1 Foundations

This is the chapter where the reader first approaches the world of networks. The preliminary step to take consists in providing a succinct introduction to the analysis of the signals that, either in the electrical or in the optical domain, carry data information across networks.

Also, before throwing onto the floor the first Lego bricks and the instructions to let the reader build the composite view of modern computer networks, a review of the "fabric" of these bricks is needed. A light excursus is thus provided, to gain a basic acquaintance with copper wires, fiber optics and the radio channel, each transmission medium being employed in specific networking contexts.

Network classification comes next: it is a useful exercise to gain familiarity with the main ideas and the terminology peculiar to the world of networks. This is followed by the introduction of the crucial notion of delay, as well as by a miscellanea of concepts that cross different areas, spanning from sources of network traffic to service classification, from performance metrics to quality of service.

Following a pattern that will shape the exposition throughout the entire book, a direct experience is proposed and commented at the end of the chapter, to let the reader confront the real world via a first-hand adventure.

1.1 Signals: time and frequency analysis

Indeed, our aim is to talk about networks. Before doing so, however, we have to recall the fundamentals of signal analysis, to understand what type of information we are moving around.

The more intuitive approach takes us to imagine – digital – information streams traveling among computer facilities and crossing network devices. If we think of the binary transmission of the single hexadecimal value 4F, i.e., byte 0100-1111, then the corresponding signal in the time domain could be represented as in Fig. 1.1, and could, for instance, be a voltage level at the transmitter output.

This simple description provides a way to introduce two parameters pertaining to a digital signal: the bit time T_{bit}, and the bit rate B_{r}. The first is the time it takes for a bit to be "pushed out" of the transmitting device; the second is its inverse, $B_{\mathrm{r}} = 1/T_{\mathrm{bit}}$. The higher the bit rate B_{r}, the more rapid are the signal fluctuations.

Beyond this intuitive analysis in the time domain, we are also interested in examining the properties of signals in the frequency domain.

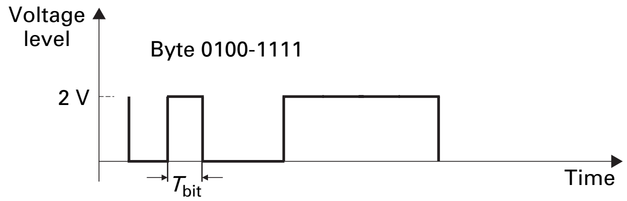


Fig. 1.1 A possible representation of byte 4F via voltage levels

To this end, we recall that, provided some conditions hold, every aperiodic, real signal $x(t)$ defined for $-\infty < t < +\infty$ can be Fourier transformed via the following relation:

$$X(f) = \int_{-\infty}^{+\infty} x(t)\mathrm{e}^{-\mathrm{j}2\pi ft}\mathrm{d}t\,, \tag{1.1}$$

$X(f)$ being the counterpart of $x(t)$ in the frequency domain.

The corresponding inverse transform is

$$x(t) = \int_{-\infty}^{+\infty} X(f)\mathrm{e}^{\mathrm{j}2\pi ft}\mathrm{d}f\,. \tag{1.2}$$

As $x(t)$ is a real function, we have that for the complex function $X(f)$ the equality

$$X(f) = X^*(-f) \tag{1.3}$$

holds, where $(\cdot)^*$ denotes the complex conjugate; equivalently,

$$\begin{cases} |X(-f)| = |X(f)| \quad \text{and} \\ \arg X(-f) = -\arg X(f) \end{cases}, \tag{1.4}$$

so that $x(t)$ can also be written as

$$\begin{aligned} x(t) &= \int_{-\infty}^{+\infty} |X(f)|\mathrm{e}^{\mathrm{j}\arg X(f)}\mathrm{e}^{\mathrm{j}2\pi ft}\mathrm{d}f \\ &= \int_{-\infty}^{+\infty} |X(f)|\left[\cos(2\pi ft + \arg X(f)) + \mathrm{j}\sin(2\pi ft + \arg X(f))\right]\mathrm{d}f \\ &= \int_{-\infty}^{+\infty} |X(f)|\cos(2\pi ft + \arg X(f))\mathrm{d}f \\ &= \int_{0}^{\infty} V(f)\cos(2\pi ft - \varphi(f))\mathrm{d}f\,, \end{aligned} \tag{1.5}$$

where

$$V(f) = 2|X(f)| \tag{1.6}$$

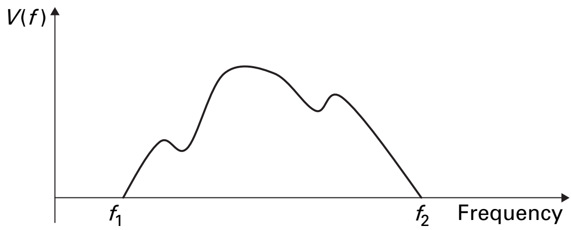


Fig. 1.2 Frequency bandwidth of a signal

and

$$\varphi(f) = -\arg X(f). \tag{1.7}$$

The $V(f)$ function provides the so-called amplitude spectrum for the original signal $x(t)$, and $\varphi = -\arg X(f)$ its phase spectrum.

More importantly for us, via $V(f)$ we introduce the concept of frequency bandwidth of $x(t)$: if $V(f)$ has the behavior that Fig. 1.2 exemplifies, then we affirm that $[f_1,f_2]$ is the frequency bandwidth B_f of signal $x(t)$, whose width is $f_2 - f_1$ hertz (Hz being the abbreviated form).

In this case Eq. (1.5) modifies into

$$x(t) = \int_{f_1}^{f_2} V(f)\cos(2\pi ft - \varphi(f))\mathrm{d}f\,, \tag{1.8}$$

which is subject to a nice physical interpretation: within the $[f_1,f_2]$ interval, $x(t)$ appears as the "sum" of infinite cosine terms, the generic term being located at frequency f, with amplitude $V(f)\mathrm{d}f$ and phase $\varphi(f)$.

Moreover, we observe that, if the function $x(at)$ is considered, $a > 0$, then its Fourier transform is given by

$$\int_{-\infty}^{+\infty} x(at)\mathrm{e}^{-\mathrm{j}2\pi ft}\mathrm{d}t = \frac{1}{a}\int_{-\infty}^{+\infty} x(z)\mathrm{e}^{-\mathrm{j}2\pi f\frac{z}{a}}\mathrm{d}z = \frac{1}{a}X\left(\frac{f}{a}\right). \tag{1.9}$$

This simple result reveals that, when $0 < a < 1$, i.e., when the signal $x(at)$ displays slower variations with respect to $x(t)$, then the low frequency components become predominant in its Fourier transform. Qualitatively, the frequency bandwidth shrinks: if it is B_f for $x(t)$, it is aB_f, $0 < a < 1$, for $x(at)$.

In a corresponding manner, we can immediately conclude that for $a > 1$ the new signal $x(at)$ exhibits more rapid fluctuations in the time domain, and that its frequency spectrum widens with respect to $x(t)$.

Let us now take advantage of this statement when examining the streams of binary digits exchanged over computer networks: recalling the notion of bit rate B_r, we can conclude that, the higher B_r, the wider the frequency bandwidth B_f of the digital signal will be. As our aim is to transmit signals bearing more and more information per unit time, we forcedly have to handle ever increasing frequency bandwidths.

Simple as it sounds, this is definitely a first conclusion to remember.

Step over now, and observe that a digital signal needs to be conveyed through a communication channel, which typically attenuates, delays and, alas, may also distort the input signal.

As a matter of fact, any communication channel is limited in bandwidth to an interval of frequencies where its effects on the conveyed signal are tolerable: typically, this constraint has to be ascribed to the physical limitations of the medium and to the components of the transmitter and the receiver.

If the bandwidth of the input signal $x(t)$, say $[f_1, f_2]$, mainly falls within such interval, what we collect at the channel output is a signal that closely resembles the input in shape. In contrast, if some of the frequency components of $x(t)$ fall outside the channel bandwidth, they will be filtered out, and this will result in a distorted output. If the communication-channel filtering is too drastic, it becomes impossible to recover the information content of the input signal out of the distorted output signal.

It is immediately concluded that, if we aim at transmitting digital signals at high bit rates, we have to employ communication channels featuring wide bandwidths.

In addition to this, we need to observe that transmission can occur either in the native frequency bandwidth of the signal, a circumstance described by the term "baseband transmission," or in a different frequency band, for which the term "bandpass transmission" is adopted. "Baseband" and "bandpass communication channels" also belong to the common glossary of this field.

So, given that a digital signal of the type shown in Fig. 1.3 exhibits a frequency bandwidth which includes frequency $f = 0$ and the nearby (low) frequencies, this signal can be sent over a baseband channel featuring an adequate bandwidth (in this case the channel is acting as a low-pass filter).

It can, however, also happen that the information-bearing signal has to be transmitted over a bandpass communication channel.

How do we handle this mismatch?

We have to require that the *transmitted* signal bandwidth falls within the communication channel bandwidth, otherwise no information delivery process will successfully take place. More explicitly, this calls for a further operation, which roughly corresponds to the translation of the digital signal spectrum to the frequency bandwidth where the channel acts as a bandpass filter: the signal has to *modulate* a suitable carrier frequency.

This is exactly what happens, for example, in radio communications.

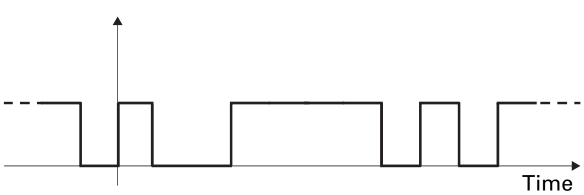


Fig. 1.3 A digital signal in the time domain

1.2 A more general notion of bandwidth

We know by now that a communication channel can support a given, maximum bit rate: it is expressed in bit/s, although Mbit/s and Gbit/s are more appropriate scales for the channels used in computer networks.

As we previously asserted, this rate directly depends on the characteristics of the transmission medium the communication channel is built upon, as well as on the hardware and software characteristics of both transmitter and receiver: confining our attention to the media, it depends on the frequency bandwidth effectively available to deliver the information signal undistorted and not significantly attenuated.

In the networking field, the transmission rate of a communication link and more generally of networks is often referred to as "bandwidth." Internet service providers (ISPs) advertise ADSL connections with 20 Mbit/s downstream bandwidth; sellers of network switches praise their apparatus as offering 1 Gbit/s bandwidth.

Purists shiver, as historically the notion of bandwidth belongs to the frequency domain, and bandwidths are measured in Hertz, not bit/s. True, without any doubt.

There is not even numerical coincidence between the maximum bit rate a channel can sustain and its frequency bandwidth: it is not true that in a communication system built on a channel whose bandwidth is 100 MHz, the maximum achievable transmission rate is exactly 100 Mbit/s (the actual ratio between rate and channel bandwidth can also be significantly greater than one).

What can settle the matter is to use the term *layer-1* (or *physical layer*) bandwidth to tag the old, familiar frequency bandwidth and *layer-2* (or *data-link layer*) bandwidth to identify the nominal transmission rate at which a link or a network carries the bits.

We are aware that the *layer-1*, *layer-2* attributes and their equivalents sound arcane at this point of the book, but in the next chapter they will gain a proper meaning. So, let us agree on their usage and simply observe that these terms inherently refer to the process of information transmission at different abstraction levels.

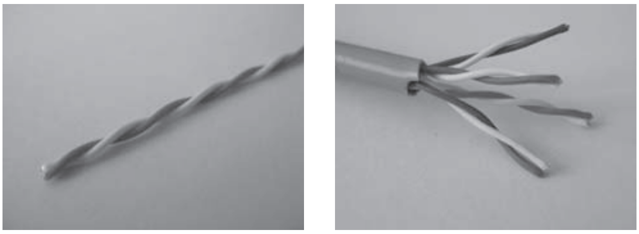
Talking about layer-1 bandwidth implies a tight attention to the media that carry the information bits; talking about layer-2 bandwidth indicates that the focus shifts onto the information itself.

The next questions we pose are: to what type of physical media do we resort to build the required communication channels? And what are their basic features? The answers come next.

1.3 Physical media

We first introduce those transmission media where the signal travels via guided propagation: among copper wires, where information bits are conveyed to their destination within the electrical domain, we mention twisted pair and coaxial cable; within the optical domain, where the information bits we aim at transferring are represented by either the presence or the absence of a pulse of light, the rails we use to move information around are optical fibers. Next, we will discuss the usage of the radio channel, a transmission medium where the signal, in the form of electromagnetic waves, propagates freely.

Cambridge University Press
978-0-521-86985-0 - Hands-On Networking: From Theory to Practice
Maria Luisa Merani, Maurizio Casoni and Walter Cerroni
Excerpt
More information



(a) (b)

Fig. 1.4 A twisted pair and a UTP cable

In telephone networks, twisted pair is by far the most widespread transmission medium: it is employed in the local loop, i.e., in that portion of the network going from the subscriber's premises to the nearest end office of the telephone company. In computer networks, and, more accurately, in local area networks (LANs), twisted pair is also very popular. Honestly though, it bears different characteristics with respect to its telephone network cousin: overall, it is of better quality, a feature not that hard to understand even at this point of the chapter, given we can imagine that the intent is to squeeze more out of a LAN wiring than out of a modest telephone line, whose wires display a bandwidth of a few hundreds of kilohertz (kHz).

What does a twisted pair look like? An example is shown in Fig. 1.4(a), which helps understand where the name comes from: two copper wires, whose diameter is of millimetrical order, are gently twisted together; for the sake of cabling, several of these pairs are then grouped within a same cable, as Fig.1.4(b) illustrates: note that each wire is insulated and that there is a protective sheath that wraps all the four pairs of this example. Having four pairs is not an accident: we will see it is exactly the number adopted by LAN cables. The term for this solution is unshielded twisted pair (UTP) cable: it is flexible and therefore easy to install, and also cheap. We will meet it again in Chapter 3, where we will discuss in detail its usage and its limits. For now we simply observe that good quality twisted pair displays a frequency bandwidth of a few megahertz; its usage is, however, confined to relatively short distances, a few hundred meters.

Coaxial cable has a different role with respect to twisted pair: no longer used in LANs, broadband coaxial cable originally survived in the separate field of analog – not digital – broadcasting of television channels. It then gained renewed interest for computer networks when cable TV companies began providing Internet connectivity: their subscribers, equipped with appropriate devices called cable modems, can transmit data on broadband coaxial cable at considerable rates.

A coaxial cable is shown in Fig. 1.5. Here too we find two copper conductors: the inner is a cylindrical core, surrounded by a dielectric insulator; the outer is a metallic shield, protected by a plastic jacket. Coaxial cable can be moderately bent and twisted without altering its main characteristics, the reason being that the electromagnetic field is confined between the inner conductor and the shield. In terms of quality, broadband coaxial cable does not shine, but it can be employed for frequencies up to several hundreds of megahertz and for distances up to one hundred kilometers.

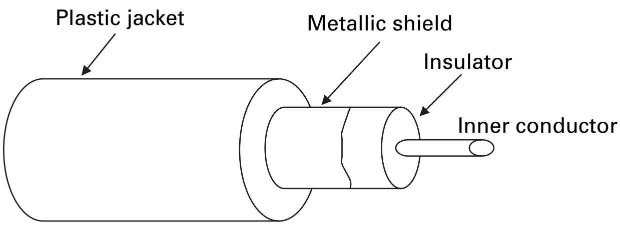


Fig. 1.5 Example of a coaxial cable

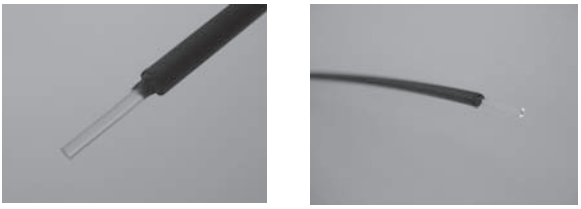


Fig. 1.6 An optical fiber

Among the guided media, optical fiber is by far the best performing choice: its geometry shows the presence of a glass core, surrounded by a glass cladding with a different propagation index and coated by a few protecting layers: some illustrative pictures of a fiber appear in Fig. 1.6. Transmission over fiber occurs in the infrared region, in three distinct wavelength bands, centered at 850 nm, 1300 nm and 1550 nm. Two main fiber categories exist: single-mode and multi-mode fibers. In the former, light pulses propagate within the core in single mode, i.e., with no reflections against the cladding; single-mode fibers commonly display a core diameter of 8 to 10 μm, and operate in the first two bands. Multi-mode fibers, on the contrary, owe their name to the multitude of paths that the light takes along the core. They operate in the second and third wavelength bands.

The attenuation that the fiber introduces is extremely low; its bandwidth is huge, as compared to the competing media, and allows very high data rates, of the order of tens of Gbit/s. Photonic devices for signal generation and detection are currently highly reliable. Moreover, fiber is lightweight, of small size, immune to electromagnetic interference, hard to tap and does not corrode. No surprise that it is the preferred transmission medium in several networking contexts!

As regards the radio channel, here the transmission medium is the ether: the signal bearing the desired information propagates freely, no sharp boundaries delimit the area where it can be heard. This is the strength of communications that rely upon such physical medium: users of a wireless LAN can readily move their laptop, free from the hassle of wiring; subscribers of UMTS mobile radio networks can experience the “always on, always connected” promise, having Internet access in places not reached by cables (we should honestly add “almost” every time and every place).

The radio frequencies employed by most wireless LANs and mobile radio systems fall within the ultra-high-frequency range (UHF), spanning from 300 MHz to 3 GHz.

The bad news is that this portion of radio spectrum is fairly crowded and its usage often requires a license; moreover, the UHF wireless channel is a tough medium, where transmissions are impaired by strong attenuation, multipath propagation, interference from simultaneous transmissions, all statistical in nature. To make things even worse, wireless users move around in a hardly predictable manner, often at high speed, so that the channel characteristics vary both in frequency and time. It is definitely not that trivial to guarantee both high transmission rates and excellent quality on a radio channel!

Finally, it is not always an advantage potentially to have anyone hear anyone else's communication: careful mechanisms have to be put in place to secure the privacy and integrity of the messages exchanged in wireless settings.

1.4 Network classification

1.4.1 The obvious starting example

Having declared what are the transmission media that allow to build the physical infrastructure of networks, let us jump to the most popular and perhaps misused example of networks: the Internet. We define it as a "network of networks" sharing one attribute: all devices over the Internet speak a common language, known as the TCP/IP protocol suite. Its discussion deserves several pages, and we have chosen the next chapter to illustrate it. For now it is important to concentrate on the salient features of the Internet infrastructure, because its heterogeneity and constructive mess already contains all the elements we need to classify networks.

Fig. 1.7 provides a first, necessarily simplified view: the intent is to portray the myriad of access networks constituting the edges of the Internet, interconnected via a tiered

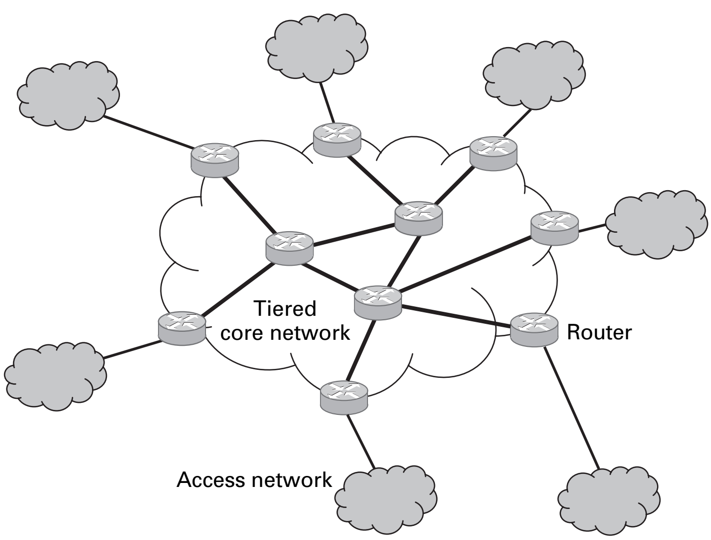


Fig. 1.7 A simplified view of the Internet

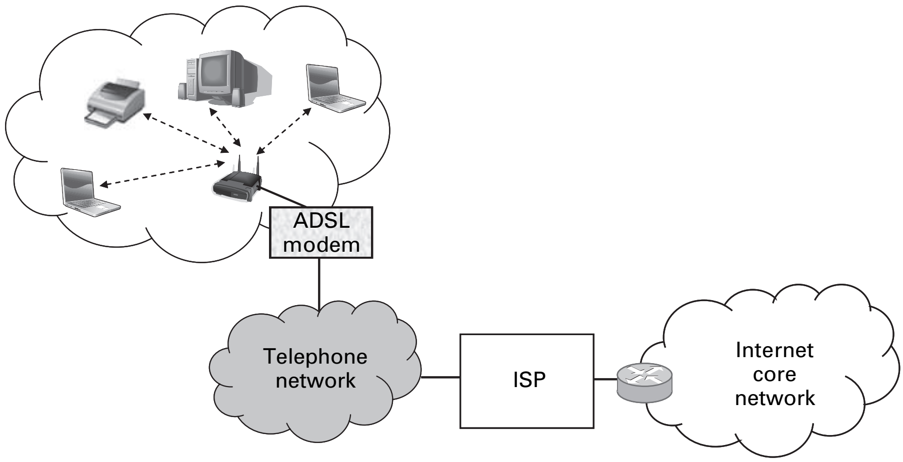


Fig. 1.8 A residential wireless LAN connected to the Internet

core network. Network devices called routers provide the desired connectivity between different domains.

Fig. 1.8 sketches a more familiar portion of the whole picture: in this instance a residential customer is equipped with a small wireless LAN, connected to the Internet through its telephone line and an ADSL modem: after a few hops covered via the telephone network, we meet the router of the ISP guaranteeing the user Internet access.

Routers: what are their main tasks? For the time being, it suffices to know that these nodes are the busy Internet servants: their role is to connect its different constituent networks and to make the information travel from source to destination, properly sized in chunks generically termed *packets*.

1.4.2 Circuit-switched versus packet-switched networks

The word packet immediately gives us a way to distinguish between two classes of networks, which radically differ in the way resources are shared among users: the alternatives are *circuit-switched* and *packet-switched* networks. The Internet is a packet-switched network.

Let us comment on circuit-switching first, according to the same chronological order that the introduction and deployment of these technologies followed. In circuit-switched networks, before any communication can take place, it is mandatory to determine a path between source and destination, and statically allocate resources for that communication along the entire path, i.e., build the circuit. Only after the resource assignment procedure has been successfully completed can the communication take place. Resources are released – the circuit is torn down – as soon as the communication process is over. If for any reason a resource is missing along the path, then the communication cannot occur.

In the past, the telephone network, often referred to as the public switched telephone network (PSTN), was the par-excellence example of a circuit-switched network. To set up a circuit in it means to dedicate resources exclusively to the call within the network switching centers, as well as to assign the call some capacity on the traversed telephone trunks. This hides a significant and time-consuming signaling effort: the circuit set-up request has to propagate along the entire path to the called party, and once successful, a call-accept signal has to find its way back to the calling party (as a result, a ringing tone is sent to the calling telephone, while the called telephone rings).

Currently, the way the PSTN operates is much more articulated: the network backbone of telephone carrier operators also resorts to packet-switching to carry voice calls, but we will not deepen this point.

The approach is completely different in packet-switched networks, where from the very beginning the end users willing to communicate were only computers: the name recalls that the information crosses the networks split in packets. At this point of the chapter, we simply state that the packet format encompasses a field for data bits and one or more fields bearing control bits.

Two different categories exist within the packet-switched family: virtual circuit networks and datagram networks. The Internet core is a datagram network.

In virtual circuit networks, a "virtual circuit" (we could read a "route") is built between source and destination before packets of the information stream can start traveling along that circuit. This concise description might suggest a similarity to circuit switching: it is not so, there is a strong conceptual novelty here, that we now explain. To set up a virtual circuit means to draw a route across the network that all packets of the virtual circuit will follow, but it does not imply that resources are deterministically reserved in advance. Instead, when packets belonging to a virtual circuit travel along the network, they *dynamically* share the capacity of the traversed links with other packets of different virtual circuits; when they arrive at the routers, where they are typically stored and then forwarded to the desired output link, they *dynamically* share the router resources (the power of the processor the router is equipped with, the input and output buffers of the router).

This *statistical multiplexing* of network resources, as opposed to the *deterministic multiplexing* that circuit-switched networks implement, has an immediate, tangible effect: packets might queue up at routers, incurring random delays, which depend on the amount of load each router has to handle. This is something that cannot happen in circuit-switched networks, where no storing at network nodes is required.

The concept behind the virtual circuit approach is represented in Fig. 1.9: in this simple example, packets belonging to three distinct virtual circuits share resources along various paths pinned through the network.

Finally, datagram technology. This is the simplest and most anarchic solution: every packet (a *datagram*, as it was originally called at the beginning of Internet history) of an information flow is individually routed through the network. No static path is built in advance, so that in principle every packet of the stream can follow a different road toward its final destination. To make this possible, the control field of the packet has to bear the ultimate destination address: each router crossed by the packet reads this address to decide the most appropriate output link to forward the datagram. The packet, therefore, proceeds hop-by-hop toward its final recipient, every