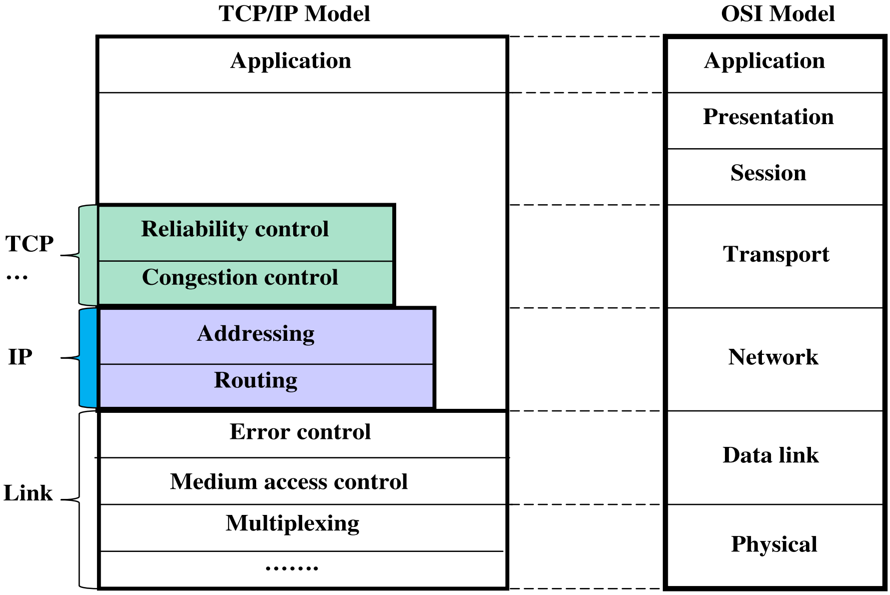


Fig. 2.2 ISO seven-layer model versus TCP/IP model

The key difference between these two models in terms of networking modes is that, with the ISO model, the connection-oriented service is provided in the network layer. Thus, packet switching can be used to accelerate packet forwarding. With the TCP/IP model, only the connectionless service using IP is provided, with which data packets can be transmitted anytime without a prior connection setup. In this case, routing has to be used. Due to its simplicity, this model has been widely implemented in the Internet. A brief comparison between these two models can be found in [2].

2.1.2 End-to-End Arguments

The end-to-end arguments [3] were among the most influential design principles for the Internet even before they were first explicitly articulated in the early 1980s. The arguments state that "functions placed at low levels of a system may be redundant or of little value when compared with the cost of providing them at that low level." Basically, these arguments as a whole strongly suggest a design principle of putting the application-level functions at the network edge rather than inside the network as much as possible in order to simplify network design and implementation. Doing so can also make the so-designed network protocols versatile for different types of networks.

One typical example following the arguments is the TCP/IP protocol stack. TCP is the most well-known protocol for end-to-end reliable transmission, and was first

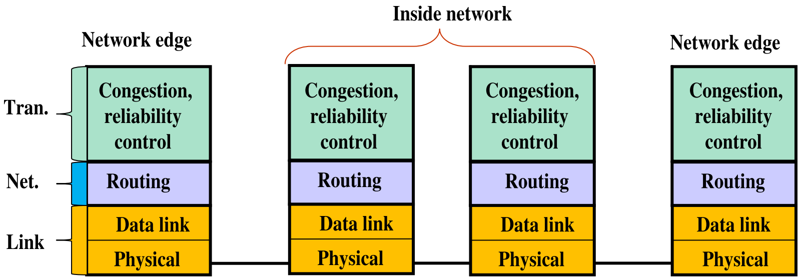


Fig. 2.3 A possible implementation without following the end-to-end arguments

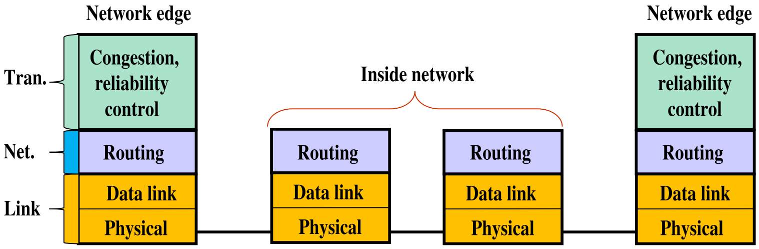


Fig. 2.4 An illustrative implementation following the end-to-end arguments

published in 1977 [4]. Although TCP was invented before the first explicit articulation of the arguments in 1980s, it follows them well [3]. As illustrated in Fig. 2.3, a possible networking mode is to allow every node to fully implement networking functions as depicted in Fig. 2.2. However, following the end-to-end arguments, only the source and destination of a TCP connection need to have the full functions as illustrated in Fig. 2.4. Similarly, the IP in the network layer is also as simple, using the connectionless networking technology and the first-in-first-out (FIFO) scheduling policy without the retransmission of lost packets. Actually, the end-to-end reliability control function is shifted to TCP in the transport layer. This design can simplify the implementation of relaying units such as routers so the protocols can run successfully over various types of networks [5].

2.2 Challenges of Wireless and Optical Networks

There is no doubt that both the layered models and the end-to-end arguments have played critical roles in the successful development of the Internet over the past three decades. These design principles allow a complex system to be decomposed into smaller subsystems for easy implementation and can guarantee the interoperability of devices produced from different manufacturers. They can also simplify the network structure and operation so that such a designed network can run over various communication systems for interconnection.

However, the user's expectations of the Internet today and in the future have been going far beyond the objectives of its original designers in terms of network coverage, capacity, and number and type of both users and applications. These new requirements greatly stimulate the development of new networking technologies for the future Internet, as discussed below.

2.2.1 New Requirements of Users and Applications

When the Internet was designed for the U.S. Department of Defense in the 1970s, the main objective was to share files in computers or storage media located in government offices and institutional organizations. These computers and storage media were mainly wired with metallic cables and were almost stationary. By default, the design of networking protocols and algorithms such as routing protocols and congestion control schemes assumes the availability of powerful computing and buffering technologies for the network service provisioning.

Today, the Internet is available almost everywhere. The ever-increasing number of users has already caused the address space of the original IP (i.e., IP version 4 or simply IPv4) to be exhausted. According to the statistics given by [6], the number of Internet users in the world was 360,985,492 on December 31, 2000, and reached 1,966,514,816 by June 30, 2010, which is equivalent to 28.7% of the world population, increasing at a rate of 444.8%.

On the other hand, mobile users, who may use mobile phones, laptop computers, or personal digital assistants, etc., constitute the largest population of electronic device users in the world. According to the report published by the United Nations (UN) on February 23, 2010 [7], two-thirds of the world's population (i.e., around 4.7 billion) were mobile subscribers, while this number was only about one billion in 2002. Meanwhile, the number of powerful and smart mobile devices such as Apple iPhone and iPad is growing rapidly. The number of smart phone users is expected to exceed one billion by 2013 [8]. Thus, more and more applications originally designed for stationary devices are expected to run on mobile devices too. This change requires the Internet be able to efficiently support mobile applications.

Regarding the Internet applications, besides the original data applications such as e-mail, FTP, and remote login, newly developed killer data applications include the World Wide Web (WWW) and e-commerce. Real-time applications have also been developed, for example, voice over IP (VoIP) and networked games as well as streaming applications like IP television (IPTV). Some other developing applications include cloud computing [9] and the Internet of Things (IoT) [10]. One can expect that more and more new applications will appear in the future as the number of Internet users is continuously increasing.

These radical changes in both the number and type of users and applications that the Internet should support pose a big challenge to networking modes. Although much incremental effort has been made to enhance the Internet, the networking modes mentioned earlier have been kept almost intact so far. The question is whether these networking modes can still guide us to foster the Internet to satisfy new requirements in terms of network capacity, pervasive networking, quality of service (QoS) provisioning, and network security support as well as green networking. These issues are discussed below.

2.2.2 Network Capacity

The ever-increasing number of users and applications requires immense network capacity to support them. For the time being, this level of capacity can only be provided by using optical communication technologies, particularly using optical fibers as communication media. For example, so far Gigabit Ethernet using metallic wires such as twisted copper pairs can only provide Gbps-level capacity over a maximum distance of less than 100 km, while a single optical fiber can provide Tbps-level (1 Tbps $= 10^3$ Gbps) transmission rates over much longer distances. Furthermore, optical fibers are much cheaper than metallic wires.

However, the ultra-high transmission rate of optical fibers cannot transfer into the same level of network speed if there is not the same level of high-speed packet forwarding technology. The successful development of the Internet is largely due to the mature electronic computer technology, which can provide high-speed computing and buffering support for realizing high-speed and quality networking. Many networking functions such as routing, QoS provisioning, and congestion control need complex computing and buffering, the key elements indispensable for networking. Unfortunately, cost-effective photonic computer and optical random access memory are not yet available, and it may still be a long time before they are available to provide ultra-high speed optical networking.

The following sections give a brief survey of optical networking technologies that aim at handling this issue.

2.2.2.1 Optical-Electrical-Optical Conversion

Today, many optical networks are implemented by jointly using electronic computers and optical fibers through optical-electrical-optical (OEO) conversion, as illustrated in Fig. 2.5. Optical fibers are used as transmission media while electronic computers are used for complex networking operations, particularly routing. To this end, optical signals are first converted into electrical signals, which are then routed electrically at the relaying units such as IP routers, and then recovered as optical signals for transmission over optical fibers. The major drawback of this approach is that the networking speed will be limited by the electronic computer speed, which will become a bottleneck of networking performance. The transmission speed of optical fibers is much faster than that of electronic computers, and the same applies to their speed increase rates, as discussed below.

According to Moore's Law [11], the speed of an electronic processor can be doubled almost every 18 months, while it can be doubled almost every nine months for

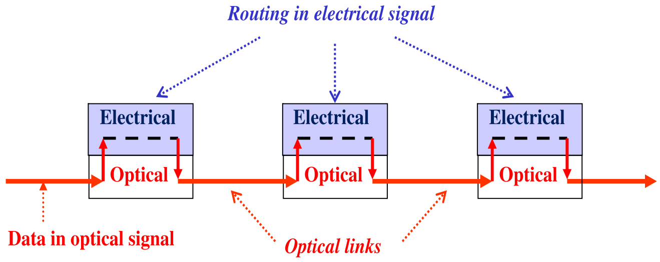


Fig. 2.5 Optical networks based on OEO conversion

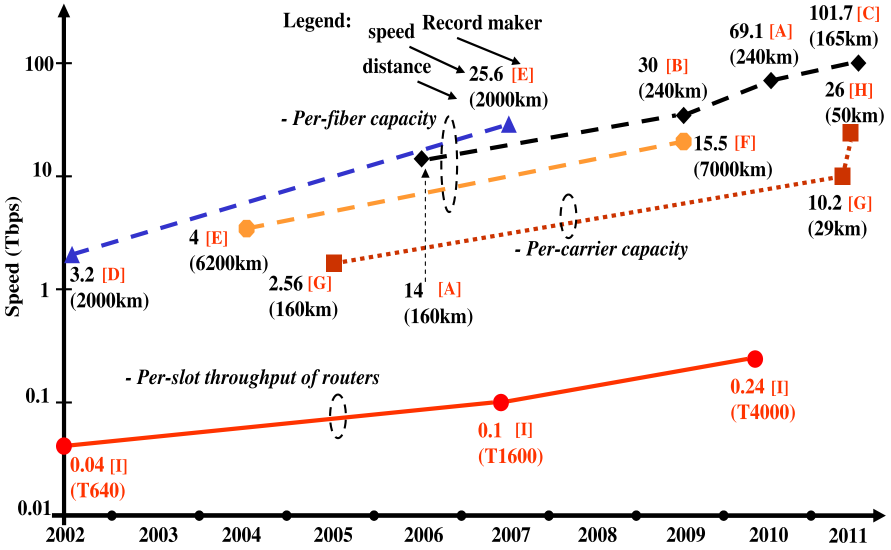


Fig. 2.6 Optical fiber capacity versus electronic router speed (Record makers: [A] = Nippon Telegraph and Telephone Corporation (NTT), [B] = KDDI R&D, [C] = NEC Laboratories in Princeton, [D] = Furukawa Electric (Japan), [E] = Alcatel-Lucent, [F] = Bell Labs, [G] = Fraunhofer Heinrich Hertz Institute (Germany), [H] = Karlsruhe Institute of Technology (Germany), [I] = Juniper Networks)

the optical transmission speed following Butters' Law [12]. Figure 2.6 depicts some optical carrier speed records and per-slot routing speeds of the electronic router T-Series from Juniper Networks [13]. We find that the routing speeds are much slower than those of the optical carrier, and the fiber speeds vary largely against the distances that the optical signal can travel without using a repeater. An inspiring fact is that the per-carrier capacity has increased rapidly recently, which will further improve the optical fiber capacity.

A carrier is associated with a wavelength or frequency. The capacity of an optical fiber is the sum of the capacities of all carriers multiplexed in this fiber by using wavelength division multiplexing (WDM) [14], which can allow multiple carriers

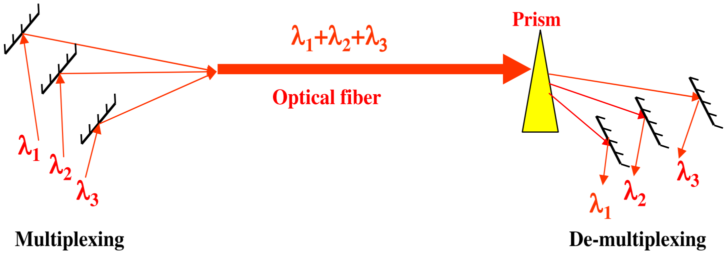


Fig. 2.7 Principle of optical multiplexing and demultiplexing

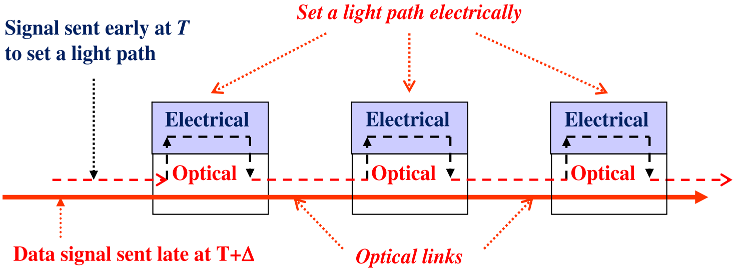


Fig. 2.8 Optical burst switching (OBS) network

to be multiplexed into one fiber and demultiplexed at the end of the fiber. As illustrated in Fig. 2.7, when it passes through a transparent medium such as a prism to another medium, light is refracted at an angle that varies with wavelength, and so a multiplexed signal can be demultiplexed in this way.

Another weakness of the OEO-based technology is its low energy efficiency due to the OEO conversion. As will be discussed in Chap. 13, energy consumption rather than transmission capacity will become one of the major hurdles that impact the Internet development, because the future Internet must be green.

2.2.2.2 Partial OEO Conversion

To avoid a full OEO conversion of every data signal for electronic processing, another approach suggests that the sender first transmit an optical signal along a dedicated light path to set a light path for data transmission. The data will then be sent subsequently in a Δ time after the path setup signal has been sent out. As illustrated in Fig. 2.8, upon arriving at a relaying unit, the path setup signal is converted into an electrical signal, which is processed electronically to set up a light path for the data transmission. In this case, no OEO conversion is carried out for the data signal, which travels in the optical domain all the way to the destination. This typical example is called optical burst switching (OBS) [15].

An OBS router can be composed of an array of mirrors, which can be tuned electrically through a micro-electro-mechanical system (MEMS) [16, 17] according

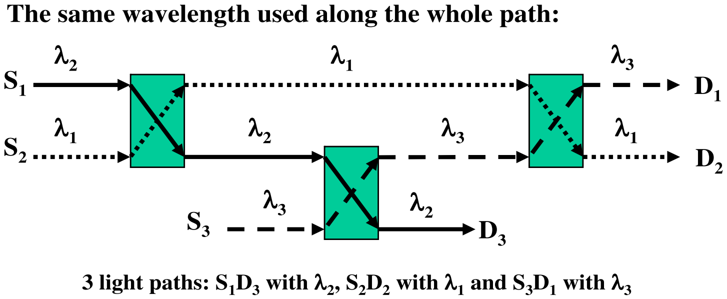

The same wavelength used along the whole path:
λ_2
λ_1
λ_3
S_1
D_1
λ_1
λ_2
λ_3
λ_1
S_2
D_2
λ_3
λ_2
S_3
D_3
3 light paths: S_1D_3 with λ_2, S_2D_2 with λ_1 and S_3D_1 with λ_3

Fig. 2.9 Wavelength routing

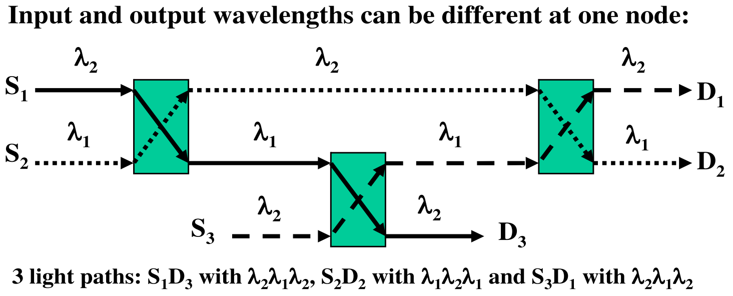
to the routing information carried by the path setup signal. The major problem with OBS is the electronic processing with the OE conversion, which may become the bottleneck of the ultra-high-speed optical fibers. The major implementation issues include how to efficiently set up a light path for successful data transmission in the optical domain, which largely depends on the Δ setting. A large Δ will waste the fiber capacity, while a small one may cause data packets to be transmitted on an unavailable light path since no ACK on path setup is returned to the sender. Furthermore, since no global coordination is carried out for such a path setup, path collisions may occur if multiple paths are competing for the same carrier on the same link, causing packet losses.

2.2.2.3 All-Optical Networking

Two methods able to fully make use of ultra-high-speed optical fibers without any OEO conversion are wavelength routing and wavelength switching. These two methods exploit the same property of light as used for the demultiplexing mentioned above. The major difference between the routing and switching here is the number of wavelengths used in constructing a light path. With routing, the same wavelength is used all the way from source to destination; with switching, the wavelength can be changed hop by hop.

Figure 2.9 shows an example of wavelength routing, where three light paths are constructed, each of which is associated with one wavelength. That is, the same wavelength must be used along the entire light path. This requirement, called the wavelength continuity constraint [18], limits the number of light paths that can be set up by a given number of wavelengths. In contrast, wavelength switching exploits wavelength converters, which can covert the wavelength associated with the arriving signal on an input link into another wavelength and transmit it through an output link. As illustrated in Fig. 2.10, for the same set of three light paths as constructed in Fig. 2.9, two wavelengths are sufficient here.

However, for both wavelength routing and switching, a major problem is that the number of wavelengths with which the light signal can efficiently travel in optical fibers is very limited in comparison with the number of source-destination pairs. In


Input and output wavelengths can be different at one node:

3 light paths: S_1D_3 with $\lambda_2\lambda_1\lambda_2$, S_2D_2 with $\lambda_1\lambda_2\lambda_1$ and S_3D_1 with $\lambda_2\lambda_1\lambda_2$

Fig. 2.10 Wavelength switching

Table 2.1 Summary of optical networking technologies

	Full OEO	Partial OEO	All-optical networking
Switching mode	Packet switching		Circuit switching
OEO level	Packet	Path setting	Control
Challenges	Slow electronic computing speed	Inefficiency of light path setting	Limited number of suitable wavelengths

this case, traffic aggregation and deaggregation need to be conducted at the network edges in order to use fewer wavelengths. To avoid OEO conversion at the time scale of a call level or packet level, a light path should be set up in advance for each node pair through a setting process with the OEO conversion at a larger time scale. However, once a light path has been set up, it cannot be adjusted dynamically on either call-level or packet-level time. Since the bandwidth of a light path is directly associated with the wavelengths, the bandwidth of a light path cannot be shared by other paths even if there is no traffic traveling along this path. This makes wavelength routing and switching essentially identical to electrical circuit switching in terms of bandwidth utilization, which cannot benefit from the traffic multiplexing that can be provided by packet switching.

2.2.2.4 Summary

Table 2.1 summarizes the major characteristics of the above-mentioned optical networking technologies. Without effective photonic computing and buffering technologies, a joint use of electronic computer and optical fibers for packet-level routing with OEO will be constrained by the electronic computing speed, which is much slower than the optical carrier speed and has a lower energy efficiency. The performance of OBS with partial OEO conversion relies on the efficiency of the light path setting. Wavelength routing and switching suffer from the limited number of wavelengths available for optical transmission.

Table 2.2 Comparison between wireless and wired networks

	Wired networks	Wireless networks
Channel reliability (BER)	About 10^{-9}	10^{-2}–10^{-5}
Channel capacity	Tbps-level or higher	Mbps-level
Channel security	Secure	Insecure
Mobility support	No	Yes
Node capability	Strong	Weak
Power supply	Unlimited	Battery-operated

2.2.3 Pervasive Networking

The ever-increasing number of mobile users greatly stimulates the development of mobile wireless networks toward supporting high mobility at high speeds as indicated in Fig. 2.11. This kind of network is usually used jointly with wired networks to provide the user with Internet access to anything anyhow, anywhere and anytime. However, as listed in Table 2.2, there are many radical differences between wired and wireless networks due to their distinct communication media. The major communication medium for wireless networks is radio, which by its nature is exposed and broadcast and thus vulnerable to interference and attack. Media for wired networks typically include metallic cables (e.g., twisted copper pairs) and optical fibers that are well protected. As listed in Table 2.2, wireless networks have the following distinguishing characteristics: low channel capacity, unreliable and insecure channels, and mobile terminals, which are less computationally powerful than stationary computers and often battery operated.

These differences are so significant that many networking issues that have been solved in wired networks must be readdressed in mobile wireless networks. A big challenge facing all wireless networks is how to efficiently use the scarce radio spectrum to support the ever-increasing number of both mobile users and mobile applications. As illustrated in Fig. 2.6, it is relatively easy for optical networks to have Tbps-level network capacity. However, for wireless networks, the fastest wireless technology available on the market is the ultra-wide bandwidth (UWB) network, which can provide up to 485 Mbps but within a distance of less than 2 meters. Furthermore, wireless network environments are usually highly dynamic due to interferences, multi-path propagation, and terminal mobility.

To achieve the above objectives in such a difficult and complex situation, much research has been conducted, and many proposals have been reported in the literature. Among them, the cross-layer design and optimization approach has recently attracted a lot of attention [19]. Unlike the layered models and the end-to-end arguments mentioned earlier, this approach tries to improve the performance of wireless networks and optimize wireless resource utilization by (i) sharing information available at different layers and (ii) redistributing networking functions or even adding new functions if necessary. In this case, changes in both the layered models and the end-to-end arguments are inevitable. Of course, some new challenging issues

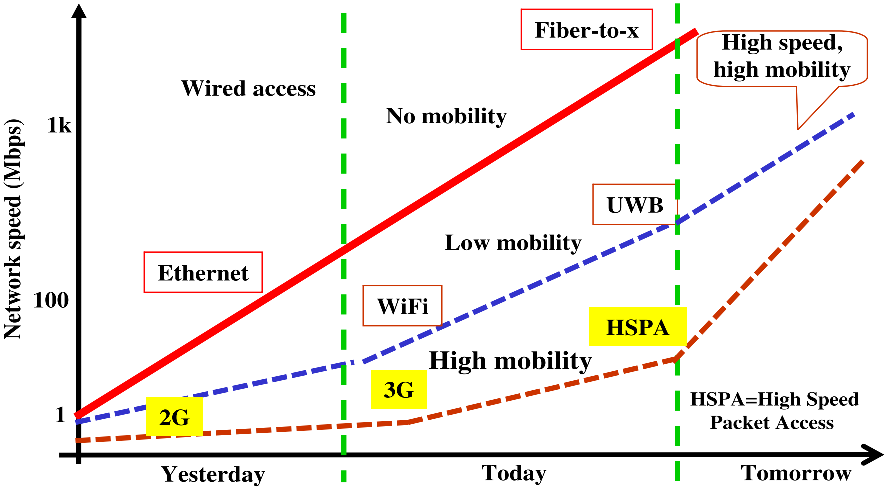


Fig. 2.11 Development of mobile wireless networks

may also arise from this approach. For example, we need to (i) study how to maintain interoperability for various cross-layer designed protocols and algorithms and (ii) determine whether the above changes are cost effective with respect to the performance gain and increased complexity caused by these changes.

2.2.4 Quality of Service

As the number of applications is continuously increasing, especially real-time and multimedia applications, an essential issue to be addressed for future networks is how to efficiently satisfy the quality of service (QoS) of various applications in a cost-effective way. QoS has been studied for more than two decades, especially for ATM and IP networks. ATM is regarded as a reference model for QoS support; many approaches originating from ATM are applied today for QoS support in other networks (e.g., IP) such as scheduling and call admission control. However, ATM is a connection-oriented network and is more complex than the connectionless IP network, leading to IP's dominance in today's Internet. This fact further stimulates the all-IP network approach, in which only IP is used as the network protocol in the network layer.

2.2.4.1 QoS Capability of IP

QoS support for the IP network has not been well resolved, since IP was originally designed to only provide a best-effort service at the packet level with the connectionless networking mode. Thus, IP has a property that is unfavorable for QoS support:

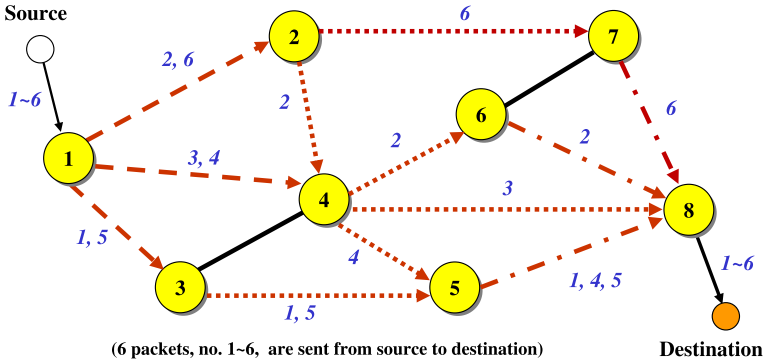


Fig. 2.12 Example of dynamics of IP routes

at the packet level, it is impossible for a router to make any resource reservation for a packet that is going to visit it, since the router cannot know whether there is a packet visiting until the packet arrives. IP has been enhanced by introducing the flow approach on the top of packets. A flow is a series of packets traveling from source to destination. However, at the flow level, the route of a flow is not pinned [20] and is often adjusted during the flow lifetime to balance traffic loads and maximize fault tolerance capability. Therefore, there is not a dedicated route for a traffic flow between a source-destination pair since each packet is routed on-the-fly according to the route dynamically set by routing protocols. Figure 2.12 gives an example of this phenomenon, which shows that a series of packets may travel along different routes between node 1 (source) and node 6 (destination).

The best-effort service of the original IP using the FIFO scheduling policy has also been enhanced by using sophisticated output schedulers such as class-based scheduling. The flow-level resource negotiation and reservation method has also been standardized by the Resource Reservation Protocol (RSVP) [21, 22]. Call admission control (CAC) has also been proposed in order to further enhance IP's QoS capability. Actually, these efforts to improve the QoS capability of IP more or less follow ATM's philosophy for QoS provisioning. Note that ATM is a connection-oriented network, but the above enhancements to IP do not change its inherent properties, i.e., those of a connectionless network with unpinned routes. In this case, it is not cost effective to make a hard resource reservation for a flow to guarantee its QoS since the reserved resource is wasted if the route of the flow is changed. Therefore, RSVP only provides soft-state resource reservation, i.e., an amount of network resource reserved for a flow may be automatically released after a pre-defined time period even when the flow is still alive.

2.2.4.2 Per-Flow IntServ Versus Per-Class DiffServ

Two typical mechanisms proposed for IP to enhance its QoS capability include the per-flow IntServ [23] and the per-class DiffServ [24]. Similar to ATM, IntServ tries to provide granular QoS at the flow level by reserving resource for each flow, while DiffServ tries to provide QoS at the class level by aggregating flows with the same type of QoS requirements. However, the per-flow IntServ has a scalability problem in the case of a large number of flows present at a router, especially those in core networks. This occurs because the router needs to store per-flow QoS information for QoS provisioning. Although the per-class DiffServ can overcome this problem through flow aggregation, it cannot cost-effectively provide QoS support for each individual flow, since it must satisfy the individual flow with the most stringent QoS requirement in the aggregated flow.

To trade off between QoS granularity and implementation complexity, a combination of IntServ and DiffServ has also been reported [25]; i.e., DiffServ is used in core networks and IntServ in edge networks, such as access networks. In this combination, QoS parameter conversion must be carried out at the boundary between these two types of networks. Thus, this combination is not a seamless end-to-end QoS solution, and the conversion will cause extra delay and energy consumption. On the other hand, both IntServ and DiffServ as well as their hybrid largely depend on sophisticated output schedulers for QoS provisioning. Both parameter conversion and output scheduling will increase the implementation complexity as QoS granularity and the number of flows increase. The conversion point may become the performance bottleneck of end-to-end QoS provisioning, while a sophisticated output scheduler may become the bottleneck of high-speed links, since they have to make a decision on-the-fly for every incoming packet.

The adoption of CAC, resource reservation, and sophisticated schedulers for QoS provisioning very much increases the implementation complexity, which violates the simplicity principle of the original IP following the end-to-end arguments. Therefore, the capacity over-provisioning approach is proposed and implemented in practice to simplify QoS provisioning for IP by using a more-than-need network capacity [26]. Obviously, this approach wastes network bandwidth and is not green.

2.2.5 Network Security

Similar to QoS support, the Internet was initially designed without considering network security. This happened because the overwhelming majority of users at that time were people from governmental and institutional organizations, who are usually well educated and trustworthy. Thus, a lot of sensitive information such as user names and passwords for FTP and Telnet were sent in a plaintext over the Internet at that time. However, as the Internet has gone public, especially for commercial applications such as e-commerce and e-banking, attacks on the Internet are continuously increasing, causing immense economic and social damage. Thus, IP Security (IPSec) has been proposed to provide "access control, connectionless integrity, data origin authentication, rejection of replayed packets and confidentiality" [27]. Basically, IPSec builds a secure sublayer on top of the connectionless IP to provide secure network services for applications in higher layers.

Actually, it is almost impossible for a network to completely prevent a packet from being intercepted or modified during the journey to its destination, especially when traveling in a shared-media network such as wireless networks—everything transmitted over the air can be received by other parts due to its broadcast nature. Therefore, the actual confidentiality and integrity protection are realized through cryptographic mechanisms. That is, for confidentiality, if a packet is intercepted, it cannot be understood without a decryption using the proper key. Similarly, for integrity, if any modification is incurred to a packet, this incident can be detected by the receiver. However, these cryptograph-based protections in the network layer can also be equally provided by the transport layer on an end-to-end basis, and similarly for other secure services such as repudiation, which ensures that the sender of a message can be identified.

The confidentiality and integrity protections provided by the network layer are in an embarrassing situation. On one hand, many applications that do not want such protection avoid using these functions due to their computational expense. On the other hand, those applications requiring these security protections prefer to adopt the same protection provided by the transport layer since they are end-to-end based. This is because the application layer cannot be assured that the same protection can be provided by every network segment all the way from source to destination. If any of them fails in doing so, the end-to-end security may be compromised. Therefore, the network layer should focus on providing security protections that cannot be provided by higher layers.

Similar to the efforts made to enhance IP's QoS capability discussed above, IPSec does not change the connectionless nature of IP, which allows packets to be transmitted anytime to anywhere without requiring a connection setup. In this case, neither the sender nor the receiver has the chance to authenticate each other before any transmission incurs between them in the network layer. Furthermore, the destination has no way to control traffic loads approaching it and cannot judge whether the IP address of a sender is genuine. These features make it easy to launch many attacks, such as denial of service (DoS) attacks [28]. Since every packet is forwarded along a route that is dynamically set on-the-fly by routing protocols, attacks to routing information such as IP addresses or routing tables may lead to incorrect delivery of IP packets [29], probably leading packets to be sent to wrong receivers or to circulate in the network.

Today the Internet has become an indispensable part of our society, and Internet security has been delegated as a part of national security in many countries. Since both the number and type of users and applications are continuously increasing, it is still an open issue whether the enhanced IP security can satisfy the security requirements of a complex, important and even green cyberspace.

2.2.6 Green Networking

Due to the increasing number of disasters caused by global warming, carbon emission has become an important issue of almost every sector in the world today. As reported in [30], the energy consumed by the information and communication technology (ICT) sector is more than that consumed by the aviation sector. There is no doubt that the Internet has achieved great success over the past three decades. However, during this development, the privileged issues addressed by network researchers and designers have been mainly networking performance and reliability—energy efficiency has been more or less ignored.

High requirements on security and energy efficiency in the future Internet pose big challenges to network design. Therefore, it is necessary to conduct a timely revisit of the current networking modes with respect to the new requirements raised by the future Internet discussed above. We will discuss this issue in more detail in Chap. 13.

2.3 Conclusion

Table 2.3 summarizes the major characteristics of the Internet at different stages of its development. It is straightforward to understand the parts of the initial and current Internet, which have been discussed earlier. For the Internet in the future, we list some possible changes in both users and applications as well as networking modes that may be incurred to satisfy new requirements in the future.

First of all, "any media" is used here to indicate that unlimited types of applications may run over the future Internet, and the users will not be limited to humans only, but also objects. For example, the Internet of Things (IoT) tries to interconnect everything through the Internet. In terms of network design principles, cross-layered design and optimization should be applied in addition to the original layered models. Simplicity is always one of the major design objectives, but sometimes a tradeoff between simplicity and other important indicators is necessary. Although we do not know yet if it is necessary to invent new protocol units, there is no doubt that the packet and flow will still be basic protocol units in the future. Since both connectionless and connection-oriented networks have their own advantages and disadvantages, they should be used jointly to efficiently support various types of applications in the future Internet. For the QoS capability, more efforts should be made to improve QoS granularity for high cost-effectiveness and energy efficiency. Meanwhile, the structure of QoS support should be scalable in order to cost-effectively satisfy the QoS requirements of any media. Therefore, both soft- and hard-state resource reservations should be provided for different resource requirements. To this end, some new mechanisms need to be devised.

Network security and mobility support as well as green networking should become a cohesive part of the core network structure. Since optical fibers can provide much more immense capacity than traditional media cost-effectively, and also provide higher energy efficiency, optical networking is a promising technology for both wired and wireless networks. For wireless, the radio over fiber (RoF) technology can be applied to improve wireless coverage by using optical fibers. However, some new technologies need to be developed in order to narrow the gap between optical

Table 2.3 Evolution of network technologies for the Internet

	Initial Internet	Current Internet	Future Internet
Applications	Data (e.g., e-mail FTP, Telnet)	Multimedia (e.g., data voice, video, cybergames)	Any media (e.g., IoT cloud computing, new apps)
User characteristics			
Type	Limited types	Anyone	Anyone+anything
Number	Small number	Large number	Enormous number
Mobility	Stationary	Stationary+mobile	
Networking modes			
Vertical	Layered models		Cross-layered models
Horizontal	End-to-end arguments		Simplicity+tradeoff
Unit	Packet	Packet+flow	
CL/CO	Connectionless (CL)		CL+Connection-oriented
Stability	Unpinned routes		Unpinned+pinned routes
QoS structure			
QoS type	Best effort	Per-flow, per-class	Granularity+scalability
Scheduling	FIFO	Output scheduler	New mechanisms
Resource	No reservation	Soft-state reservation	Soft+hard reservation
Additional features			
Mobility	Not	Improved	Fully
Security	Supported	On top of IP	Integrated
Green	Not considered or optional		Compulsory
Enabling technologies			
Processing	Electronic computing+buffer		Optical processing+buffer
Media	Metallic cables	Cables+optical fiber	Optic fiber
Routing	Routing	OEO-based routing	All-optical switching
Switching	Switching	Wavelength routing	New methods

networking requirements and immature photonic computing and buffering technologies that are unable to effectively support optical networking. More discussions on these issues will be provided in the remainder of the monograph.

References

1. Stallings, W.: Data and Computer Communications, 5th edn. Prentice-Hall, New York (1997)
2. Tanenbaum, A.S.: Computer Networks, 4th edn. Pearson Education, Upper Saddle River (2003)

3. Saltzer, J.H., Reed, D.P., Clark, D.D.: End-to-end arguments in system design. ACM Trans. Comput. Syst. **2**(4), 277–288 (1984)
4. Postel, J.: Transmission control protocol. IETF RFC 793 (1977)
5. Clark, D.: The design philosophy of the DARPA Internet protocols. In: SIGCOMM '88: Symposium Proceedings on Communications Architectures and Protocols, pp. 106–114 (1988)
6. Internet World Stats – Usage and Population Statistics. http://www.internetworldstats.com
7. Available on line at http://www.mobilemarketingwatch.com
8. Available on line at http://www.telecoms.com
9. Hayes, B.: Cloud computing. Commun. ACM **53**(7), 9–11 (2008)
10. International Telecommunication Union (ITU): ITU Internet reports 2005: the Internet of things – executive summary (2005). http://www.itu.int/internetofthings
11. Schaller, R.R.: Moore's law: past, present and future. IEEE Spectr. **34**(6), 52–59 (1997)
12. Robinson, G.: Speeding net traffic with tiny mirrors. EE|Times News & Analysis **26** (2000)
13. Available on line at http://www.juniper.net
14. Mukherjee, B.: Optical WDM Networks. Springer, Berlin (2006)
15. Qiao, Q.M., Yeo, M.: Optical burst switching (OBS) – a new paradigm for an optical Internet. J. High Speed Netw. **8**(1), 69–84 (1999)
16. Wu, M.C., Li, F., Lee, S.S.: Optical MEMES: huge possibilities for lilliputian-sized devices. Opt. Photonics News, 25–29 (1998)
17. Rebeiz, G.M.: RF MEMS: Theory, Design, and Technology. Wiley, Hoboken (2003)
18. Borelia, M.S., Jue, J.P., Bankerjee, D., Ramamurthy, B., Mukherjee, B.: Optical components for WDM lightwave networks. Proc. IEEE **85**(8), 1274–1307 (1997)
19. Lin, X.J., Shroff, N.B., Srikant, R.: A tutorial on cross-layer optimization in wireless networks. IEEE J. Sel. Areas Commun. **24**(8), 1452–1463 (2006)
20. Handley, M.: Why the Internet only just works. BT Technol. J. **24**(3), 119–129 (2006)
21. Braden, R., Zhang, L., Berson, S.: Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification. RFC 2205, Internet Engineering Task Force (1997)
22. White, P.P.: RSVP and integrated services in the Internet: a tutorial. IEEE Commun. Mag. **35**(5), 100–106 (1997)
23. Braden, R., Clark, D., Shenker, S.: Integrated services in the Internet architecture: an overview. IETF RFC 1633 (1994)
24. Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., Weiss, W.: An architecture for differentiated services. IETF RFC 2475 (1998)
25. Chang, I.C., Chen, S.F.: An end-to-end QoS adaptation architecture for the integrated IntServ and DiffServ networks. In: IFIP Int. Federation for Information Processing, pp. 365–376 (2007)
26. Menth, M., Martin, R., Charzinski, J.: Capacity overprovisioning for networks with resilience requirements. In: Proc. ACM SIGCOMM, Pisa, Italy (2006)
27. Kent, S., Atkinson, R.: Security architecture for the Internet protocol. IETF RFC 2401 (1998)
28. Zhou, C.F., Leckiea, C., Karunasekera, S.: A survey of coordinated attacks and collaborative intrusion detection. Comput. Secur. **29**(1), 124–140 (2010)
29. Butler, K., Farley, T.R., McDaniel, P., Rexford, J.: A survey of BGP security issues and solutions. Proc. IEEE **98**(1), 100–122 (2010)
30. Tucker, R.S.: A green Internet. In: Proc. Annual Meeting of the IEEE Lasers and Electro-Optics Society, Acapulco, Mexico, pp. 4–5 (2008)