There is no doubt that the decades-long evolution ofmicroprocessor and storage technologies, computer

architecture and software systems, and parallel algorithms and distributed control strategies paved the

way for cloud computing. Yet we have to recognize that the fascinating developments in networking

are at the heart of this new technology. In this section we only discuss those features of the Internet

affecting cloud computing.

The Internet is continually evolving under the pressure of its own success and the need to accommodate

new applications and a larger number of users; initially conceived as a data network, a network

supporting only the transport of data files, it has morphed into today’s network for data with real-time

constraints for multimedia applications. To understand the architectural consequences of this evolution,

we examine first the relations between two networks:

• Peering, in which the two networks freely exchange traffic between each other’s customers.

• Transit, in which a network pays another one to access the Internet.

• Customer, in which a network is paid money to allow Internet access.

Based on these relations the networks are commonly classified as Tier 1, 2, and 3 (Figure 7.4). A

Tier 1 network can reach every other network on the Internet without purchasing IP transit or paying

settlements; examples of Tier 1 networks are Verizon, ATT, NTT and Deutsche Telecom.

A Tier 2 network is an Internet service provider (ISP) that engages in the practice of peering with

other networks but that still purchases IP transit to reach some portion of the Internet. Tier 2 providers

are the most common providers on the Internet. A Tier 3 network purchases transit rights from other

networks (typically Tier 2 networks) to reach the Internet. A point-of-presence (POP) is an access point

from one place to the rest of the Internet.

An Internet exchange point (IXP) is a physical infrastructure allowing ISPs to exchange Internet

traffic. The primary purpose of an IXP is to allow networks to interconnect directly via the exchange

rather than through one or more third-party networks. The advantages of the direct interconnection

are numerous, but the primary reasons to implement an IXP are cost, latency, and bandwidth. Traffic passing through an exchange is typically not billed by any party, whereas traffic to an ISP’s upstream

provider is.

IXPs reduce the portion of an ISP’s traffic that must be delivered via their upstream transit providers,

thereby reducing the average per-bit delivery cost of their service. Furthermore, the increased number

of paths through the IXP improves routing efficiency and fault tolerance. A typical IXP consists of one

or more network switches, to which each of the participating ISPs connects.

New technologies such as Web applications, cloud computing, and content-delivery networks are

reshaping the definition of a network, as shown in Figure 7.5 [203]. TheWeb, gaming, and entertainment

are merging, and more computer applications are moving to the cloud. Data streaming consumes an

increasingly large fraction of the available bandwidth as high-definition TV (HDTV) sets become

less expensive and content providers such as Netflix and Hulu offer customers services that require a

significant increase in network bandwidth.

Does the network infrastructure keep up with the demand for bandwidth? The Internet infrastructure

in the United States is falling behind in terms of network bandwidth, as illustrated in Figure 7.6. A

natural question to ask is:Where is the actual bottleneck that limits the bandwidth available to a typical

Internet broadband user? The answer is: the “last mile,” the link connecting the home to the ISP network.

Recognizing that the broadband access infrastructure ensures continual growth of the economy and

allows people to work from any site, Google has initiated the Google Fiber Project, which aims to

provide 1 Gb/s access speed to individual households through FTTH.3