1 Introduction: telecommunications evolution and the set of actors

1.1 The evolution of telecommunications and the associated economic models

Telecommunication networks are occupying an increasing role in our daily life: almost everything is now available from the Internet (possibly via a mobile phone), and getting this kind of access has even become compulsory for some administrative operations, without mentioning the social pressure to be part of the trend. While the telephone network started commercially in 1877 following the birth of the Bell Telephone Company, and its development slowly democratized, mobile networks have since the 1970s quickly come to occupy a major place, with an estimated 5.3 billion mobile subscriptions worldwide at the end of 2010. It is also estimated by the International Telecommunication Union (ITU), in its report "The World in 2010," that 90% of the global population can reach a wireless network. In the same report, the Internet is also shown to be an increasing market worldwide, with the number of subscribers having doubled between 2005 and 2010, reaching more than two billion at the end of 2010 (1.6 billion having access at home). To highlight this growth, the average household telecommunications budget has considerably increased, counting wired and wireless telephony subscriptions (with often a wireless subscription per member of the family), Internet access, television, etc., with a role that is becoming compulsory for operations such as access to information and to documentation, and making declarations (taxes, etc.). E-commerce, namely the buying and selling of products or services over electronic systems such as the Internet and other computer networks, is also exploding: according to Goldman Sachs, global Internet sales are growing at about 19% a year, and are expected to reach almost one trillion dollars in 2013; see Figure 1.1. Also, the average time spent on the Internet keeps increasing: comScore Media Metrix estimates that the average American spent 32 hours per month online in 2010, and this is now surpassing the time spent watching TV.

While the above illustrates the increasing importance of telecommunications from the users' and content points of view, the networks themselves have evolved considerably [238]. The Internet is the key network of our analysis, being the one over which all networks converge. Although we do not intend to provide a complete description of the Internet's history, we wish to give a broad idea of it, since its economic model is highly related to the evolution of the network structure. The general idea of the Internet started in the early 1960s, with the principle of connecting computers in order to share

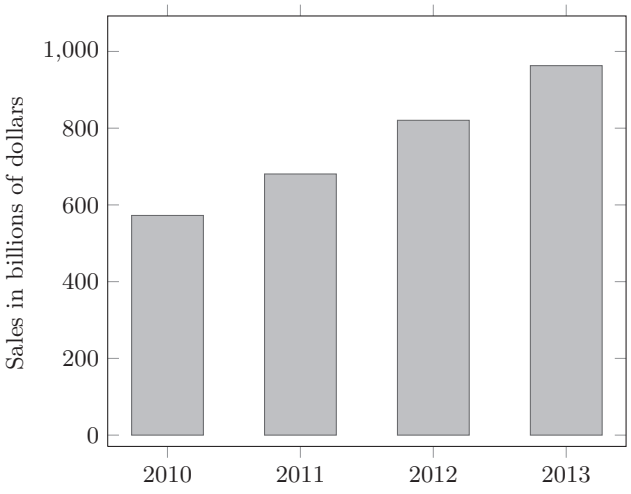

Figure 1.1 Expected evolution of e-commerce sales (Goldman Sachs data).

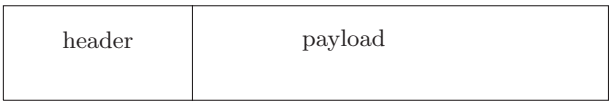

Figure 1.2 Packet description.

information and resources for research, but also in order to provide a robust network to the US military. The project was led by J. C. R. Licklider, and developed by the Defense Advanced Research Projects Agency (DARPA). The key **packet-switching** method for transmitting data consists in cutting the information into **packets** sent one after the other. Basically, a packet is made of two parts: the information (also called *payload*) and the **header**; the header contains fields such as source and destination address, data length and type, etc.; see Figure 1.2. See [154, 295] for more details. This is in contrast to the **circuit-switching** method used in telephony, for which a **circuit** is a communication channel along one line, which is fully used by the conversation. Packets have the advantage of being easily built and transmitted over the telephone network, and of allowing more reliable transmission, thanks to retransmission if needed. Two computers were effectively connected between the Massachusetts Institute of Technology and California in 1965 using dial-up telephone lines and packet switching, and the network sponsored by DARPA was called ARPAnet. ARPAnet allowed universities and research labs to share the cost of long-distance telephone lines. Still-in-use applications (such as FTP for example) and routing and addressing protocols were then imagined. Owing to the success of the network, other universities wanted to get connected, and created their own networks using the same protocols, so that the networks could interoperate. Thanks to the creation of Ethernet, which permitted one to cheaply interconnect computers on campuses, the network could expand, but was still just for non-commercial use. In the early 1990s, the Internet was opened to commercial organizations, which stepped in by

Table 1.1 IP traffic worldwide, measured in petabytes per month (data from Cisco Systems)

Year	Total global IP traffic	Fixed Internet traffic	Mobile IP traffic
1990	0.001	0.001	–
1995	0.18	0.17	–
2000	84	75	–
2005	2,426	2,055	0.9
2010	20,197	14,929	256
2011	27,483	20,634	597

building their own networks (especially telecommunication operators) and/or using the existing network. As illustrations, we have network operators that propose access at home to end-users, and connect them to the core network; content providers, which propose some content and get revenue by direct sales or through advertising; and more recently content delivery networks that provide resources to content providers for delivering their content on the network, social network applications such as Facebook, etc.

We can remark that, in most countries, telephony was initially a government monopoly, a situation that has changed almost everywhere. This analogy with the Internet and its interconnection of academics not interested in a business model, and now transformed into a commercial network and supply chain (from content creation to delivery to users), can be investigated to understand the pitfalls to avoid. In terms of business models for communications, it is interesting to note that Bell System had already compared the advantages and drawbacks of flat rate and metered rates for telephony in the 1880s, and preferred metered rates in large cities over flat rates, to cover the high marginal costs. Even if this situation is not necessarily true for the Internet now, studying the most "appropriate" charging scheme can lead to changes in the economic vision of the network.

We have mentioned the telephone network and the Internet, but wireless telephony (and communications) has grown and continues to grow significantly, and represents a prominent part of the economic business and technological works currently developed. Wireless communications are made through cellular networks (first 2G, then 3G encompassing data and Internet), WiFi, and now 4G/LTE transmissions. Those technologies will also be discussed in the book.

The success of the Internet and wireless networks, with an increasing number of subscribers but also because of more and more demanding applications in terms of bandwidth and resources, has seen a tremendous increase of traffic worldwide, as illustrated by Table 1.1. The traffic growth is often considered to be exponential, though it is estimated by Cisco that the yearly Internet traffic growth in the USA will fall from 42% in 2010 to 18% in 2014, and from 42% to 30% worldwide (a smaller reduction due to the later uptake in developing countries). To better check the evolution of this traffic increase, we plot the volume of traffic year by year from 1990 to 2011 in Figure 1.3. One can see that, even though the traffic is still increasing, the slope is slowing down (except for mobile communications, which are still in their infancy in terms of data

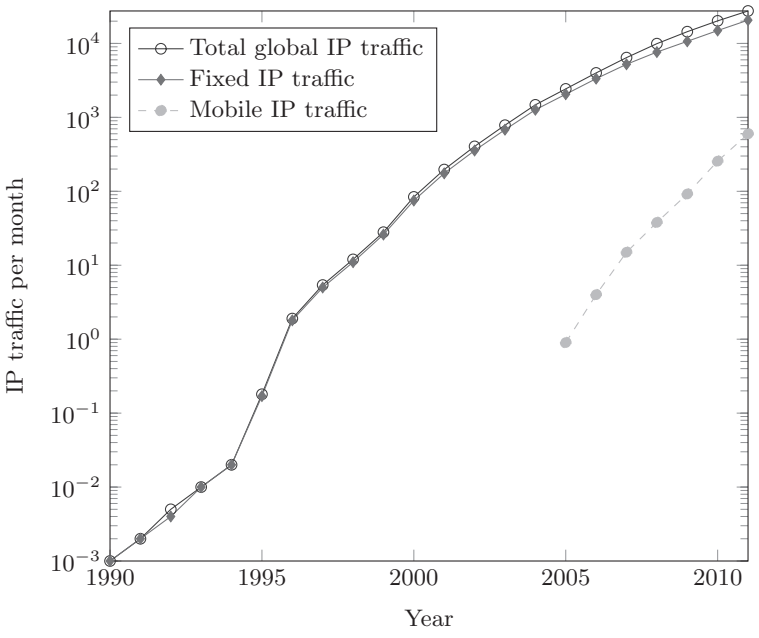

Figure 1.3 IP traffic evolution in petabytes per month, on a logarithmic scale (Cisco data).

transfer). The generally accepted statement that there is a doubling of traffic every year is no longer valid.

As a consequence of all that we have described, the initial (actually, non-existent) business model of the Internet, with free interconnection between academics, which still partly applies, is more and more being questioned: some ways to make revenues have to be defined, in order to at least cover the capacity expansion costs. Another complication is that the economics of the Internet are very related to the economics of (wired) telephony. Indeed, the Internet has experienced an easy development, and a rapid success, thanks to the use of the telephony infrastructure, even with its bandwidth restrictions. Therefore, a change in the business model of the Internet has to take into account the history of pricing mechanisms of the Internet, but also of telephony, in order to seek a better acceptance from users.

Among the many questions that the telecommunication network actors need to answer, we can non-exhaustively mention the following.

- *Determining the most relevant and profitable network access pricing scheme for end-users*, in a competitive context between network providers, also called Internet service providers (ISPs).

  One of the main aspects that could explain the success of the Internet is the fact that users just pay a subscription fee to the network, and can freely use it as much as they want, the so-called *flat-rate* pricing scheme. This scheme is/was very attractive to users, since they know exactly what they are going to pay and do not have to bother about usage. But, due to the tremendous development of traffic volume, users started

to complain around the 1990s about data transfer times becoming too large without there being any possibility to improve this situation by themselves, by paying more, for instance. The problem arose from the congestion on communication links, and many people thought that link capacity expansion was not a viable solution because the network must respond to an increasing demand, and they claimed that experience had shown that demand for bandwidth has always been ahead of supply. Furthermore, in this context of congestion, flat-rate pricing is considered unfair since users with a huge consumption level pay the same amount as "light" users, while congesting the network and lowering the quality of service for all users. It was therefore suggested that one should replace the flat-rate fees by usage-based fees. Both flat-rate and usage-based schemes have their own respective advantages: first, flat-rate pricing is easy to implement, not requiring any costly measurement, and well-accepted by users. On the other hand, usage-based pricing allows better control of network usage, and makes it possible to differentiate services for potentially higher revenues and user satisfaction. There was a really strong research activity in the 1990s and early 2000s directed towards the design of various usage-based pricing mechanisms (some of which will be described in Chapter 3), but this activity slowed down afterwards, especially with the over-provisioning of the core network with optic fiber. However, usage-based pricing is coming back because of the increase of data traffic in wireless networks, as highlighted by Figure 1.3, for which bandwidth is scarce and more difficult to increase. Similarly, there is a push (mainly from Internet service providers) towards usage-based pricing: for instance in Canada in early 2011, usage-based pricing was enforced by the regulator, the Canadian Radio-television and Telecommunications Commission (CRTC), following the request of providers Bell, Rogers, Videotron, Cogeco, and Shaw, which were expecting to earn more. A lot of protests from user associations and content providers followed. Indeed, there is a strong public preference for flat-rate pricing [89], since users are more comfortable with this principle even though it is often said that usage-based pricing would mean a lower overall bill. Indeed, people always think about the possibility of consuming more, even if they will eventually not do so.

Another issue arising from the pricing scheme definition is related to network convergence: now wired and wireless telephony, Internet, and television are provided by the same operators. Because of competition, those operators merge the services into single offers, called *bundles*. Triple play offers that combine telephony, television over Internet, and Internet access, have become the norm, wireless telephony being added, too, for a quadruple play offer. Those offers and their associated prices have to be studied, taking care that a bundle offer would not cannibalize other potential offers, at the expense of the provider's profits.

- *Determining the best investments for network providers.*

  Network service providers have several strategic decisions to make, among which are those concerning investment in new technologies, and capacity and infrastructure expansions, but also participation in spectrum auctions. Those decisions are very important ones, because of the enormous costs involved, and the consequences of bad strategies can be economically very damaging for a company.

Indeed, providers should not only decide carefully the access price they will impose (the topic of the previous item) but also on which technologies to operate. They need, for example, to decide whether to invest in terms of infrastructure in new technologies such as the third generation (3G) of wireless networks (antennas, routers), and/or the fourth generation (4G) such as LTE (Long Term Evolution) or WiMAX. These investments should allow them to attract customers in order to get new sources of revenue. The LTE technology which is starting to be deployed is expected to lead to new broadband services, encouraged by mobile devices with larger screens, better batteries, and better performance. For instance, standard e-mails and SMS are moving towards photo messages, instant messaging, and video messaging, as well as social networking thanks to Facebook and Twitter, among others. As a consequence, investing should foster customers' subscriptions, but it has to be investigated whether the expected gains exceed the costs. Looking at the full picture, taking into account potential investments of competitors, would an investment by other providers reduce the market share too much? Will it help to segment the market and propose specific offers? Is it worth it for an operator to pay a license and devote resources to infrastructure in order to be present in a new technology? Will the return on investment be sufficient, and will it not be at the expense of other technologies already implemented and with limited maintenance costs?

Again, and to emphasize this issue, infrastructures are not the only required investment when dealing with wireless networks. Radio spectrum is indeed sold in most countries through auctions. Bidding to get a part of this spectrum is a costly and very strategic decision that has to be taken with caution. The auction run in 1999 for the radio spectrum in the USA was considered a disaster and had to be re-run after being declared null and void, because all the major bidders defaulted and declared bankruptcy (they could not bear the costs): the industry was far too speculative. Designing a fair auction mechanism is a key issue for regulators, in order for the governments to get as much revenue as possible but also to make sure that the providers' operations will not be endangered. Another complicated situation occurred also in France in 2009, when the country opened a fourth 3G license in order to foster competition (this license was won by the operator called Free). The price and conditions had to be decided in such a way as to let the entrant survive in confrontation with incumbents with existing infrastructure and licenses. This new license was offered at a lower cost than the initial ones, but the decision had to be as fair as possible for all actors.

Investing in new technologies is not the only strategic question. When a technology has already been implemented, providers have to decide whether or not, and, if so, when, to invest in capacity expansion, and when to cease a service. This decision has to forecast the evolution of demand and of capacities, and ponder the available options.

- *Defining the economic relations between network operators.*

  There are different types of network operators, organized in tiers and defining a hierarchical Internet. The characterization of an operator is not very easy, but it is considered that a Tier-1 operator is an operator network that is able to reach any other portion of the Internet without paying settlements. In other words, it peers with all other Tier-1 networks. A Tier-2 operator, on the other hand, is a network that peers with some networks, but pays transit fees to reach at least some portion of the Internet. Finally, a

Tier-3 operator is a network that is only buying transit fees to reach the Internet. The transit costs between those providers have to be carefully computed, and here too the regulator can intervene to ensure that a dominating network does not impose its rules on others at the expense of competition and users. These charging agreements structure the Internet in a tiered architecture, at the top of which Tier-1 operators dominate the market.

Among new questions, there is now a huge asymmetry of traffic exchanges due to some content providers, such as Google via YouTube, producing a non-negligible proportion of total Internet traffic. As a consequence the standard peering agreements are now less relevant. The economic model between providers is now being rethought. The network providers also think about which other provider they should be connected to. For instance, Tier-2 providers are starting to connect to each other to avoid the costs of going through Tier-1 providers. Similarly, big content providers are starting to become network providers to avoid paying the connection fees. Therefore we are leaning towards a flatter (or more meshed) Internet with more direct interconnections because there is more traffic originating from very few content providers or content delivery networks such as YouTube and Akamai (with the increasing importance of video streaming). This may reshape the Internet economy.

- *Understanding the relations between content providers and ISPs.*

  The previous item naturally leads to the question of relations between competitive and profit-seeking content and network service providers. Up to now, because of the historic non-profit organization of the Internet, the main principle driving the network has been the *universal access* principle, meaning that all consumers are entitled to reach meaningful content, whatever the technical limitations of their service. Here also, the fact that there are large and bandwidth-consuming content providers (e.g., YouTube) connected to a limited number of ISPs means that the competitive ISPs to which they are not connected, which therefore are not directly getting money from them, are starting to wonder why distant content providers should not be charged by them, with the threat that their traffic will not be delivered if they refuse to pay [168]. YouTube, for instance, is accessed by all users while being hosted by a single Tier-1 ISP. This issue was the starting point of the *network neutrality* debate which was launched at the end of 2005 by Ed Whitacre (CEO of AT&T) saying that content providers should also be charged by ISPs to which they are not directly connected:

  > How do you think they're going to get to customers? Through a broadband pipe. Cable companies have them. We have them. Now what they would like to do is use my pipes free, but I ain't going to let them do that because we have spent this capital and we have to have a return on it. So there's going to have to be some mechanism for these people who use these pipes to pay for the portion they're using. Why should they be allowed to use my pipes? The Internet can't be free in that sense, because we and the cable companies have made an investment and for a Google or Yahoo! or Vonage or anybody to expect to use these pipes free is nuts.

  The underlying concern is that investments are made by ISPs, but content providers appropriate a large part of the dividends. The revenue arising from online advertising (i.e. showing graphical ads on regular web pages) is estimated at approximately

$24 billion in 2009, while textual ads on search pages have led to a combined revenue of $8.5 billion in 2007, with those figures increasing every year. Meanwhile, transit prices – which constitute the main source of revenues for transit ISPs – are decreasing and predicted to be under $1 per Mbps (per month) by 2014. ISPs argue that there is insufficient incentive for them to continue to invest in the network infrastructures if most benefits go to content providers. Another behavior of ISPs has been to lower the quality of traffic coming from distant content providers. For instance, Comcast, one of the main ISPs in the USA, started in 2007 to block P2P applications such as BitTorrent, using the argument that P2P is mostly used to share illegal content. Advocates of neutrality argued here that P2P has legitimate uses and that other types of initiatives should be imagined. In all cases, the goal of ISPs is to change the current behavior of the Internet, where all users (of whatever type) have full access to the network with the same quality at a flat-rate fee.

This threat of modifying the current Internet model has led to a lot of protests from content providers and user associations arguing that charging for content or blocking some types of flow is an impingement of freedom of speech and/or human rights that will impact the network development. The relevance of arguments from both sides needs to be investigated, to see whether service differentiation should be allowed, and, if the answer is in the affirmative, at which level and how far it should be implemented. This issue has already been a subject of debate at the legal and political level. In the USA the trend was first to go with imposing network regulation on ISPs to ensure neutrality, but it is not clear who, if anyone, has the authority to regulate the Internet (for instance, the Federal Communications Commission (FCC) lost a lawsuit where the Supreme Court found that it (the FCC) lacked the authority). The Federal Trade Commission (FTC) released in 2007 a report not supporting neutrality constraints, increasing the debate at the political level. This debate is also active in the European Union, as illustrated by the open consultation on network neutrality launched in 2010 at the EU level but also in each country.

- *Defining the economic model of content and (application) service providers.*

  Content and (application) service providers have two main options for getting revenue: either they charge users for access to the content, or the access is free, but advertisements are inserted thanks to a banner on the displayed web pages, and it is the advertisers who remunerate the content provider. A typical example of a service provider with an access charge is Netflix, an American provider of on-demand Internet streaming media that allows users to watch an unlimited amount of movies and TV episodes over the Internet for a monthly fee. On the other hand, we have, for instance, many newspapers, which are used to provide articles freely on the web, but display ads on their web pages. However, that trend is changing, as some newspapers start to charge for the content found on their web sites. A trade-off (as applied by the French newspaper *Le Monde*) is to publish a limited amount of information for free, while the full newspaper is available for a price (15 euros per month in 2012). According to Peter Barron of Google, "the future is going to be a mixture of paid-for content behind pay walls and free content." As we said, ads were inserted in horizontal banners, but now you can find them on other areas of the websites, even within the content of the page, in vertical banners (which are popular among advertisers because they

are permanently seen even when the reader is scrolling down) and on small buttons. There are tools to optimally choose the ads that are displayed, the most well-known being Google AdSense, but the competition is organizing itself, like Microsoft with AdCenter. Those tools are easy to use. For instance, webmasters can place Google AdSense JavaScript code on their web pages in order to allow Google's servers to show context-sensitive advertisements (Google Adwords). The history of users' browsing can be used to target ads even more; this is especially the case for social networks.

A similar issue occurs for search engines, such as Google, Yahoo! and Bing. Search engines are used to reach desired content when its address or even its existence is not known by end-users. Those search engines propose a ranked list of web sites corresponding to the keywords that have been typed. But, being commercial entities, they make money by additionally presenting advertisement links, usually at the top and/or on the right of the page. The advertisement links are selected from an auction to which an advertiser can submit a bid if she wants her ad to be displayed, because she believes the keyword search may result in a financial transaction when the user sees the ad. Search engine advertising has become an important business, with the combined revenue of the two main actors in the area, Yahoo! and Google, amounting to more than $11 billion in 2009, this business being expected also to count for about 40% of total advertising revenue.

Generalizing to all types of content/applications, when ads are thought to correspond to specific content and users (for web sites), or keywords (for search engines), the selection of the displayed ads is made through auctions. There are different ways to perform an auction. The advertisers submit bids, but the highest bids are not necessarily selected, since the selection depends also on the paying mechanism. Several principles can be applied: pay-per-view, such that the advertiser has to pay each time the ad is displayed; pay-per-click, where it pays each time the ad is clicked through; and pay-per-transaction, where it pays each time the click is transformed into a sale. There are also different possibilities regarding the amount to be paid: the first-price principle, according to which the advertiser pays its bid; the second-price principle, where it pays the bid of the second-highest bidder, etc. All this will be detailed in the book, with the respective interests and properties of the different alternatives.

Those items clearly illustrate some of the stakes for all actors in the telecommunications business. Since telecommunication networks keep evolving, the question about the most favorable economic models always comes up. Some other recent economic questions that we could have developed here, but will describe more later, are, non-exhaustively, app stores business models, grids/clouds economy, content delivery networks (CDNs) and caching, etc.

1.2 The need for modeling and analysis

There are therefore a lot of questions to be answered in terms of optimal decisions for the actors, but we wish now to highlight that mathematical modeling and analysis is an important way to avoid pitfalls that can have dramatic consequences. We aim at giving

examples of such situations and of some typical paradoxes that may occur if a model is not properly defined and analyzed.

1.2.1 The tragedy of the commons

An argument that is often advocated for changing Internet pricing is the so-called *tragedy of the commons*. The basic idea is that several individuals acting independently, rationally, and selfishly can actually deplete a shared limited resource. This may seem surprising, since at first sight it is not in the interest of any of the individuals to let this happen. This type of counter-intuitive outcome can be understood, and maybe avoided, thanks to modeling and analysis through *game theory*.

A *common* is a resource owned by no one, but to which all have access, i.e., it is in some sense a public good. The word *tragedy* follows here the definition of the philosopher Whitehead: "The essence of dramatic tragedy is not unhappiness; it resides in the solemnity of the remorseless working of things."

The tragedy of the commons has been highlighted by Hardin [129], taking as an example the case of herders sharing a parcel of land on which they can let their cows graze. In this situation, it is in each herder's interest to put as many cows as possible on the land, with the outcome that the quality of the grass decreases, because of overgrazing. Why is that the best strategy for a herder? This comes from the fact that the herder gets all the benefits from each additional cow put on the land, whereas the degradation is shared by all of the herders.

There are several other examples of the tragedy of the commons to consider, before discussing telecommunications. A second example we can mention is fishing, for which each fisher has an interest in fishing as much as possible, but this results in overfishing and the resource depletion which can be observed. Global warming is another striking example, with countries having a selfish interest in developing their industry and over-exploiting the soil, at the expense of the Earth's climate.

There are two commonly accepted solutions for solving this problem in general: (i) management of the common goods by a regulatory authority, such as a government; and (ii) privatization of the common goods. Regulation allows one to control the common resources by limiting their use through rules defined to drive the system to a socially optimal outcome. Converting the common resources into private ones (when possible; it is hardly possible when one is talking about oxygen, for example) is another way to produce incentives to use those resources efficiently. Regulation is used for fishing, for example, by limiting the amount which can be fished, with the risk of fines if the rule is not obeyed. The same happens for the amount of pollutant that can be released. Privatizing the goods allows one to avoid the outcome that the negative effect of using a resource is shared by all individuals, hence privatization provides an incentive not to overuse it. Interestingly, the Internet network, which was initially public, has been privatized, and the question is how to define the associated business model.

Though it is interesting, we need to mention that Hardin's work has been argued to be inaccurate by other researchers, claiming that the common land was effectively managed to prevent overgrazing and that self-interested individuals often find ways to cooperate.