local area networks and intranets

3.1 Introduction

As we saw in the last chapter, when a person is at home access to the Internet is through an Internet service provider (ISP). The ISP is located remotely from the subscribers and access to it is obtained using a PSTN/ISDN access network and either a low bit rate modem or, in many instances, a broadband modem. In the case of low bit rate modems, the cost of a session is determined by normal telephone charges; that is, by the duration of the session and the distance from the home to the ISP site. In the case of broadband modems, normally the connection to the ISP is permanent and the subscriber then pays a regular monthly subscription. This mode of access is also used in small offices and businesses.

In the case of a person in a large business/enterprise, in addition to accessing the Internet, he/she needs to communicate with other members of the business/enterprise. In the case of a large enterprise, this may involve multiple sites that are physically distributed around a single country or, with the largest enterprises/corporations, around the world.

To provide a telephony service at a site, a private branch exchange (PBX) that is similar to a (small) LE/EO is used. To provide a data communications service, a (private) site network is used which, because of its physical scope, is called a local area data network or simply a **local area network** or **LAN**. Examples are a campus, a hospital complex and a large office building.

For a large multi-site enterprise, normally the PBX at each site are interconnected together using leased lines so extending the telephony service to the whole enterprise. Similarly, the site LANs are also interconnected using portions of the bandwidth of the same leased lines. The total network is then known as an **enterprise network** and these provide each member of the enterprise with a communication facility for both telephony/voice and data.

In the case of telephony, each PBX, in addition to the leased-line connections to the other sites, also has a connection to a LE/EO owned by a public telecommunications provider. In this way, a member of the enterprise can also speak to clients/customers, etc., in other businesses and enterprises. Similarly, in the case of data communications, the enterprise network has a connection to the Internet/ISP through an access gateway so enabling the staff at each site to send e-mail, access the Web, and so on. Also for other organizations to access Web sites maintained by the enterprise. In the past, many large enterprises/corporations used proprietary protocols in their networks. Increasingly, however, these have been replaced with the Internet protocols. For this reason, therefore, large multi-site enterprise networks that use the Internet protocols are called **intranets**.

As we have just indicated, LANs are used to interconnect distributed communities of end systems – referred to as **stations** in the context of LANs – including multimedia PCs, workstations, servers, and so on. Typically, these are physically distributed around an office, a single building, or a localized group of buildings, all of which belong to a single company/enterprise. The international standards body responsible for LANs is the IEEE and all the different LAN types are part of the IEEE 802 series. They operate using a shared, high bit rate, transmission network to which all stations are attached and the information frames relating to all sessions transmitted. To ensure the transmission bandwidth is shared fairly between all of the attached stations, a number of different **medium access control** (**MAC**) methods are used. These include **carrier-sense multiple-access with collision detection** (**CSMA/CD**) and **control token**, both of which have a defined maximum number of attached stations and length of transmission medium associated with them. As we shall see, in practice the maximum distance is relatively small and hence most LANs of this type comprise multiple (LAN) **segments** that are interconnected together using either electrical **repeaters** or devices known as **bridges** and high bit rate switches.

An example of a LAN that operates using the CSMA/CD MAC method is **Ethernet** and one that operates using a control token is **Token ring**. Ethernet LANs are defined in IEEE802.3 and Token ring LANs in IEEE802.5. However, IEEE802.3 LANs are now the dominant LAN type, primarily because the

CSMA/CD MAC method can operate at the much higher bit rates that have become possible owing to the advances in transmission technology over the past few years.

All the IEEE802.3 LANs utilize fixed wiring such as coaxial cable, twisted-pair wire and, more recently, optical fibre as the transmission medium. However, the advent of sophisticated pocket PCs and notebook computers means that, in addition to operating solely as portable devices, they often need to communicate with computers that are attached to a wired LAN. To facilitate this, a number of wireless LANs have been developed. The one that is compatible with IEEE802.3 LANs is defined in IEEE802.11. However, since wireless LANs use radio as the transmission medium, we shall defer studying them until the next chapter where we describe the various types of wireless access networks. In this chapter, therefore, we shall limit our discussion to the operation of the various IEEE802.3 LANs and the methods that are used to create both small and large single-site LANs and multi-site enterprise networks and intranets.

Ethernet/IEEE802.3 networks are used extensively in technical and office environments. As we shall see, Ethernet has gone through many phases of development since its first introduction and, in general, the same basic MAC method is still used. All frame transmissions between all the stations that are attached to the same LAN segment take place over a shared high bit rate transmission cable. The CSMA/CD MAC method is then used to share the use of the transmission bandwidth of the cable in an equitable way. We describe this in some detail in Section 3.2.

More recently, higher bit rate versions of the older LAN types – now known as **legacy LANs** – have become available. To obtain the higher network throughputs that are required with multimedia applications, the central **hubs** associated with the earlier LANs have been upgraded to operate at much higher bit rates. We explain the different LAN interconnection technologies in Section 3.3. Also, as we shall explain, the older hubs operate in a half-duplex mode and support only a single frame transfer at a time. Hence the newer hubs operate in a duplex mode and allow the frames relating to multiple sessions to be transmitted concurrently. Examples include **fast Ethernet** hubs, **Ethernet switching** hubs, and **Gigabit Ethernet**, all of which we describe in Section 3.4. In addition, to improve security and obtain higher throughput, **virtual LANs** have been defined and these are described in Section 3.5.

In terms of the link layer protocol associated with LANs, the various LAN types all use a standard link control sublayer and there is a different MAC sublayer for each of the LAN types. We describe the structure and the user services offered by each sublayer in Section 3.6. Finally, we describe the different interconnection methods that are used to create large enterprise-wide networks in Section 3.7.

3.2 Ethernet/IEEE802.3

Ethernet/IEEE803.2 networks are used extensively in technical and office environments. Also, as we indicated earlier, Ethernet has gone through many phases of development since its first introduction but, in general, the same basic MAC method is still used. All frame transmissions between all the stations that are attached to the same LAN segment take place over a shared high bit rate transmission cable. The CSMA/CD MAC method is then used to share the use of the transmission bandwidth of the cable in an equitable way.

3.2.1 CSMA/CD

Since all the stations are attached directly to the same cable/bus, it is said to operate in a **multiple access** (**MA**) **mode**. To transmit a block of data, the source station first encapsulates the data in a frame with the address of the destination station and its own address in the frame header and an FCS field at the tail of the frame. The bus operates in the **broadcast mode**, which means that every frame transmitted is received by all the other stations that are attached to the bus. Hence as each of the other stations receives the frame, it first checks that the frame is free of errors using the FCS and, if it is, it compares the destination address in the header with its own address. If they are different, the station simply discards the frame; if they are the same, the frame contents are passed up to the link control sublayer for processing together with the address of the source station.

With this mode of operation, two (or more) stations may attempt to transmit a frame over the bus at the same time. Because of the broadcast mode, this will result in the contents of the two (or more) frames being corrupted and a **collision** is said to have occurred. Hence in order to reduce the possibility of a collision, prior to sending a frame the source station first determines whether a signal/frame is currently being transmitted on the bus. If a signal – known as the **carrier** – is **sensed** (**CS**), the station defers its own transmission until the current frame transmission is complete and only then does it attempt to send its own frame. Even so, in the event of two (or more) stations waiting to send a frame, both will start to transmit their frame simultaneously on detecting that the transmission of the current frame is complete. When this happens, however, it is necessary for the two (or more) stations involved, to detect a collision has occurred before each has finished transmitting its own frame. In practice, because of the possibly large signal propagation delay of the bus and the high transmission bit rate used (10 Mbps), this is not as straightforward as it might seem.

A station detects that a collision has occurred by simultaneously monitoring the signal that is present on the cable all the time it is transmitting a frame. Then, if the transmitted and monitored signals are different, a

collision is assumed to have occurred – **collision detected** (**CD**). As we show in Figure 3.1, however, a station can experience a collision not just at the start of a frame but after it has transmitted a number of bits. The worst-case time delay – and hence maximum number of bits that have been transmitted – before detecting that a collision has taken place is known as the **collision window** and occurs when the two colliding stations are attached to opposite extremities of the bus, as we show in the figure.

In the figure, station *A* has determined that no transmission is in progress and hence starts to transmit a frame – part (i). As we explained in Chapter 1, in the section on signal propagation delay, irrespective of the bit rate being used, the first bit of the frame will take a small but finite time to propagate over the transmission medium determined by the length of the cable, l, and the signal propagation velocity, v. The maximum length of cable is set at 2.5 km. Hence, assuming a v of 2×10^8 m s^{-1}, the worst-case signal propagation delay time, T_p, going from one end of the cable to the other, is given by:

$$T_p = l/v = 2.5 \times 10^3 / 2 \times 10^8 = 12.5\,\mu s$$

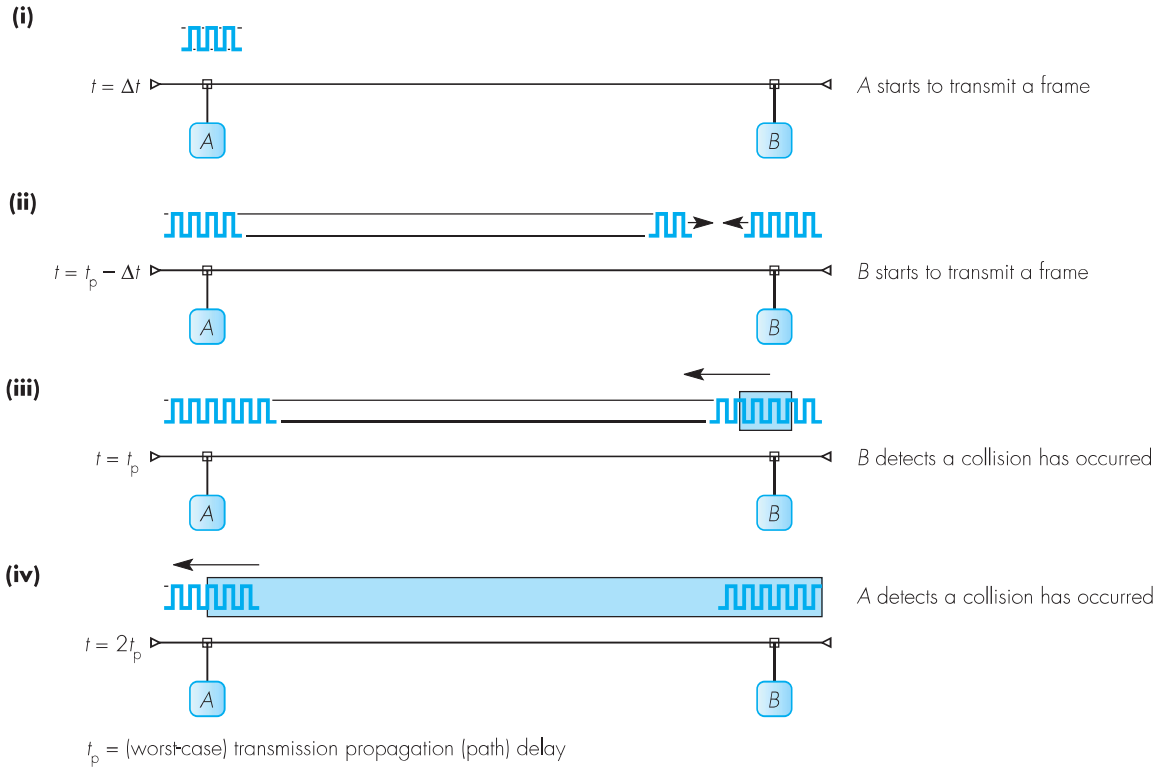


Figure 3.1 CSMA/CD worst-case collision detection.

Now assume that, just prior to the first bit of the frame arriving at its interface, station *B* determines the transmission medium is free and starts to transmit a frame – part (ii).

As we show, after *B* has transmitted just a few bits, the two signals collide – part (iii) – and the collision signal then continues to propagate back to station *A* – part (iv). Hence the worst-case time before station *A* detects that a collision has occurred, $2T_p$, is 25 μs. In addition, as we shall expand upon later, in order to transmit the signal over this length of cable, the cable is made up of five 500 m *segments*, all interconnected together by means of four devices called **repeaters**. Each repeater introduces a delay of a few microseconds in order to synchronize to each new frame. Hence the total worst-case time is set at 50 μs or, assuming a bit rate of 10 Mbps, after *A* has transmitted:

$$10 \times 10^6 \times 50 \times 10^{-6} = 500 \text{ bits}$$

A safety margin of 12 bits is then added to this and the minimum frame size is set at 512 bits or 64 bytes/octets which takes 51.2 μs to transmit. This is called the **slot time** and ensures that station *A* will have detected a collision before it has transmitted its smallest frame. Also, to ensure that the collision signal persists for sufficient time for it to be detected by *A*, on detecting the collision, *B* continues to send a random bit pattern for a short period. This is known as the **jam sequence** and is equal to 32 bits.

After detecting a collision, the two (or more) stations involved then wait for a further random time interval before trying to retransmit their corrupted frames. As we shall explain later, the maximum frame size including a four-byte CRC is set at 1518 bytes and hence a collision will occur if two (or more) stations create a frame to send during the time another station is currently transmitting a maximum sized frame. This is equal to a time interval of:

$$1518 \times 8/10 \times 10^6 = 1.2144 \text{ ms}$$

Clearly, the probability of this occurring increases with the level of traffic (number of frames) being generated and the maximum throughput of the LAN occurs when this limit is reached. Hence if a second collision should occur when a station is trying to send a frame, this is taken as a sign that the cable is currently overloaded. To avoid further loading the cable, therefore, the time interval between trying to retransmit a frame is increased exponentially after each new attempt is made using a process known as **truncated binary exponential backoff**.

When a collision first occurs, each station waits for a random time of either 0 or 1 slot times before attempting to retransmit its frame. Clearly, if both stations select the same number then a second collision will occur the probability of which is 0.5. In the event of a second collision occurring, the degree of randomness is increased by each station waiting for one of 0, 1, 2 or 3 slot times. Hence the probability of a collision occurring is now halved to

0.25. In the event of a third collision, each station waits for one of 0, 1, 2, 3, 4, 5, 6, 7 or 8 slot times, so again halving the probability of a collision occurring to 0.125.

As we can deduce from this, after n successive collisions a random number of slot times between 0 and $2n - 1$ is selected. This continues for up to ten collisions after which the random number of slot times selected remains fixed at between 0 and 1023. The number ten is known as the backoff limit. As we can deduce from the operation of the algorithm, it ensures that only a small delay is incurred when the cable/LAN is lightly loaded and the access time increases in a controlled way as the load increases. As we explain later in Section 3.6.2, in the event of a set number of attempts to send a frame failing – called the **attempt limit** – then the sending LLC sublayer is informed that the transmission of the frame has failed owing to excessive collisions.

Finally, it should be noted that the CSMA/CD access method is only concerned with sharing the physical transmission medium in an equitable way. It does not guarantee that a frame that does not incur a collision arrives at its intended destination free of errors. This is the role of the FCS at the tail of the frame and, if bit errors are detected by the FCS, the frame is discarded.

3.2.2 Wiring configurations

There are a number of different types of cable that have been used with Ethernet. These are listed in historical order:

- **10Base5**: thick-wire (0.5 inch diameter) coaxial cable with a maximum segment length of 500 m.
- **10Base2**: thin-wire (0.25 inch diameter) coaxial cable with a maximum segment length of 185 m;
- **10BaseT**: hub (star) topology with twisted-pair drop cables of up to 100 m;
- **10BaseF**: hub (star) topology with optical fibre drop cables of up to 2 km.

Although different types of cable are used, they all operate using the same CSMA/CD MAC method.

At the time the first Ethernet installations were carried out, the only transmission medium available that could operate at 10 Mbps was coaxial cable. Initially, thick-wire coaxial cable was installed since this can be used in relatively long lengths of up to 500 m before the transmitted/broadcast signal needs to be repeated. As we explained in Section 1.3.1, this involves the attenuated signal received at the extremity of the cable segment being amplified and restored to its original form before it is retransmitted – repeated – out onto the next cable segment. Up to five cable segments – and hence four repeaters – can be used in this way. Hence the maximum length of cable the signal propagates is 2.5 km plus 4 repeaters, which is the origin of the slot-time figure used in the standard.

The disadvantage of thick-wire coax is that it is relatively difficult to bend and hence install. To overcome this, thin-wire coax was used but, because of the increased (electrical) resistance associated with it, the maximum length of cable for each segment is reduced to 185 m.

More recently, as we explained in Section 1.3.2, with the arrival of inexpensive adaptive crosstalk canceller circuits to overcome near-end crosstalk (NEXT), it is possible to obtain bit rates of tens of Mbps over unshielded twisted-pair cable of up to 100 m in length. Also, it was found that, in a vast majority of offices, the maximum length of cable used for telephony to reach each desktop from the wiring closet was less than 100 m. Hence unshielded twisted-pair (UTP) cable – as used for telephony – became the standard for use with Ethernet. The configuration used for each segment is shown in Figure 3.2(a).

Since the cable forms a physical bus, both thick and thin wire coaxial cable installations involve the cable passing near to each attached station. As we can see in the figure, however, with twisted-pair cable a star configuration is used with the hub located in the wiring closet and each station connected to it by means of twisted-pair drop cables. Initially, category three (CAT3) UTP cable was used as for telephony. Each cable contains four separate twisted-pairs. In the case of Ethernet, just two pairs are used: one pair for transmissions from the station to the hub and the second pair for transmissions in the reverse direction. More recently, higher bit rate versions of the basic Ethernet have been introduced that use the higher-quality category five (CAT5) unshielded twisted pair (UTP) cable. Also, as we shall see later in Section 3.3, optical fibre is now used extensively to interconnect hubs together since, in many instances, this is over larger distances than those possible with twisted pair cables.

To emulate the broadcast mode of working associated with CSMA/CD, as we show in Figure 3.2(b), the repeater electronics within the hub repeats and broadcasts out the signal received from each of the input pairs onto all of the other output pairs. Hence the signal output by any of the stations is received by all the other stations and, as a result, the carrier sense function simply involves the MAC unit within each station determining whether a signal is currently being received on its input pair. Similarly, the collision detection function involves the station determining if a signal arrives on its input pair while it is transmitting a frame on the output pair.

Because of their mode of operation, this type of hub is called a **repeater hub** and typical numbers of attached stations – and hence sockets – are from 8 through to 16. Above this number multiple hubs are stacked together and are connected by repeaters or, as we shall explain in Section 3.3, bridges or switches. In the case of repeaters, the maximum length of cable between any two stations – including the 100 m drop cables – must not exceed 1.5 km. To achieve this coverage/distance, however, normally it is necessary to use a central hub to which each twisted-pair hub is connected by means of optical fibre cables.

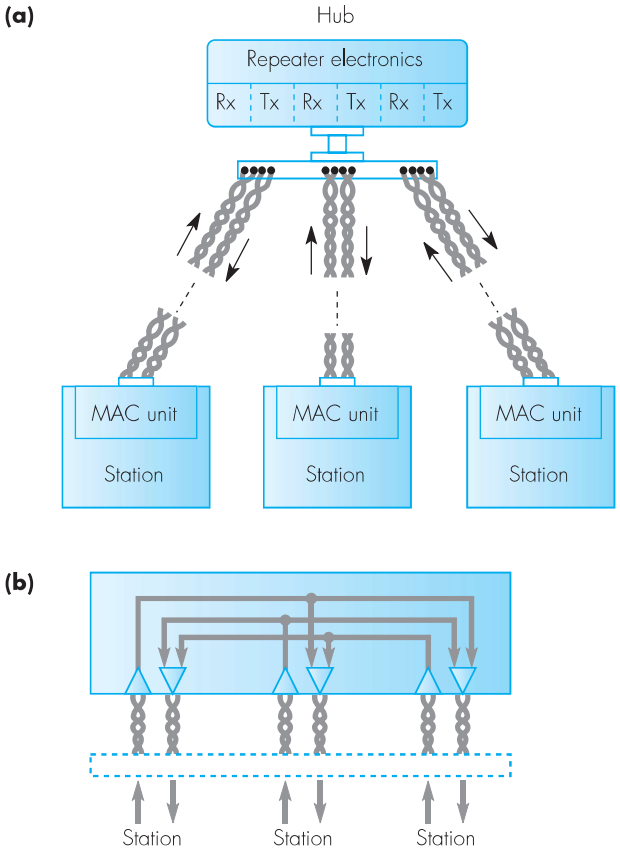


Figure 3.2 Hub configuration principles: (a) topology; (b) repeater schematic.

3.2.3 Frame format and operational parameters

The format of a frame and the operational parameters of a CSMA/CD network are shown in Figure 3.3. The *preamble* field is sent at the head of all frames. Its function is to allow the receiving electronics in each MAC unit and repeater to achieve bit synchronization before the actual frame contents are received. The preamble pattern is a sequence of seven bytes, each equal to the binary pattern 10101010. All frames are transmitted on the cable using Manchester encoding. Hence, as we explained in Section 1.3.3, the preamble results in a periodic waveform being received by the receiver electronics in each station, which acts as a reference clock. The *start-of-frame delimiter* (*SFD*) is the single byte 10101011 that immediately follows the preamble and signals the start of a valid frame to the receiver.

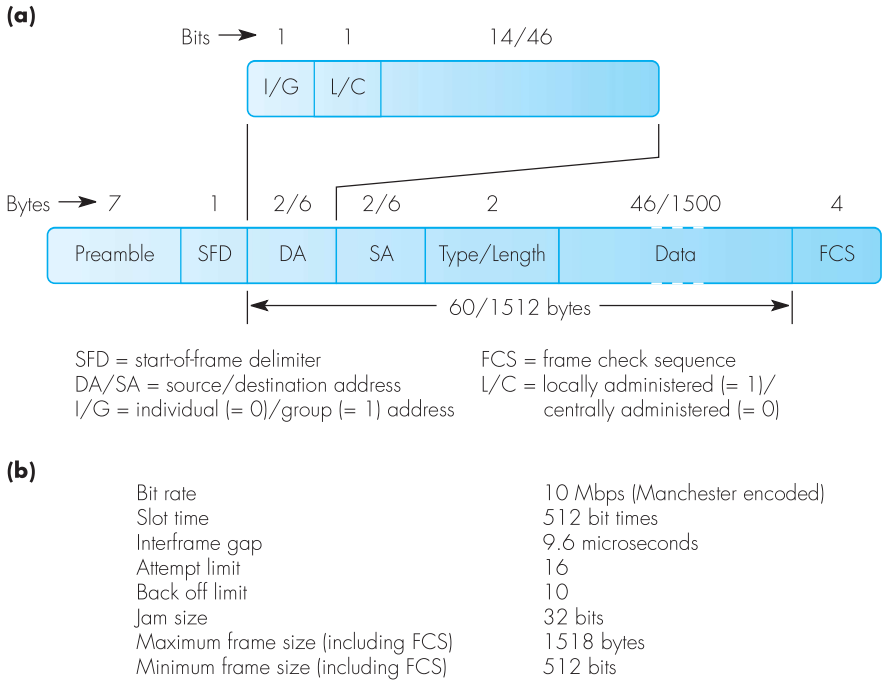


Figure 3.3 Ethernet/IEEE802.3 characteristics: (a) frame format; (b) operational parameters.

The *destination* and *source addresses* – also known as **MAC addresses** because they are used by the MAC sublayer – specify the identity of the hardware interface of both the intended destination station(s) and the originating station, respectively. Each address field can be either 16 or 48 bits, but for any particular LAN installation the size must be the same for all stations. The first bit in the destination address field specifies whether the address is an **individual address** (= **0**) or a **group address** (= **1**). If an individual address is specified, the transmitted frame is intended for a single destination. If a group address is specified, the frame is intended either for a logically related group of stations (group address) or for all other stations connected to the network (**broadcast** or **global address**). In the latter case, the address field is set to all binary 1s and, for a group address, the address specifies a previously agreed group of stations. The type of grouping is specified in the second bit and can be locally administered (=1) or centrally administered (=0). Group addresses are used for multicasting and the MAC unit/circuit associated with each station in the multicast group is then programmed to read all frames with this group address at its head.

With the original Ethernet standard, the two-byte *type* field immediately follows the address fields and indicates the network layer protocol that cre-

ated the information in the data field. With the more recent IEEE802.3 format, the next two bytes are used as a *length indicator* which indicates the number of bytes in the data field. If this value is less than the minimum number required for a valid frame (minimum frame size), a sequence of bytes is added, known as **padding**. The maximum size of the data field – normally referred to as the **maximum transmission unit** (**MTU**) – is 1500 bytes. Hence to enable the same field to act as a type field, any value greater than 1500 is interpreted as indicating a frame type. We shall illustrate the use of the type field in later sections. Finally, the *frame check sequence* (*FCS*) field contains a four-byte (32-bit) CRC value that is used for error detection. Note that with the original Ethernet standard, the end of a frame is detected when signal transitions end.

3.2.4 Frame transmission and reception

The frame transmission sequence is summarized in Figure 3.4(a). When a frame is to be transmitted, the frame contents are first encapsulated by the MAC unit into the format shown in Figure 3.3(a). To avoid contention with other transmissions on the medium, the MAC unit first monitors the carrier sense signal and, if necessary, defers to any passing frame. After a short additional delay (known as the **interframe gap**) to allow the passing frame to be received and processed by the addressed station(s), transmission of the frame is initiated.

As the bitstream is transmitted, the transmitter simultaneously monitors the received signal to detect whether a collision has occurred. Assuming a collision has not been detected, the complete frame is transmitted and, after the FCS field has been sent, the MAC unit awaits the arrival of a new frame, either from the cable or from the link control layer within the station. If a collision is detected, the transmitter immediately turns on the collision detect signal and enforces the collision by transmitting the jam sequence to ensure that the collision is detected by all other stations involved in the collision. After the jam sequence has been sent, the MAC unit terminates the transmission of the frame and schedules a retransmission attempt after a short randomly computed interval.

Figure 3.4(b) summarizes the frame reception sequence. The MAC unit first detects the presence of an incoming signal and switches on the carrier sense signal to inhibit any new transmissions from this station. The incoming preamble is used to achieve bit synchronization and, when the start-of-frame delimiter has been detected, with an IEEE802.3 LAN, the length indicator is read and used to determine the number of bytes that follow. The frame contents including the destination and source addresses are then received and loaded into a frame buffer to await further processing. The received FCS field is first compared with the computed FCS and, if they are equal, the frame content is further checked to ensure it contains an integral number of bytes and that it is neither too short nor too long. If any of these checks fail then the frame is discarded. If all checks pass, then the destination address is read

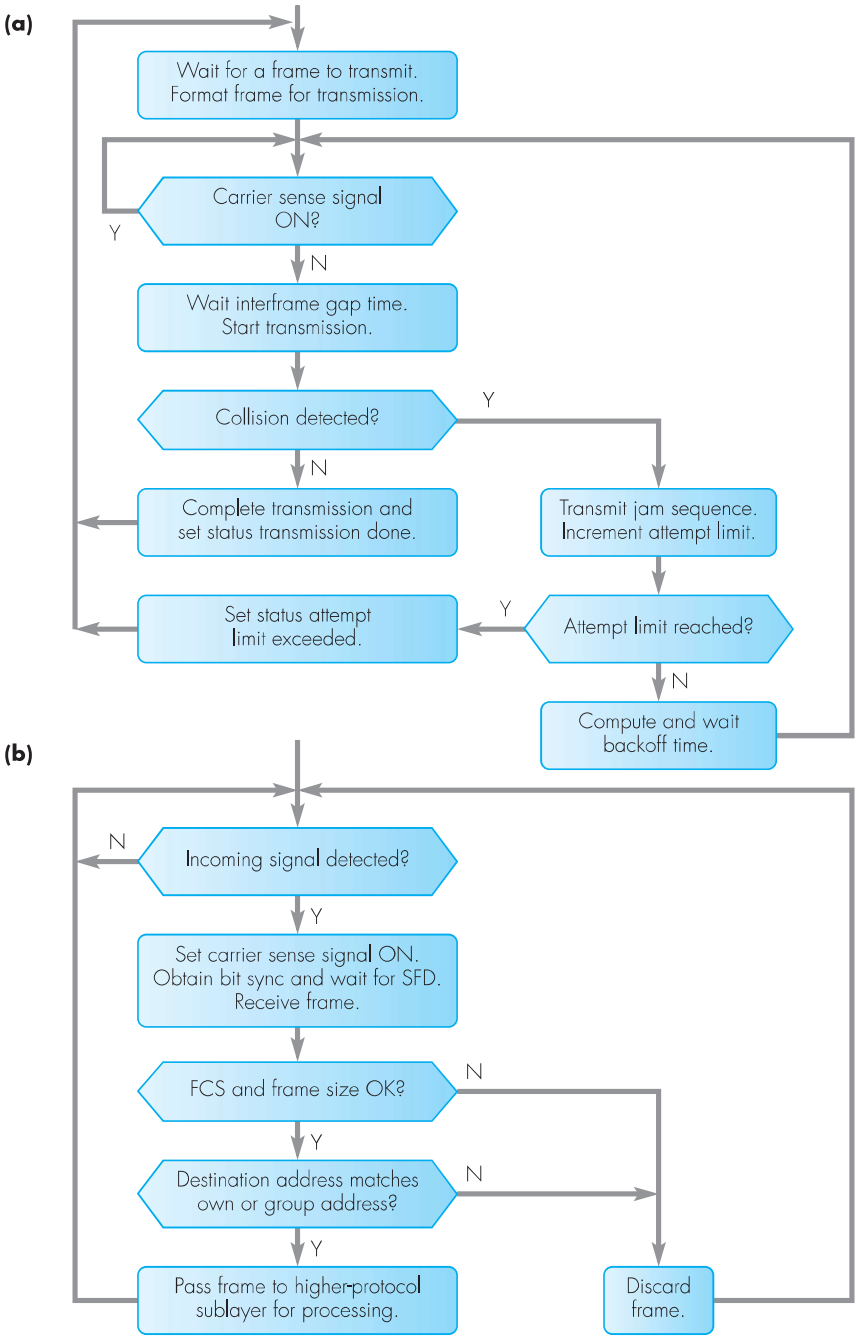


Figure 3.4 CSMA/CD MAC sublayer operation: (a) transmit; (b) receive.

from the head of the frame and, if the frame is intended for this station – that is, the address of the station is the same as that in the frame or, if it is a group address, the station is a member of the specified group – the frame contents are passed to the link control layer for processing.

3.3 LAN interconnection technologies

In general, within a small business/site, the basic communications requirement – in addition to telephony – is to enable a number of users, each with a desktop PC/workstation, to access a server computer that is used as, say, a print server and an e-mail server for the site. To achieve the latter function, the server is connected to the Internet through an Internet service provider that also provides access to the Web. As we explained in Section 2.6 and illustrate in Figure 3.5, the connection to the ISP can be by means of either a broadband modem or, if higher throughput is required, a primary rate leased line.

With a larger business that requires multiple hubs, to enable all users to access the site server(s), the hubs must be interconnected together. This can be done in three alternative ways:

- repeater hubs,
- bridging hubs,
- switching hubs.

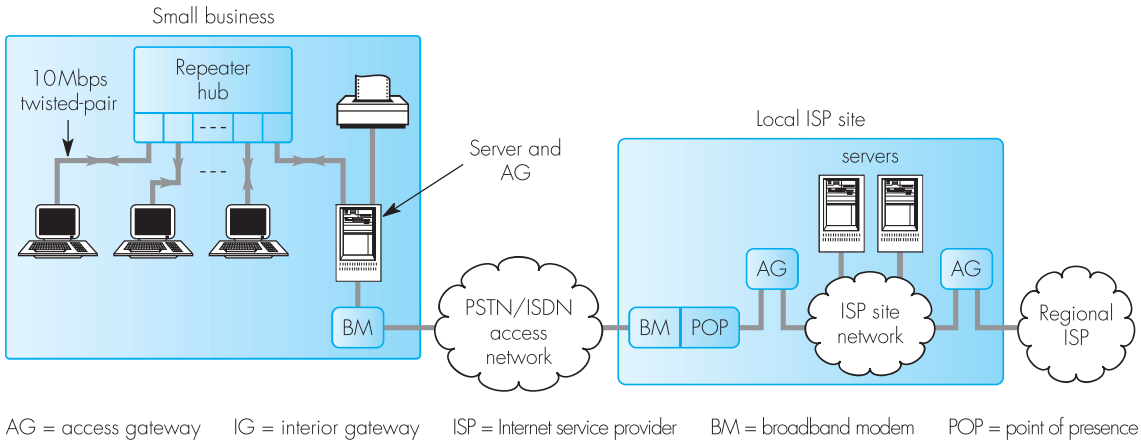


Figure 3.5 Access to the Internet for a small business.

3.3.1 Repeater hubs

A repeater hub, in addition to repeating each frame received at each of its input ports out onto all of its other ports, also repeats each frame it receives out to a higher level repeater if one is present. Hence, as we can deduce from the example network topology shown in Figure 3.6, each frame transmitted by a member of any of the workgroups is repeated to all the other segments – and hence stations – in the total network. This means, therefore, that in terms of the available bandwidth, the network behaves like a single LAN segment. In many instances, however, there is no necessity for the frames generated within a workgroup to be transmitted beyond the hub/segment on which they are attached. Bridging hubs were introduced therefore to limit the forwarding of frames to those that are intended for a different segment/workgroup.

3.3.2 Bridging hubs

The function of a bridging hub is similar to a repeater hub in that it is used for interconnecting a set of repeater hubs. However, when a bridging hub is used, all frames received from a lower-level hub are first buffered (stored) and error checked before they are repeated (forwarded). Moreover, only those frames that are free of errors and addressed to a station that is attached to a different repeater hub/segment from the one on which the frame was received are forwarded. Consequently, all frames that are addressed to another member of the same workgroup are not forwarded and hence do

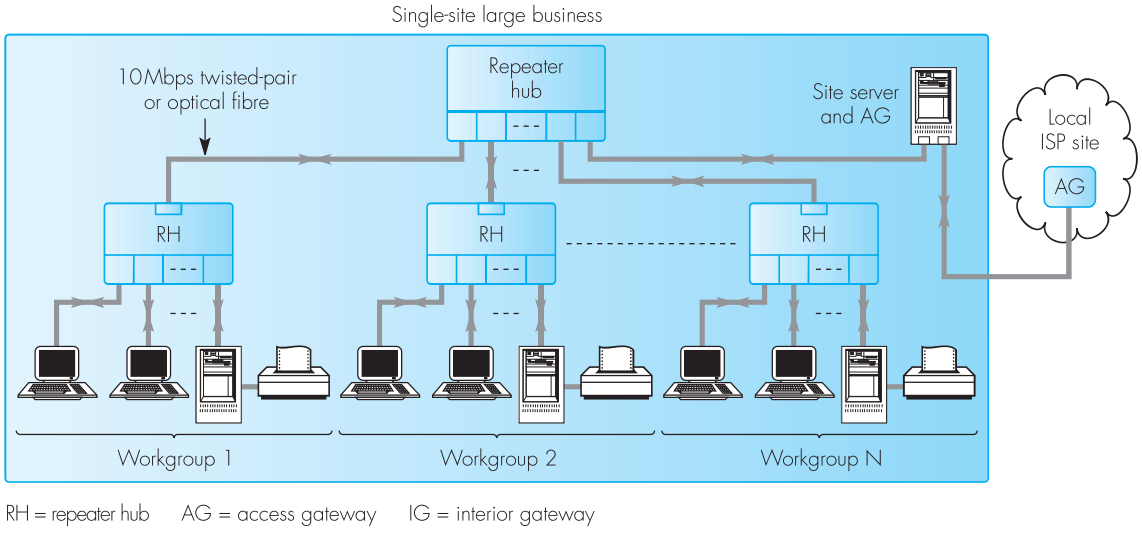


Figure 3.6 Access to the Internet for a single-site large business using repeater hubs only.

not load the rest of the network so increasing significantly the overall network throughput.

The presence of a bridging hub in a path between two communicating stations is transparent to the two stations and their associated repeater hubs. For this reason, bridging hubs are also known as **transparent bridges** and are often abbreviated to simply bridges. All routing decisions are made exclusively by the bridge(s). Moreover, a bridge automatically initialises and configures itself (in terms of its routing information) in a dynamic way after it has been put into service. A schematic of a bridge is shown in Figure 3.7(a) and a simple bridged LAN in Figure 3.7(b).

A LAN segment is physically connected to a bridge through a **bridge port**. A basic bridge has just two ports whereas a **multiport bridge** has a number of connected ports (and hence segments). In practice, each bridge port comprises the MAC integrated circuit chipset associated with the particular type of LAN segment – Ethernet – together with some associated port management software. The software is responsible for initializing the chipset at start-up – chipsets are all programmable devices – and for buffer management. Normally, the available memory is logically divided into a number of fixed-size units known as buffers. Buffer management involves passing a free buffer (pointer) to the chipset ready for receiving a new frame and passing the pointer of a full buffer to the chipset for onward transmission (forwarding).

Every bridge operates in the **promiscuous mode**, which means it receives and buffers all frames received on each of its ports. When a frame has been received at a port and put into the assigned buffer by the MAC chipset, the port management software prepares the chipset for a new frame and then passes the pointer of the memory buffer containing the received frame to the **bridge protocol entity** for processing. Since two (or more) frames may arrive concurrently at the ports and two or more frames may need to be forwarded from the same output port, the passing of memory pointers between the port management software and the bridge protocol entity software is carried out via a set of queues.

Frame forwarding (filtering)

As we show in Figure 3.7(b), a bridge maintains a **forwarding database** (also known as a **routing directory**) that indicates, for each port, the outgoing port (if any) to be used for forwarding each frame received at that port. If a frame is received at a port that is addressed to a station on the segment (and hence port) on which it was received, the frame is discarded; otherwise it is forwarded via the port specified in the forwarding database. The normal routing decision involves a simple look-up operation: the destination address in each received frame is first read and then used to access the corresponding port number from the forwarding database. If this is the same as the port on which it was received, the frame is discarded, else it is queued for forward transmission on the segment associated with the accessed port. This process is also known as **frame filtering**.

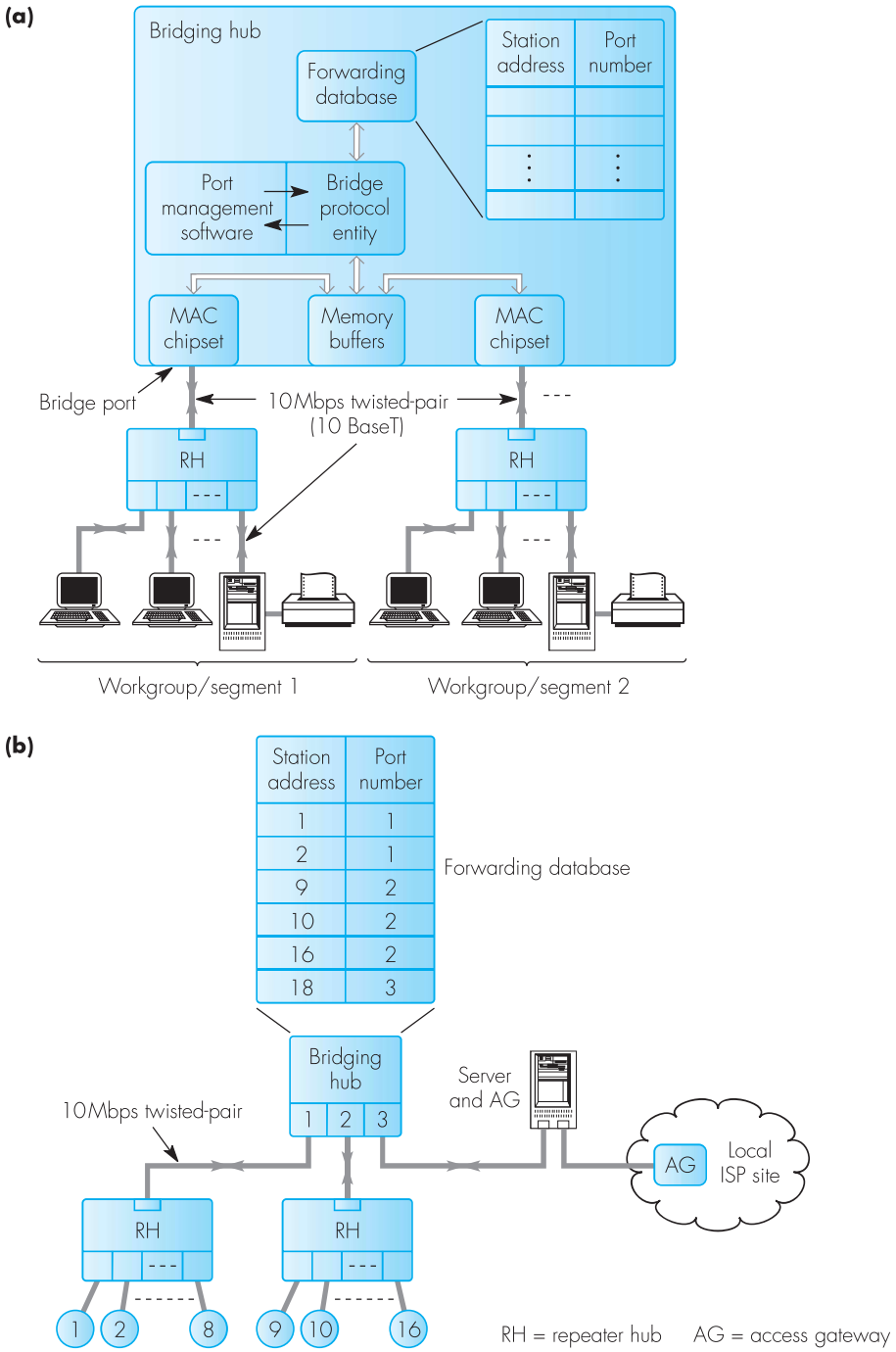


Figure 3.7 Transparent bridging hub schematic: (a) general topology; (b) multiple bridge example.

Bridge learning

A problem with transparent bridges is the creation of the forwarding database. One approach is for the contents of the forwarding database to be created in advance and held in a fixed memory, such as programmable read-only memory (PROM). The disadvantage is that the contents of the forwarding database in all bridges have to be changed whenever the network topology is changed – a new segment added, for example – or when a user changed the point of attachment (and hence segment) of his or her station. To avoid this, in most bridged LANs the contents of the forwarding database are not statically set up but rather are dynamically created and maintained during normal operation of the bridge. This is accomplished using a learning process, an overview of which is as follows.

When a bridge first comes into service, its forwarding database is initialized to empty. Whenever a frame is received, the *source address* within it is read and the incoming port number on which the frame was received is entered into the forwarding database. In addition, since the forwarding port is not known at this time, a copy of the frame is forwarded on all the other output ports of the bridge. This action is referred to as **flooding** since it ensures that a copy of each frame transmitted is received on all segments in the total LAN. During the learning phase this procedure is repeated for each frame received by the bridge. In this way, all bridges in the LAN rapidly build up the contents of their forwarding databases.

The MAC address associated with a station is fixed at the time of its manufacture. If a user changes the point of attachment to the network of his or her PC/workstation, the contents of the forwarding database in each bridge must be periodically updated to reflect such changes. To accomplish this, an **inactivity timer** is associated with each entry in the database. Whenever a frame is received from a station within the predefined time interval, the timer expires and the entry is removed. Whenever a frame is received from a station for which the entry has been removed, the learning procedure is again followed to update the entry in each bridge with the (possibly new) port number. In this way the forwarding database in a bridge is continuously updated to reflect the current LAN topology and the addresses of the stations that are currently attached to the segments it interconnects. The inactivity timer also limits the size of the database since it contains only those stations that are currently active. This is important since the size of the database influences the speed of the forwarding operation.

3.3.3 Switching hubs

Switching hubs – normally abbreviated to switches – are very similar to bridges in so much that they use the MAC addresses at the head of each frame for routing/switching purposes. The main difference is that a bridge discards frames that come from the same workgroup/segment whereas a switch switches/routes all the frames it receives. In addition, in order to improve the throughput rate, switches use duplex working over the cables that connect lower-level hubs to the switch. An example of a site network configuration using a switching hub is shown in Figure 3.8.

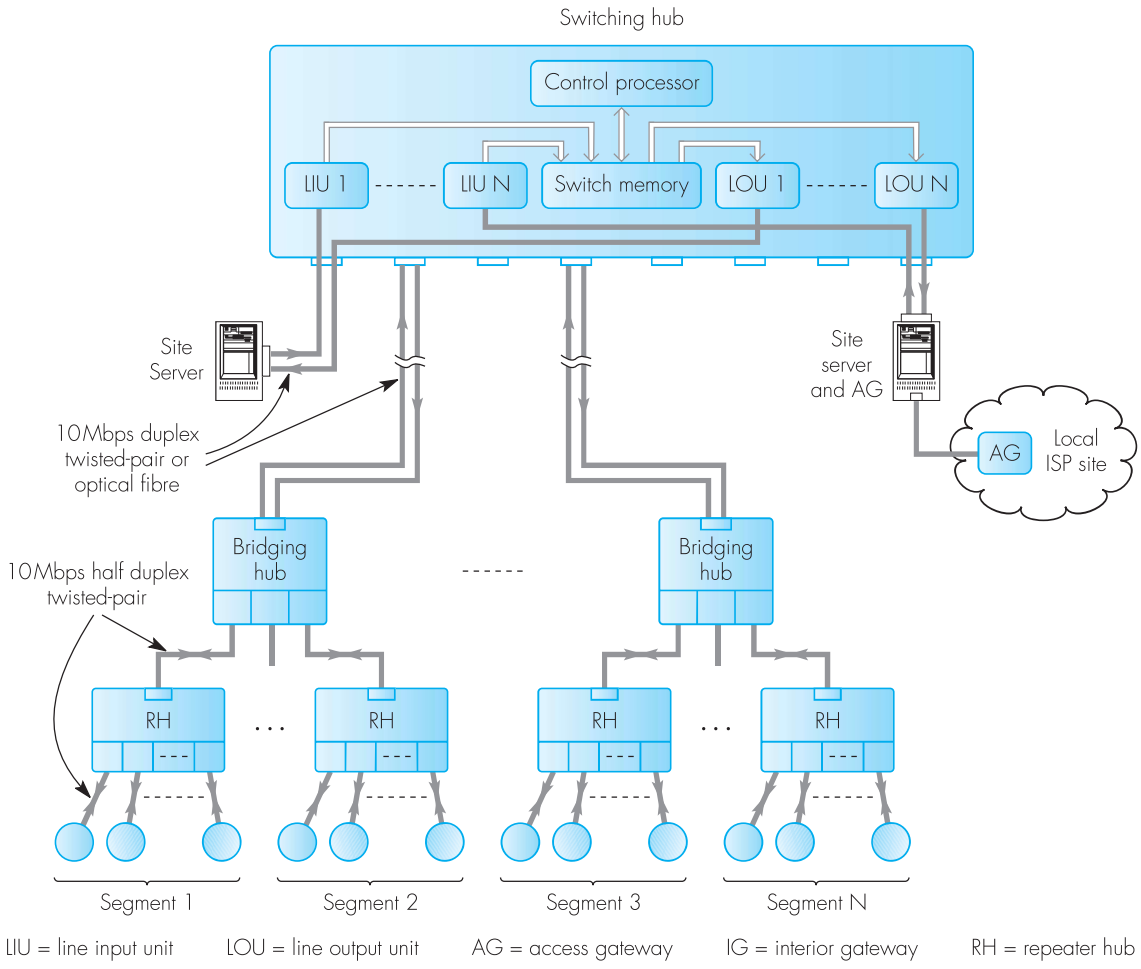


Figure 3.8 Example network configuration using a switching hub.

As we can see, lower-level hubs are connected to the switch by means of a pair of (duplex) cables that, typically, are implemented as UTP (or STP) cables or, for longer cable runs, dual multimode fibre cables. Also, with a switching hub CSMA/CD is not used and instead all lower-level hubs can transmit and receive frames concurrently. However, since repeater hubs operate in the half-duplex (repeating) mode, these cannot be connected directly to the switch and instead, bridging hubs must be used. Also, in order to obtain a high level of throughput, the two servers are connected directly to the switch by means of duplex cables.

Because each lower-level hub can transmit frames simultaneously, a frame may be received at multiple input ports of the switch – and hence require

processing – simultaneously. Similarly, two or more frames may require the same output line simultaneously. Hence associated with each input and output line is a memory buffer that can hold several fames waiting to be either processed (input) or transmitted (output). The frames – memory pointers to the start of the frame in practice – are stored in a FIFO queue. The control processor then reads the pointer to the frame at the head of each input queue in turn, obtains the destination MAC address from its head, and transfers the frame pointer to the tail of the required output queue to await transmission.

In order to retain the same connectionless mode of operation, when the switch is first brought into service – and subsequently at periodic intervals – the control processor enters a learning state similar to that used in transparent bridges. Hence when in the learning state, the switch simply initiates the onward transmission of a copy of each frame received from an input line onto all output lines. Prior to doing this, however, the control processor reads the source address from the head of the frame and keeps a record of this, together with the input port number of the port number on which the frame was received, in a routing table. The contents of the routing table are then subsequently used to route each received frame to specific output port. As we can deduce from this, there is a store-and-forward delay associated with a switch. Also, as with a bridge, the FCS at the tail of each frame is used to check for the presence of transmission errors prior to the frame being forwarded and corrupted frames are discarded.

Alternatively, some switches do not use a store-and-forward mode of operation and instead, providing the required output line is available, the switch starts to forward the frame as soon as the destination MAC address has been received. A switch that operates in this way is called a **cut-through switch**. However, since there is no provision for a frame to be stored, the frame is discarded if the required output line is busy.

3.4 High-speed LANs

As the application of LANs has become more diverse, so the demands on them in terms of information/data throughput have increased. As we have just described, by using a combination of bridges and a high bit rate backbone, the throughput of the total LAN is determined by the maximum throughput of each LAN segment. As we explained in Section 3.2, the maximum throughput of an Ethernet LAN is only a fraction of the 10 Mbps bit rate that is used. Hence in order to meet the higher throughput requirements of the newer multimedia applications, a number of high-speed LAN types have been developed. These include three variations of the basic Ethernet LAN: **Fast Ethernet**, **Switched Fast Ethernet**, and **Gigabit Ethernet**.

3.4.1 Fast Ethernet

The aim of Fast Ethernet was to use the same shared, half-duplex transmission mode as Ethernet but to obtain a ×10 increase in operational bit rate over 10BaseT while at the same time retaining the same wiring systems, MAC method, and frame format. As we explained in Section 3.2.2, when using hubs with unshielded twisted-pair (UTP) cable, the maximum length of drop cable from the hub to a station is limited to 100 m by the driver/receiver electronics. Assuming just a single hub, this means that the maximum distance between any two stations is 200 m and the worst-case path length for collision detection purposes is 400 m plus the repeater delay in the hub. Clearly, therefore, a higher bit rate can be used while still retaining the same CSMA/CD MAC method and minimum frame size of 512 bits. In the standard, the bit rate is set at 100 Mbps over existing UTP cable.

Line code

The major technological hurdle to overcome with Fast Ethernet was how to achieve a bit rate of 100 Mbps over 100 m of UTP cable. Category 3 UTP cable – as used for telephony, and the most widely installed – contains four separate twisted-pair wires. To reduce the bit rate used on each pair, all four pairs are used to achieve the required bit rate of 100 Mbps in each direction. Hence the standard is also known as **100Base4T**.

With the CSMA/CD access control method, in the absence of contention for the medium, all transmissions are half-duplex, that is, either station-to-hub or hub-to-station. In a 10BaseT installation, just two of the four wire pairs are used for data transfers, one in each direction. Collisions are detected when the transmitting station (or hub) detects a signal on the receive pair while it is transmitting on the transmit pair. Since the collision detect function must also be performed in 100Base4T, the same two pairs are used for this function. The remaining two pairs are operated in a bidirectional mode, as shown in Figure 3.9(a).

The figure shows that data transfers in each direction utilize three pairs – pairs 1, 3, and 4 for transmissions between a station and the hub and pairs 2, 3, and 4 for transmissions between the hub and a station. Transmissions on pairs 1 and 2 are then used for collision detection and carrier sense purposes as with 10BaseT. This means that the bit rate on each pair of wires need only be 100/3 = 33.33 Mbps.

If we use Manchester encoding, a bit rate of 33.33 Mbps requires a baud rate of 33.33 Mbaud, which exceeds the 30 Mbaud limit set for use with such cables, as above this, unacceptably high levels of crosstalk are obtained. To reduce the baud rate, a 3-level (**ternary**) **code** is used instead of straight (2-level) binary coding. The code used is known as **8B6T,** which means that, prior to transmission, each set of 8 binary bits is first converted into 6 ternary (3-level) symbols. From the example shown in Figure 3.9(b), we can deduce that this yields a symbol rate of:

$$\frac{100 \times 6/8}{3} = 25\,\text{Mbaud}$$

which is well within the set limit.

The three signal levels used are +V, 0, –V, which are represented simply as +, 0, –. The codewords are selected such that the line is DC balanced, that is, the mean line signal is zero. This maximizes the receiver's discrimination of the three signal levels since these are then always relative to a constant 0 (DC) level. To achieve this, we exploit the inherent redundancy present in the use of 6 ternary symbols. The 6 ternary symbols means that there are 729 (3^6)

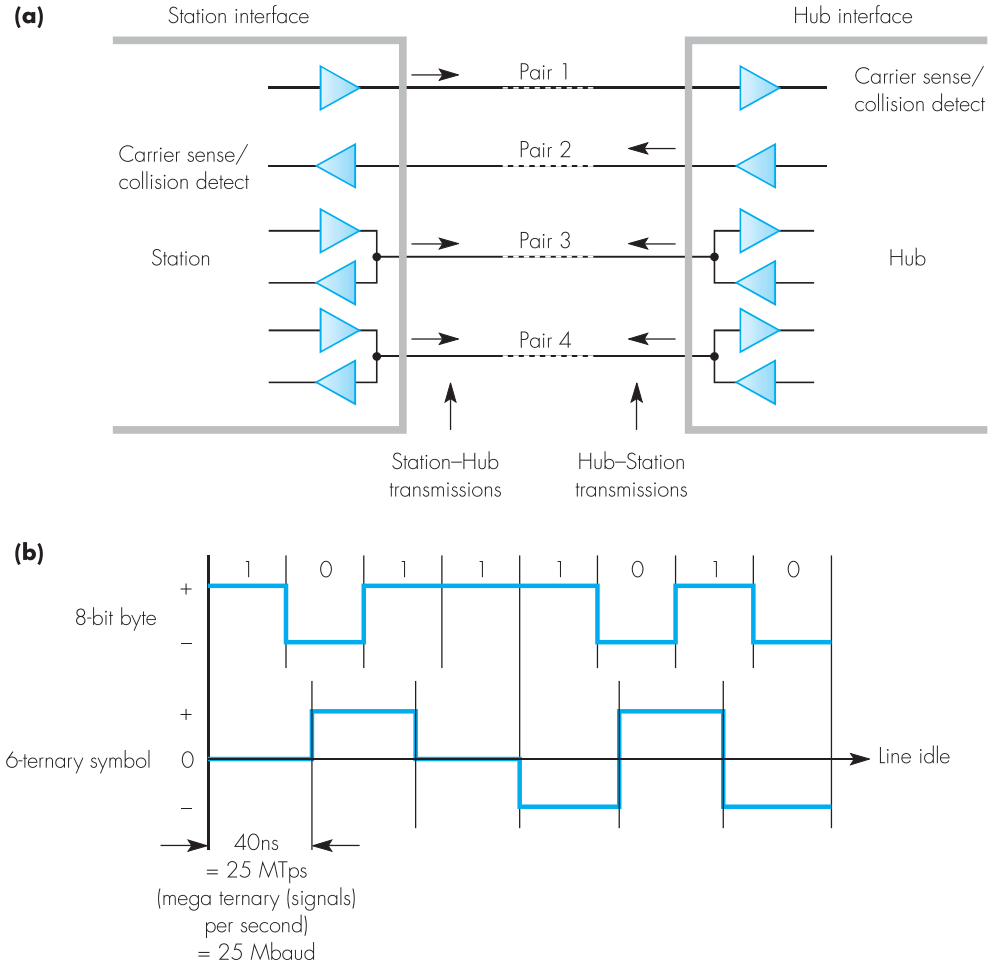


Figure 3.9 100Base4T: (a) use of wire pairs; (b) 8B6T encoding.

possible codewords. Since only 256 codewords are required to represent the complete set of 8-bit byte combinations, the codes used are selected, firstly, to achieve DC balance and secondly, to ensure all codewords have at least two signal transitions within them. This is done so that the receiver DPLL maintains clock synchronization.

To satisfy the first condition, we choose only those codewords with a combined weight of 0 or +1 and 267 codes meet this condition. To satisfy the second condition, we eliminate those codes with fewer than two transitions – five codes – and also those starting or ending with four consecutive zeros – six codes. This leaves the required 256 codewords, the first 128 of which are listed in Table 3.1.

DC balance

As we have just indicated, all the codewords selected have a combined weight of either 0 or +1. For example, the codeword +−−+0 0 has a combined weight of 0 while the codeword 0+++−− has a weight of +1. Clearly, if a string of codewords each of weight +1 is transmitted, then the mean signal level at the receiver will move away rapidly from the zero level, causing the signal to be misinterpreted. This is known as **DC wander** and is caused by the use of transformers at each end of the line. The presence of transformers means there is no path for direct current (DC).

To overcome this, whenever a string of codewords with a weight of +1 is to be sent, the symbols in alternate codewords are inverted prior to transmission. For example, if a string comprising the codeword 0+++−− is to be sent, then the actual codewords transmitted will be 0+++−−, 0−−−++, 0+++−−, 0−−−++, and so on, yielding a mean signal level of 0. At the receiver, the same rules are applied and the alternative codewords will be reinverted into their original form prior to decoding. The procedure used for transmission is shown in the state transition diagram in Figure 3.10(a).

To reduce the latency during the decoding process, the 6 ternary symbols corresponding to each encoded byte are transmitted on the appropriate three wire pairs in the sequence shown in Figure 3.10(b). This means that the sequence of symbols received on each pair can be decoded independently. Also, the frame can be processed immediately after the last symbol is received.

End-of-frame sequence

The transmission procedure adopted enables further error checking to be added to the basic CRC. We can deduce from the state transition diagram in Figure 3.11(a) that the running sum of the weights is always either 0 or +1. At the end of each frame transmission – that is, after the four CRC bytes have been transmitted – one of two different **end-of-stream** (**EOS**) codes is transmitted on each of the three pairs. The code selected effectively forms a checksum for that pair. The principle of the scheme is shown in Figure 3.10(c).

In this figure, we assume the last of the four CRC bytes (CRC-4) is on pair 3. The next codeword transmitted on pair 4 is determined by whether the running

Table 3.1 First 128 codewords of 8B6T codeword set.

Data byte	Codeword	Data byte	Codeword	Data byte	Codeword	Data byte	Codeword
00	−+00−+	20	−++−00	40	−00+0+	60	0++0−0
01	0−+−+0	21	+00+−−	41	0−00++	61	+0+−00
02	0−+0−+	22	−+0−++	42	0−0+0+	62	+0+0−0
03	0−++0−	23	+−0−++	43	0−0++0	63	+0+00−
04	−+0+0−	24	+−0+00	44	−00++0	64	0++00−
05	+0−−+0	25	−+0+00	45	00−0++	65	++0−00
06	+0−0−+	26	+00−00	46	00−+0+	66	++00−0
07	+0−+0−	27	−+++−−	47	00−++0	67	++000−
08	−+00+−	28	0++−0−	48	00+000	68	0++−+−
09	0−++−0	29	+0+0−−	49	++−000	69	+0++−−
0A	0−+0+−	2A	+0+−0−	4A	+−+000	6A	+0+−+−
0B	0−+−0+	2B	+0+−−0	4B	−++000	6B	+0+−−+
0C	−+0−0+	2C	0++−−0	4C	0+−000	6C	0++−−+
0D	+0−+−0	2D	++00−−	4D	+0−000	6D	++0+−−
0E	+0−0+−	2E	++0−0−	4E	0−+000	6E	++0−+−
0F	+0−−0+	2F	++0−−0	4F	−0+000	6F	++0−−+
10	0−−+0+	30	+−00−+	50	+−−+0+	70	000++−
11	−0−0++	31	0+−−+0	51	−+−0++	71	000+−+
12	−0−+0+	32	0+−0−+	52	−+−+0+	72	000−++
13	−0−++0	33	0+−+0−	53	−+−++0	73	000+00
14	0−−++0	34	+−0+0−	54	+−−++0	74	000+0−
15	−−00++	35	−0+−+0	55	−−+0++	75	000+−0
16	−−0+0+	36	−0+0−+	56	−−++0+	76	000−0+
17	−−0++0	37	−0++0−	57	−−+++0	77	000−+0
18	−+0−+0	38	+−00+−	58	−−0+++	78	+++−−0
19	+−0−+0	39	0+−+−0	59	−0−+++	79	+++−0−
1A	−++−+0	3A	0+−0+−	5A	0−−+++	7A	+++0−−
1B	+00−+0	3B	0+−−0+	5B	0−−0++	7B	0++0−−
1C	+00+−0	3C	+−0−0+	5C	+−−0++	7C	−00−++
1D	−+++−0	3D	−0++−0	5D	−000++	7D	−00+00
1E	+−0+−0	3E	−0+0+−	5E	0+++−−	7E	+−−−++
1F	−+0+−0	3F	−0+−0+	5F	0++−00	7F	+−−+00

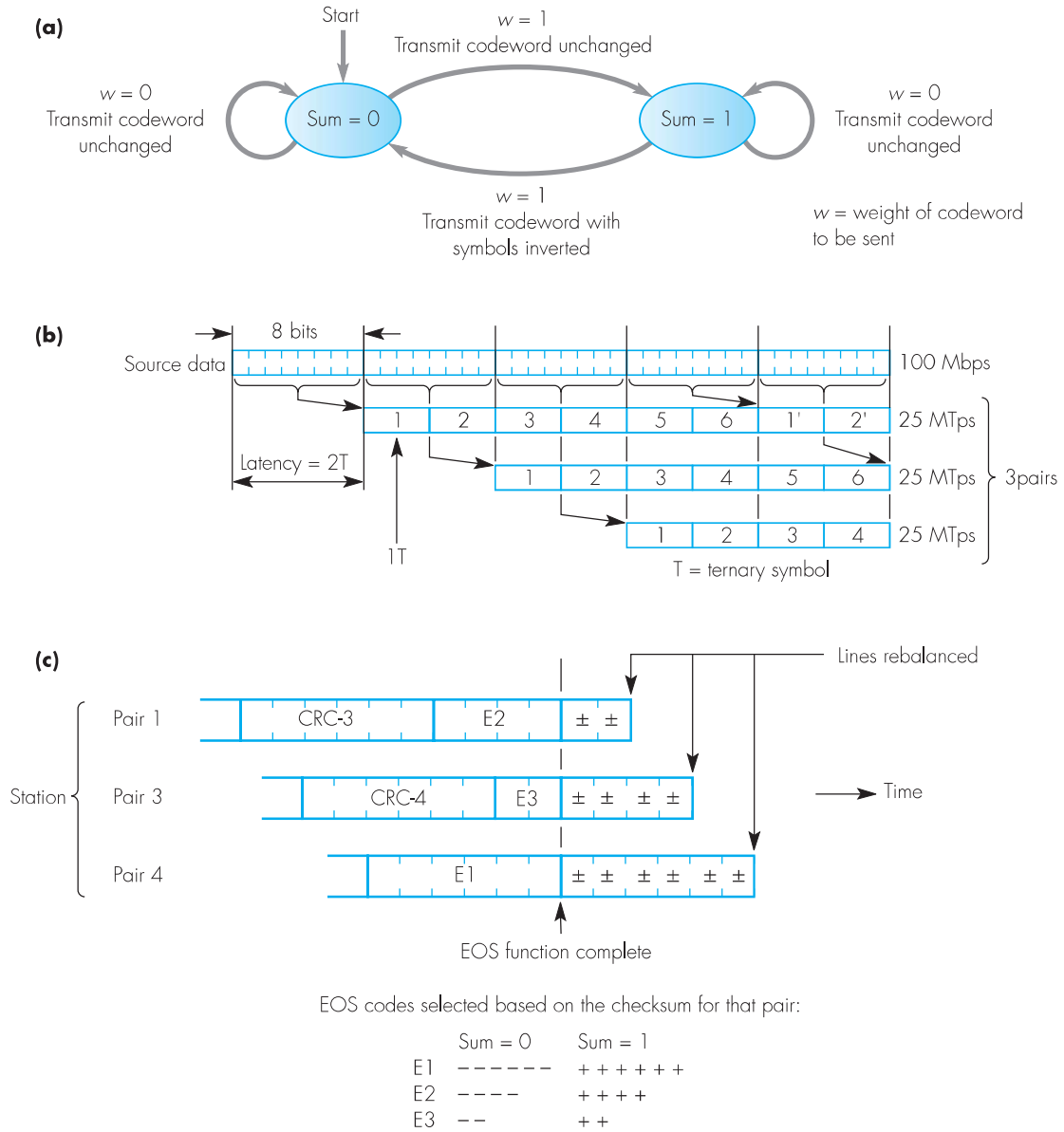


Figure 3.10 100Base4T transmission detail: (a) DC balance transmission rules; (b) 8B6T encoding sequence; (c) end of stream encoding.

sum of the weights on that pair – referred to as the checksum – is 0 or +1. The EOS function is complete at the end of this codeword and the length of the other two EOS codes are reduced by two or one times the latency, that is, 4T or 2T. This means the receiver can detect reliably the end of a frame since all signals should cease within a short time of one another. This allows for very small variations in propagation delay on each pair of wires.

Collision detection

An example station–hub transmission without contention is shown in Figure 3.11(a). Recall that a station detects a collision by detecting a signal on pair 2 while it is transmitting and, similarly, the hub detects a collision by the presence of a signal on pair 1. However, as Figure 3.11(a) shows, the strong (unattenuated) signals transmitted on pairs 1, 3, and 4 from the station side each induce a signal into the collision detect – pair 2 – wire. This is near-end crosstalk (NEXT) and, in the limit, is interpreted by the station as a (collision) signal being received from the hub. The same applies for transmissions in the reverse direction from hub to station.

To minimize any uncertainty the preamble at the start of each frame is encoded as a string of 2-level (as opposed to 3-level) symbols, that is, only positive and negative signal levels are present in each encoded symbol. This increases the signal-level amplitude variations, which, in turn, helps the station/hub to discriminate between an induced NEXT signal and the preamble of a colliding frame.

The preamble pattern on each pair is known as the **start of stream** (**SOS**) and is made up of two 2-level codewords, SOS-1 and SFD. The complete pattern transmitted on each of the three pairs is shown in Figure 3.11(b) and, as we can see, the SFD codeword on each pair is staggered by sending only a single SOS-1 on pair 4. This means that the first byte of the frame is transmitted on pair 4, the next on pair 1, the next on pair 3, and so on. An acceptable start of frame requires all three SFD codes to be detected, and the staggering of them means that it takes at least four symbol errors to cause an undetectable start-of-frame error.

On detecting a collision, a station transmits the jam sequence and then stops transmitting. At this point, the station must be able to determine when the other station(s) involved in the collision cease transmitting in order to start the retry process. In practice, this is relatively easy since, in the nontransmitting (idle) state with 8B6T encoding, a zero signal level is present on the three data wires. This means that there is no induced NEXT signal in the collision detect wire, which, in turn, enables the completion of the jam sequence from the hub side to be readily determined. Also, to improve the utilization of the cable, the interframe gap time is reduced from 9.6 μs to 960 ns.

100BaseX

In addition to the 100Base4T standard, a second Fast Ethernet standard is available which is known as **100BaseTX**. Unlike 100Base4T, which was

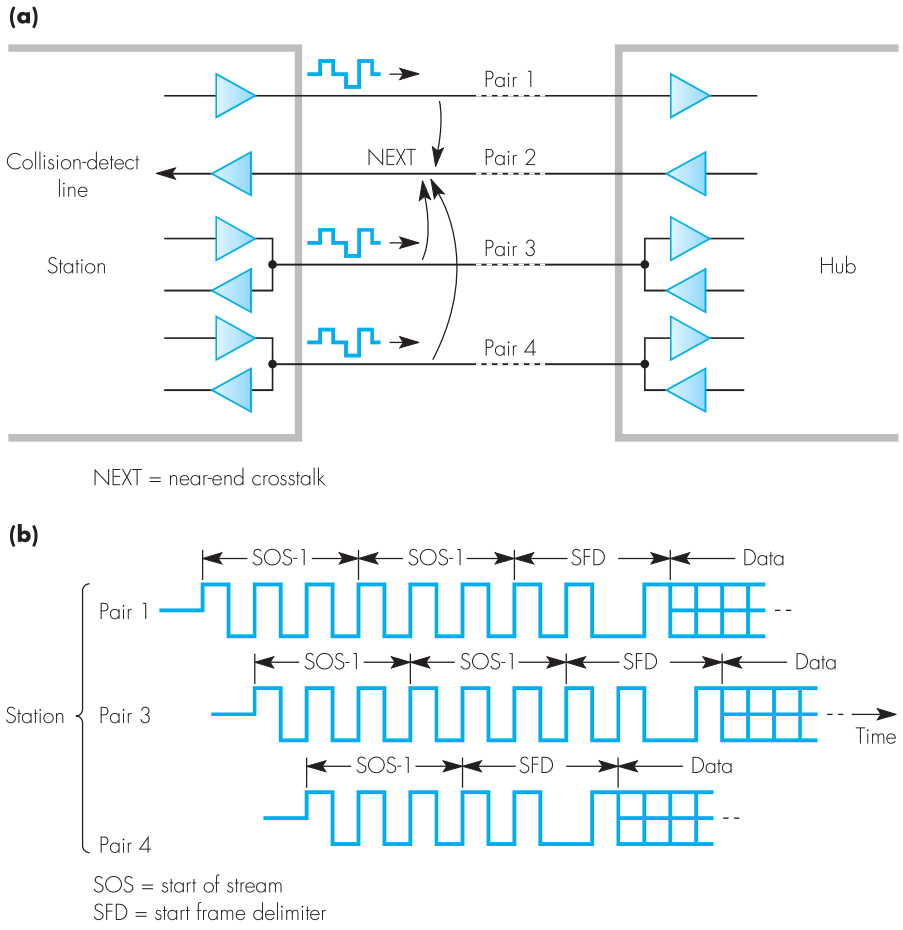


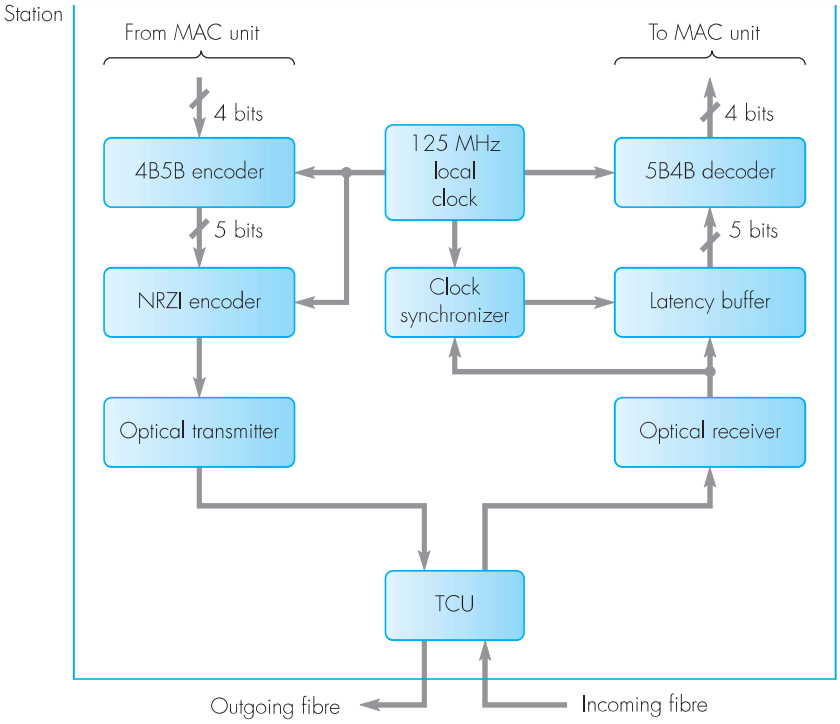
Figure 3.11 Start-of-frame detail: (a) effect of NEXT; (b) preamble sequence.

designed for use with existing category 3 UTP cable, 100BaseTX was designed for use with the higher quality category 5 cable now being used in most new installations. In addition, it is intended for use with STP. The use of various types of transmission media is the origin of the "X" in the name.

Each different type of transmission medium requires a different physical sublayer. The first to be developed was that for use with multimode optical fibre cable. It uses a bit encoding scheme known as 4B5B (sometimes written 4B/5B) and is known as **100BaseFX**. A schematic diagram showing the physical interface to the fibre cable is shown in Figure 3.12(a).

Each interface has its own local clock. Outgoing data is transmitted using this clock while incoming data is received using a clock that is frequency and phased locked to the transitions in the incoming bitstream. As we shall see,

(a)



(b)

Data symbols

4-bit data group	5-bit symbol
0000	11110
0001	01001
0010	10100
0011	10101
0100	01010
0101	01011
0110	01110
0111	01111
1000	10010
1001	10011
1010	10110
1011	10111
1100	11010
1101	11011
1110	11100
1111	11101

Control symbols

IDLE	11111
J	11000
K	10001
T	01101
R	00111
S	11001
QUIET	00000
HALT	00100

Figure 3.12 100BaseFx:(a) physical interface schematic; (b) 4B/5B codes.

all data is encoded prior to transmission so that there is a guaranteed transition in the bitstream at least every two bit-cell periods. This ensures that each received bit is sampled/clocked very near to the bit cell centre.

All data to be transmitted is first encoded prior to transmission using a **4 of 5 group code**. This means that for each 4 bits of data to be transmitted, a corresponding 5-bit codeword/symbol is generated by what is known as a **4B5B encoder**. The 5-bit symbols corresponding to each of the sixteen possible 4-bit groups are shown in Figure 3.12(b). As we can see, there is a maximum of two consecutive zero bits in each symbol. The symbols are then shifted out through a further NRZI encoder, the operation of which we described in the subsection on synchronous transmission in Chapter 1. This produces a signal transition whenever a 1 bit is being transmitted and no transition when a 0 bit is transmitted. In this way, there is a guaranteed signal transition at least every two bits.

The use of 5 bits to represent each of the sixteen 4-bit groups means that there are a further sixteen unused combinations of the 5 bits. Some of these combinations/symbols are used for other (link) control functions such as indicating the start and end of each transmitted frame. A list of the link control symbols is shown in Figure 3.12(b).

The cable comprises two fibres, one of which is used for transmissions between the station and hub and the other for transmissions between the hub and the station. As with 10BaseT, collisions are detected if a (colliding) signal is present on the receive fibre during the time the station is transmitting a frame. However, because of the additional cost of both the electrical-to-optical conversion circuits and the associated optical plugs and sockets that are required per port, the cost of the MAC unit associated with the NIC is higher than that used with 100Base4T. Hence the most popular type of Fast Ethernet is 100Base4T and 100BaseFX is used primarily when longer drop cables are required.

3.4.2 Switched Fast Ethernet

As we explained at the start of Section 3.4.1, Fast Ethernet uses the same shared, half-duplex transmission mode as Ethernet. Hence in applications that involve access to, say, large enterprise Web servers, even though the server can handle multiple transfers concurrently, the overall access time and throughput experienced by the various stations using the server is limited by the shared access circuit connecting the server (station) to the hub.

In order to allow multiple access/transfers to be in progress concurrently, two developments have been made: the first, the introduction of a switched hub architecture, and the second, duplex working over the circuits that connect the stations to the hub. The resulting type of hub is known as a **Fast Ethernet switch**.

Switch architecture

The general architecture of a switching hub is shown in Figure 3.13. As we can see, each station is connected to the hub by means of a pair of (duplex) lines that, typically, are implemented as dual UTP (or STP) cables or dual multimode fibre cables. Recall from the last section, each UTP (and STP) cable contains four separate twisted pairs. In the case of 100Base4T, three pairs are used to transmit the 100 Mbps bitstream – in a half-duplex mode – and the fourth pair is used to perform the carrier sense and collision detection functions. As we described earlier in Section 3.3.3, with a switching hub CSMA/CD is not used and instead all stations can transmit and receive frames concurrently. Hence, as with 100Base4T, three pairs in each cable are used collectively to transmit frames (at 100 Mbps) in each direction.

In the case of dual multimode fibre cables, as we described earlier, each fibre is used to transmit at 100 Mbps over several kilometers, one in each direction of transmission. Since the 4B5B coding scheme is used, the line signalling rate is 125 Mbaud. In addition, an active signal is maintained on each fibre continuously by transmitting an idle symbol during the idle period between frames. This ensures the receiver DPLL can maintain clock synchronism between successive frame transmissions.

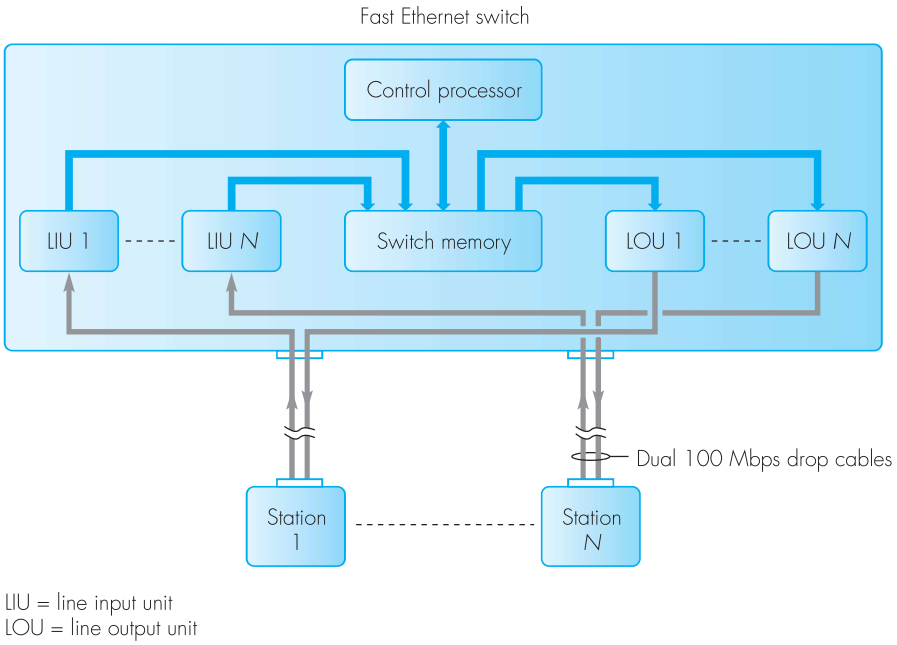


Figure 3.13 Fast Ethernet switch schematic.

Flow control

As we can see from the above with a store-and-forward switch, under heavy load conditions it is possible for all the frame buffers within the switch to become full. At this point, therefore, the control processor must discard any new frame(s). Alternatively, an optional feature associated with switched hubs is to incorporate flow control into the switch. When using flow control, should the control processor find that the level of memory in use reaches a defined threshold, it initiates the transmission of what is called a **Pause** frame on all of its input ports.

On receipt of a Pause frame, the attached station must then stop sending any further frames to the switch until either a defined time has expired or it receives a notification from the switch that the overload condition has passed. Having sent a Pause frame, the control processor monitors the level of memory in use and, when this falls below a second level, it sends out a **Continue** frame on all input ports to inform the attached stations that they can now resume sending new frames.

The two control frames are normal Ethernet frames with a defined code in the *type* field of the frame header. The first two bytes of the data field then give the command – Pause/Continue – and, in the case of the Pause command, succeeding bytes are then used to indicate the duration of the pause in multiples of the minimum frame transmission time. For fast Ethernet (100 Mbps), this is 5.12 microseconds and the maximum pause duration is 336 milliseconds.

Network configurations

An example network configuration that includes a Fast Ethernet switching hub is shown in Figure 3.14. As we can see, in order to obtain a high level of throughput, the two servers are connected directly to the switch by means of duplex 100 Mbps lines. All the end-user stations then gain access to the servers through either a 10BaseT or a 100Base4T hub. As we explained in the last section, both these types of hub operate in the half-duplex repeating mode. Hence the duplex uplink port connecting each hub to the switch is through a bridging hub since this has bridging circuitry within the hub to temporarily buffer all frame transfers to and from the switch and to perform the CSMA/CD MAC protocol associated with the shared medium hub ports. The switch also automatically configures the operating speed of each of its ports to be either 10 or 100 Mbps.

3.4.3 Gigabit Ethernet

As the name implies, the drop cables associated with Gigabit Ethernet hubs operate at 1000 Mbps (1Gbps). The standard is defined in IEEE802.3z and has been introduced to meet the throughput demands of an increasing number of servers that hold files containing multimedia information; examples include Web pages comprising very high resolution graphics, motion video and general audio. The hub can be either a simple repeater hub – that

is, one that has no memory associated with it and operates in the half-duplex mode – or a switched hub that operates in the duplex mode.

Repeater hub

An example application of a repeater hub is to distribute the output of a powerful supercomputer (performing, say, 3D scientific visualizations) to a localized set of workstations. The main issue when operating in the half-duplex – CSMA/CD – mode is to ensure that the round-trip delay between any two stations – the slot time – exceeds the time required to transmit the smallest allowable frame of 512 bits. The time to transmit a 512-bit frame at 1 Gbps is 0.512 μs. However, with a maximum cable length of 2.5 km, the signal propagation delay would still be 12.5 μs. Hence, as we computed earlier in Section 3.2.1, the slot time would still be in the order of 50 μs. Hence at 1 Gbps, the minimum frame size to detect a collision would be 50 000 bits or 6 250 bytes.

Clearly, this is not acceptable and the maximum length of drop cable has to be reduced significantly. The initial maximum length was set at 25 m which, with a hub topology, means that the worst-case signal propagation distance is then 4 × 25 or 100 m. Hence, assuming the velocity of propagation of a signal through the transmission medium is, say, 2×10^8 m s^{-1}, the worst-case time – the slot time – is $100/2 \times 10^8 = 0.5\,\mu s,^{-1}$ which is the same as the time to transmit a 512-bit frame at 1 Gbps.

The choice of 25 m, however, was rejected by the standards committee as being too small and, after much debate and lobbying, the maximum length of drop cable was set at 200 m. This has a nominal slot time of $800/2 \times 10^8 = 4$ μs and, to ensure the sender is still transmitting when a collision/noise-burst is detected, the minimum frame size must be in excess of $4 \times 10^{-6} \times 1 \times 10^9 =$ 4 000 bits or 500 bytes. The standard has set 512 bytes as the minimum frame size compared with the existing 64 bytes.

When a frame of less than 512 bytes is being transmitted – determined by the length indicator in the frame header – the sending MAC interface hardware adds padding bytes to extend the frame to 512 bytes. The receiving MAC interface then removes the added parity bytes before passing the frame (memory pointer) to the required output MAC interface.

This procedure is known as **carrier extension** and, as we can deduce from the foregoing, the link utilization can be as low as 12.5%. Hence, in addition, a second scheme known as **frame bursting** is used. This allows the sending station to transmit a set of smaller – should these be queued and awaiting transmission – in a single block. Again, if necessary, padding bytes are added to ensure the total block size is greater than 512 bytes.

Switching hub

As with a Fast Ethernet switch, CSMA/CD is not used and instead duplex transmission is used using two separate cables. Each frame is transmitted – by a lower level hub or station – and, on arrival at the switch MAC interface port,

the frame is first stored/buffered and then processed before it is queued for onward transmission at the required output MAC interface port. Hence, since CSMA/CD is not used, the length of the drop cables is determined solely by the attenuation characteristics of the transmission cable that is used.

Cabling

In relation to cabling, repeating hubs can use either category 5 UTP cable with a drop cable length of up to 100 m (1000Base4T) or STP cable providing the length of the drop cable is limited to 25 m (1000BaseCX). In the case of a switching hub – used as a backbone for example – optical fibre cable is used that supports drop cable lengths of up to 550 m using multimode fibre (1000BaseSX) or up to 5 km using monomode fibre (1000BaseLX).

Signal encoding

The signal encoding scheme used has the same aims as those of Fast Ethernet, which we studied in detail in Section 3.4.1. However, two new encoding schemes have been selected for use with the two types of cable. The scheme selected for use with fibre cable is called 8B/10B; that is, each 8-bit byte – in the bitstream – is encoded into a 10-bit symbol. Hence, since the line bit rate is 1 Gbps, the line signalling rate is extended to 1.25 Gbaud. The use of 10-bit symbols means there are 1024 different symbols available to represent each of the 256 different 8-bit bytes. The choice of symbol for each byte is such that no symbol has more than four of the same bit – 0 or 1 – in a row and no more than six 0s or six 1s. In this way each symbol has enough signal transitions to ensure the receiver clock stays in synchronism with the incoming bitstream. In addition, in order to keep the mean DC level of the bitstream near to zero – so maximizing the amplitude of the signal transitions at the receiver – because there are more than 256 codewords that meet the above conditions, many of the 256 bytes have two alternative codewords assigned to them. Then, when these bytes are present in the byte stream, the codeword chosen is such that the number of 0 or 1 bits in the overall encoded bitstream is equalized.

As we indicated earlier, there is also a new encoding scheme used with the 1000Base4T twisted-pair cable. Since category 5 UTP cable has four twisted-pairs within it, all four pairs are used to transmit each byte, two bits per pair. There are four combinations of two bits: 00, 01, 10 and 11. The pair of bits that are to be transmitted on each pair is represented by one of five different voltage levels, the fifth level indicating that the bits on the four pairs form either a data byte or a control byte such as a start-of-frame and end-of-frame delimiter. In this way, 8 bits are transmitted every clock cycle, which reduces the line signalling rate from 1 Gbaud to 125 Mbaud.

Flow control

As with Fast Ethernet, a flow control scheme is used based on the use of the Pause and Continue control frames. This is the same as that used with Fast

Internet, which we described in Section 3.4.2. The only difference is the minimum frame size, which, at 1 Gbps, is 0.512 μs. As we explained, this is also the minimum pause time and larger pauses are multiples of this up to 33.6 ms.

Network configurations

This can be the same as that shown in Figure 3.14 with Fast Ethernet, the only difference being the replacement of the Fast Ethernet switch with a Gigabit Ethernet switch. Alternatively, multiple Fast Ethernet switches can be used each with a gigabit uplink to the Gigabit Ethernet switch.

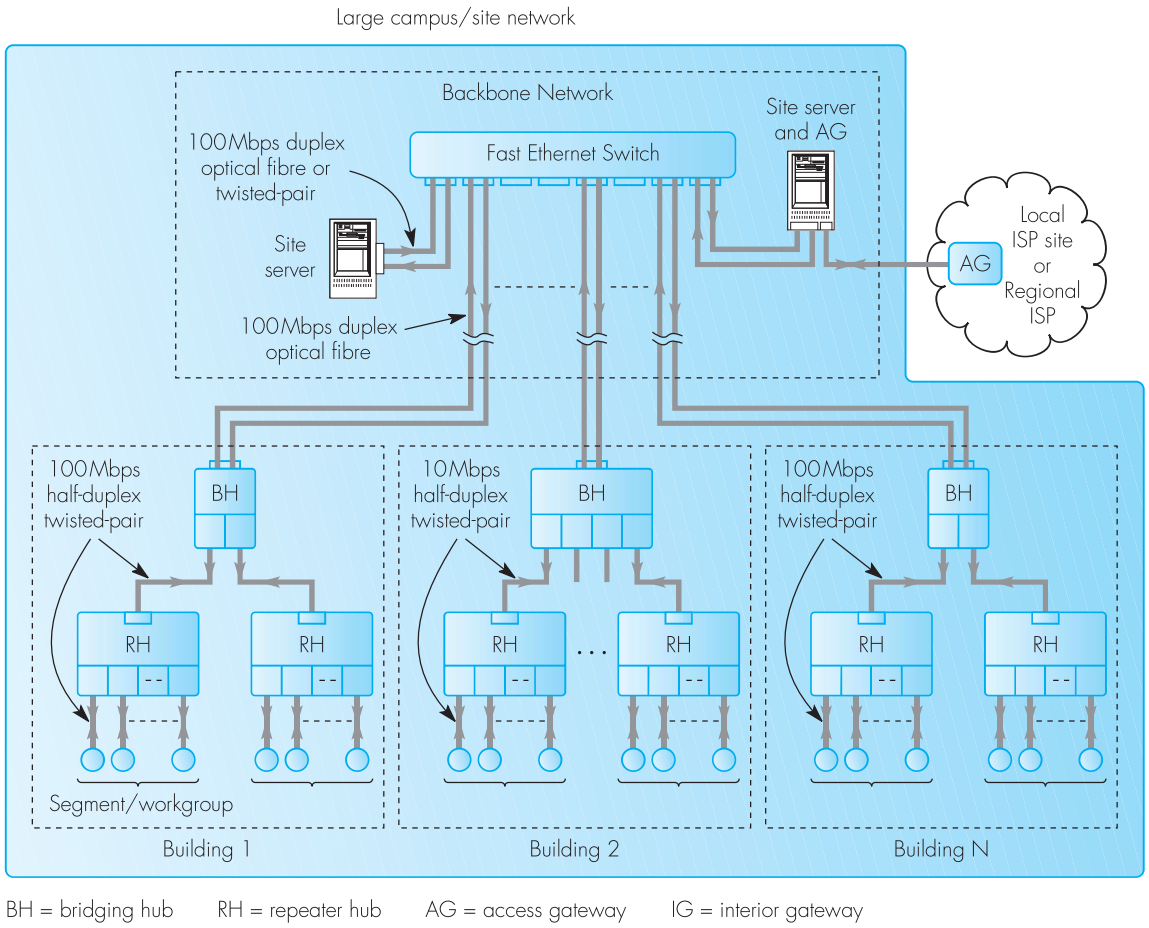


Figure 3.14 Example network configuration for a large site/campus using a mix of repeater hubs, bridging hubs and a Fast Ethernet switch.

Further developments

A standard is also available for 10-gigabit Ethernet. This is known as **IEEE802.3ae** and demonstrates the constantly increasing demands for higher bandwidths with multimedia applications involving, for example, streamed audio and video. In addition to being used for higher bit rate LANs, however, the maximum distance supported by the new standard is up to 40 km (24 miles) – 10GBase-E (extended) – using single-mode fibre. This means that it can be used in new application domains such as **metropolitan area networks** (**MANs**) that we shall discuss later in Section 3.7.5.

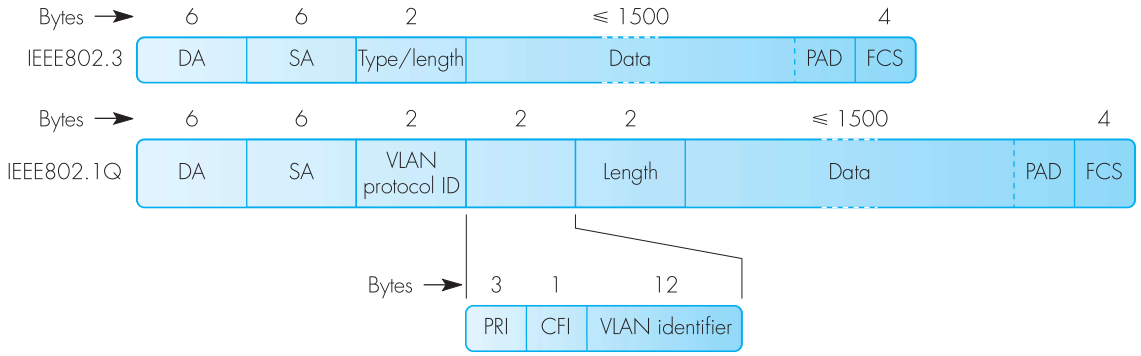
3.5 Virtual LANs

As we saw in the network configuration shown in Figure 3.14, with a LAN composed of hierarchical hubs, each workgroup – for example, within an office or department – can be physically separated from all the other workgroups. In practice, however, this is not always possible. For example, in a college/university department often a member of staff is responsible for managing the accounts relating to research grants, general teaching and research funding. Hence to do this, he/she needs access to the finance department LAN to, say, check on payments from grants, salaries of research staff, purchases, etc. However, since the servers that are attached to the finance department LAN hold sensitive information, for security reasons it is necessary for the PC/workstation of the member of staff responsible for financial matters to be attached to the finance department LAN even though the PC is physically located/attached to the department LAN.

Clearly, one approach is to relocate the member of staff to the finance office. In many instances, however, financial matters are only part of his/her job description and hence it is preferable for him/her to be located locally within the department but logically linked to the finance department LAN. To overcome this type of problem, therefore, the IEEE has produced a standard that allows a machine that is physically attached to one LAN to be a member of a workgroup associated with a different LAN. The total LAN is then known as a **virtual LAN** or **VLAN** and the standard is defined in **IEEE 802.1Q**. In this section we shall describe the operation of VLANs and identify the advantages that they can bring.

3.5.1 IEEE802.1Q

As we indicated earlier in Figure 3.3 and explained in the accompanying text, the *type* field in the original Ethernet specification was replaced with a *type/length* field when the IEEE 802.3 standard was introduced. Since that time, however, all subsequent standards have retained the same frame format. The approach adopted with the VLAN standard, however, was to introduce a new frame format.



DA = destination MAC address

SA = source MAC address

VLAN protocol ID = 8100 Hex

VLAN identifier: this identifies the VLAN to which the frame belongs

PRI = priority field (for possible future use)

CFI = canonical format identifier – used to enable a frame relating to a token ring LAN to be embedded within the data field of the frame

Note: 1. The VLAN protocol ID of 8100 Hex is greater than 1500 and hence is interpreted as a type value
2. The maximum frame size is now 1522 bytes

Figure 3.15 IEEE802.1Q frame format and field descriptions.

Frame format

The new frame format – together with descriptions of the various fields – is shown in Figure 3.15.

As we can see, the new frame format has two new 2-byte fields. Following the source and destination MAC address fields is the existing *type/length* field. This is set to 8100Hex and, since this is greater than 1500, it is interpreted as a type value and 8100Hex indicates the **VLAN protocol ID**.

The next 2-byte field is composed of three subfields. The first is a 3-bit **priority** (**PRI**) field and has been introduced to enable (future) frames to be assigned a priority value in order to define the order that frames are transmitted. For example, a frame carrying real-time speech/video could be given a higher priority than a frame carrying a string of textual information. Although this may seem an attractive feature, as we shall see in Chapter 6, it is already present in the header of IP packets.

The second subfield is called the **canonical format identifier** (**CFI**) and is used to enable a frame relating to a Token ring LAN to be embedded within the data field of the frame. Token ring LANs are still found in some legacy networks and hence this feature allows such frames to be transmitted within the total VLAN network.

The third subfield is a 12-bit **VLAN identifier**. As we shall see, each workgroup is allocated a separate VLAN ID and each frame transmitted by members of the same workgroup has the same identifier in this field.

The next 2-byte field is then a new *length* field and this indicates the number of bytes in the data field.

Frame forwarding

In order to explain the frame forwarding operation associated with a VLAN network, consider the simple LAN topology shown in Figure 3.16. It is based on bridging hubs with both the lowest tier of three hubs and the single hub in the upper tier all being IEEE802.1Q compliant. Also, all the stations that are attached to the total network have network interface cards (NICs) that are compliant; that is, they generate and process frames that are in the new format. This means that each station is assigned a specific VLAN identifier that indicates the VLAN to which the station belongs and, since this is done when the station is first initialized, it can be readily changed should modifications to the current workgroups become necessary.

In the standard, each PC/server can be identified by either its port number, MAC address or, in some instances, its (Internet) IP address. As we shall see in Chapter 6, the IP address is found within the data field of the MAC frame and hence can lead to problems should alternative network address formats from IP be present. Also, since the port number associated with a station changes whenever a PC/server is relocated, this also can create problems. Hence in most cases each PC/server is identified by its MAC address.

As we showed in Figure 3.7 and explained in the accompanying text, at startup a bridge learns the port number to which each PC/server is attached by reading the source (MAC) address in the header of each frame received at a port before the frame is forwarded out onto all the other ports. The port number, together with the related MAC address, is then entered into the routing table of the bridge.

The same procedure is followed with an IEEE802.1Q compliant bridging hub with the addition that the VLAN identifier in the header of each frame is also entered into the routing table. Similarly, during the learning phase, a copy of the frame is forwarded to the hub at the higher level and this in turn creates its own routing table. As an example, the routing tables of the four bridges are as shown in Figure 3.16.

Once the learning phase has been carried out, the routing of frames between and within each VLAN can then start. For example, assuming a PC with a MAC address of 52 sends a frame to, say, a server with a MAC address of 57, since the VLAN identifier is the same for both the PC and the server, hub BH1 carries out the routing of the frame directly without any further transmissions. If now a PC with a MAC address of 58 and a VLAN identifier of G sends a frame to, say, a server with a MAC address of 67, BH1 first forwards the frame to BH0. The latter then consults its routing table and, after determining that the server with a MAC address of 67 is also a member of VLAN G, it forwards the frame out on port 2.

In this way, frames are routed not just on their MAC address but also on their VLAN identifier and, because of this, the load on the total network is significantly reduced by including the VLAN identifier. Also, if a frame has a

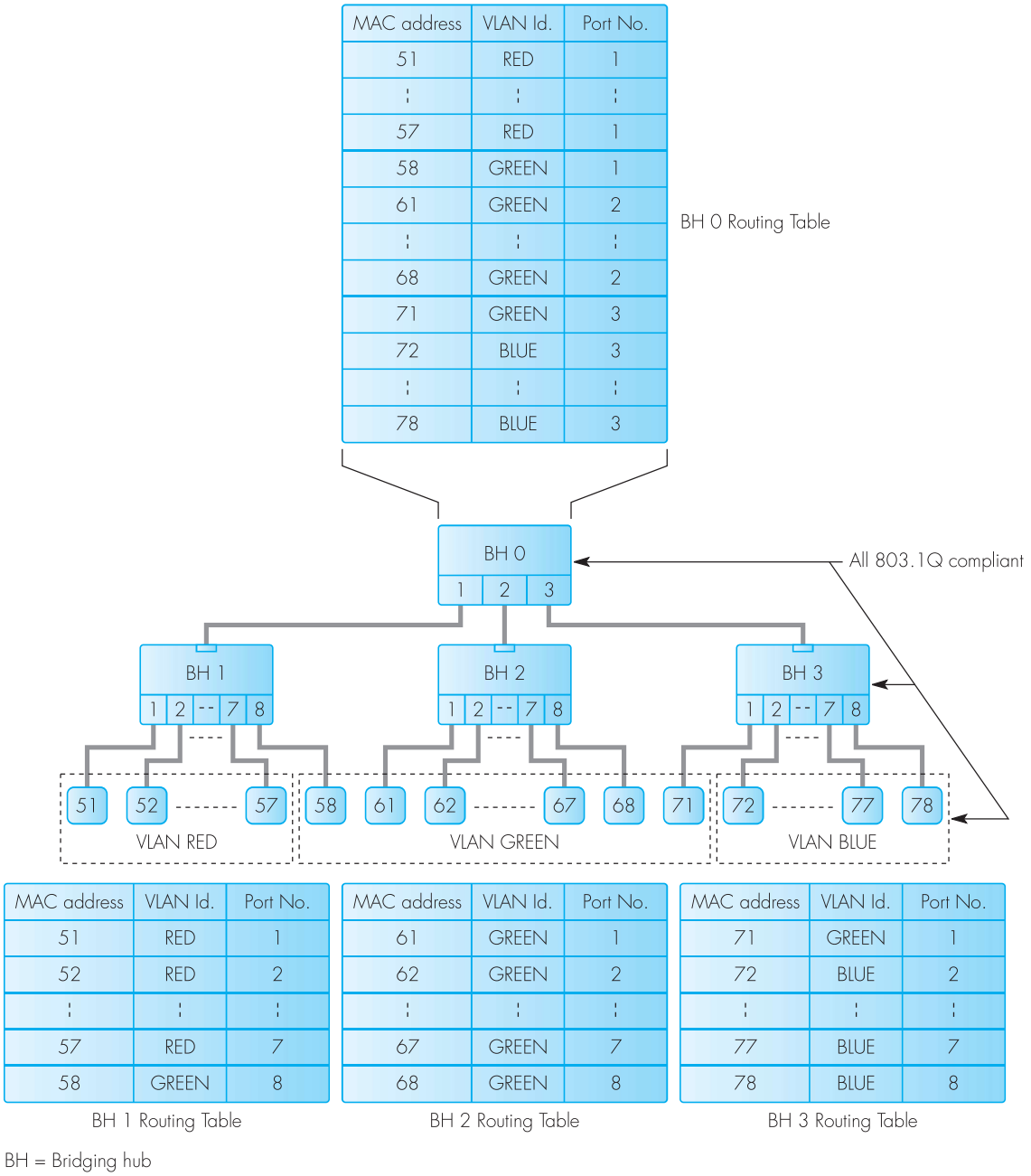


Figure 3.16 Example VLAN configuration and routing tables when all devices are IEEE802.1Q compliant.

VLAN identifier that is different from that in the routing table, then the frame is discarded, so improving the security of the network. Note also that the same procedure holds for both broadcast and multicast frames

Interworking

Problems arise when interworking between IEEE802.1Q compliant stations/hubs/switches and one or more legacy devices because the latter have no knowledge of VLAN identifiers. For example, assume the four hubs in the example network configuration shown in Figure 3.16 are all compliant but all the stations have legacy NICs. In this case the frames received at each bridge port have no VLAN identifier and hence the routing of frames can only be based on their MAC address.

To overcome this problem, therefore, the routing table in bridging hubs BH1, 2 and 3 have to be entered by network management software within each hub. Once this has been carried out, each lower-level bridge, on receipt of a legacy frame at one of its ports, reformats the frame into the new format with the related VLAN identifier obtained from the routing table.

Once this has been done, bridge BH0 is unaware of the legacy frames and hence learns and routes frames as in the previous example. Then, on receipt of a (reformatted) frame by a lower-level bridge, the bridge first determines the required output port number using the destination MAC address and VLAN identifier from its routing table. It then proceeds to reformat the frame into the legacy format and forwards this to the station that is attached to the port number obtained from its routing table. The same procedure is followed if the upper-level hub was a switch but, in this case, the routing table of the switch must be entered by network management.

Clearly, there are a number of other potential problems that can arise when interworking between the old and new frame formats and compliant and non-compliant devices. However, as PCs and servers are upgraded, the newer devices will now have an IEEE802.1Q compliant NIC. The same applies for hubs and switches. Indeed, the most recent devices such as gigabit switching hubs are already compliant and, after a period of time, VLAN will become the standard mode of working in many application domains.

3.6 LAN protocols

As we have learnt, there is a range of different types of Ethernet LAN, each of which uses a different transmission mode and transmission medium. In the context of our reference model, however, these differences only manifest themselves at the physical layer. The link control sublayer – known as the **logical link control** (**LLC**) **sublayer** in the context of LANs – then offers a standard link layer service to the network layer above it.

As we have seen, the various protocol standards associated with LANs are all part of the **IEEE 802 series**. The framework used for defining the various standards is shown in Figure 3.17(a) and a selection of the protocols that we have described in the previous sections are listed in Figure 3.17(b).

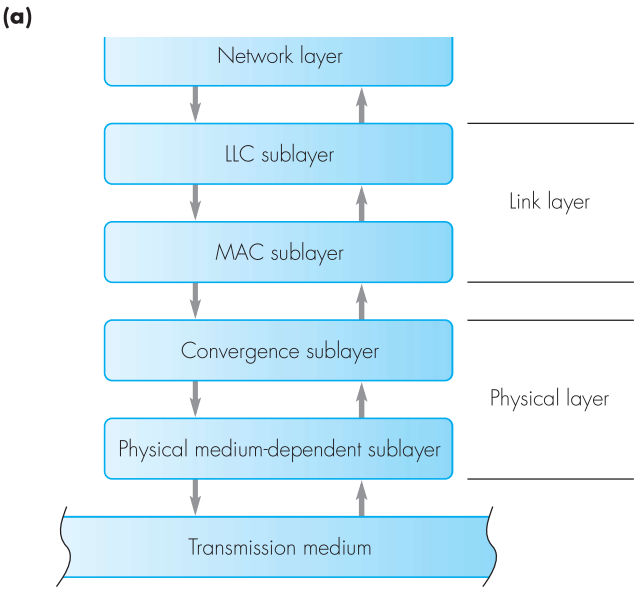


Figure 3.17 LAN protocols: (a) protocol framework; (b) examples.

3.6.1 Physical layer

To cater for the different types of media and transmission bit rates, the physical layer has been divided into two sublayers: the **physical medium-dependent** (**PMD**) **sublayer** and the (physical) **convergence sublayer** (**CS**). To facilitate the use of different media types, a **media-independent interface** (**MMI**) has been defined for use between the convergence and PMD sublayers. The role of the convergence sublayer is then to make the use of different media types and bit rates transparent to the MAC sublayer.

The set of signals associated with both interfaces for the different types of Ethernet (802.3) are shown in Figure 3.18.

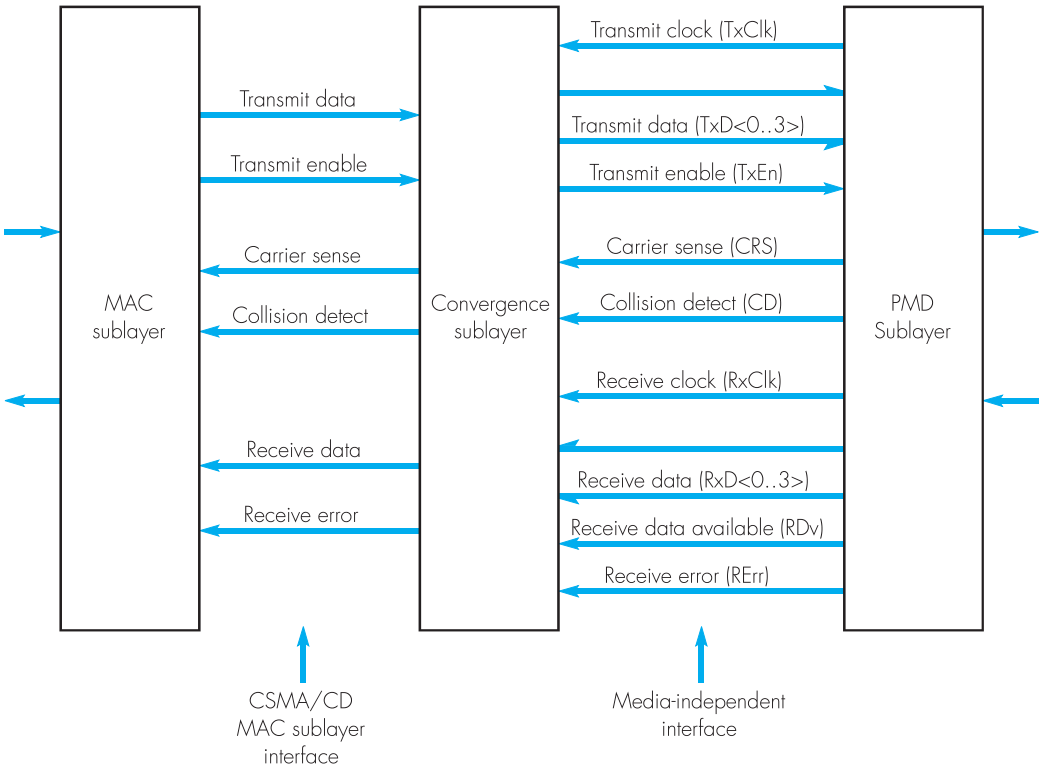


Figure 3.18 Fast Ethernet media-independent interface.

As we explained in Section 1.4.1, at bit rates in excess of 10 Mbps it is not possible to use clock encoding – for example Manchester – because the resulting high line signal transition (baud) rate would violate the limit set for use over UTP cable. Instead, bit encoding and a DPLL are used and, to ensure the transmitted signal has sufficient transitions within it, the encoding schemes use one or more groups of 4 bits in each transmitted symbol. For example, one 4-bit group when 4B5B coding is used (100BaseSX), and two 4-bit groups when 8B/10B coding is used (1000BaseSX). Hence to accommodate this, all transfers over the MII are in 4-bit nibbles. The other control lines are concerned with the reliable transfer of these nibbles over the interface. The major functions of the convergence sublayer, therefore, are to convert the transmit and receive serial bitstream at the MAC sublayer interface into and from 4-bit nibbles for transfer across the MII, and, when half-duplex transmission is being used, to relay the carrier sense and collision detect signals generated by the PMD sublayer to the MAC sublayer.

3.6.2 MAC sublayer

Irrespective of the mode of operation of the MAC sublayer, a standard set of user service primitives is defined for use by the LLC sublayer. These are:

MA_UNITDATA.request
MA_UNITDATA.indication
MA_UNITDATA.confirm

A time sequence diagram illustrating their use is shown in Figure 3.19. For a CSMA/CD LAN, the confirm primitive indicates that the block of data associated with the request has been successfully (or not) transmitted.

Each service primitive has parameters associated with it. The MA_UNITDATA.request primitive includes: the required destination MAC address (this may be an individual, group, or broadcast address) and a service data unit (containing the data to be transferred – that is, the LLC PDU).

The MA_UNITDATA.confirm primitive includes a parameter that specifies the success or failure of the associated MA_UNITDATA.request primitive. However, as we show in the figure, the confirm primitive is not generated as a result of a response from the remote LLC sublayer but rather by the local MAC entity. If the parameter indicates success, this simply shows that the MAC protocol entity (layer) was successful in transmitting the service data unit onto the network medium. If unsuccessful, the parameter indicates why the transmission attempt failed. As an example, with a CSMA/CD bus, "excessive collisions" may be a typical failure parameter.

3.6.3 LLC sublayer

The LLC protocol is based on the high-level data link control (HDLC) protocol that we described earlier in Section 1.4.9. It therefore supports both a connectionless (best-effort) and a connection-oriented (reliable) mode. In almost all LAN networks, however, only the connectionless mode is used. Hence, since this operational mode adds only minimal functionality, when the older Ethernet MAC standard is being used – see the discussion on frame formats in Section 3.2.3 – the LLC sublayer is often not present. Instead, the network layer – for example the Internet protocol (IP) – uses the services provided by the MAC sublayer directly.

When the newer IEEE802.3 MAC standard is being used, the LLC sublayer is present. However, since it operates in the connectionless mode, the only user service primitive is L_DATA.request and all data is transferred in an unnumbered information (UI) frame. The interactions between the LLC and MAC sublayers are as shown in Figure 3.20.

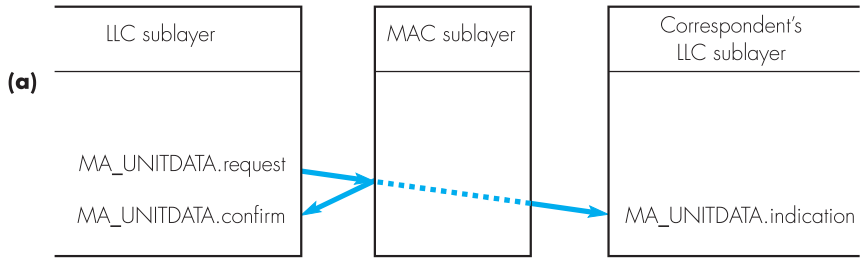


Figure 3.19 MAC user service primitives for CSMA/CD.

The L_DATA.request primitive has parameters associated with it. These are: a specification of the source (local) and destination (remote) addresses and the user data (service data unit). The latter is the network layer protocol data unit (NPDU). The source and destination addresses are each a concatenation of the MAC sublayer address of the station and an additional service access point (SAP) address. In theory, this can be used for interlayer routing purposes within the protocol stack of the station. In applications such as the Internet, however, this feature is not used and both the destination SAP (DSAP) and the source SAP (SSAP) are set to AA (hex). In addition, two further fields are added. Collectively, the two fields form what is called the **subnet access protocol** (**SNAP**) header. The first is a 3-byte field known as the *organisation* (*org*) *code* – which, with the Internet, all three bytes are set to zero – and a two-byte *type* field. This is the same as that used in the original Ethernet standard and indicates the network layer protocol that created the NPDU.

A more detailed illustration of the interactions between the LLC and MAC sublayers is shown in Figure 3.21. The LLC sublayer reads the destination and source LLC service access point addresses (DSAP and SSAP) – from the two address parameters in the **event control block** (**ECB**) associated with the L_DATA.request service primitive – and inserts these at the head of an LLC PDU. It then adds an 8-bit *control* (*CTL*) field – set to 03 (hex) to indicate it is an unnumbered information (UI) frame – the 3-byte org code, the type field, followed by the network layer protocol data unit in the user data field. The resulting LLC PDU is then passed to the MAC sublayer as the user data parameter of an MA_UNITDATA.request primitive in a MAC ECB. Other parameters include the MAC sublayer destination and source addresses (DA and SA) and the number of bytes (length indicator) in the user data field.

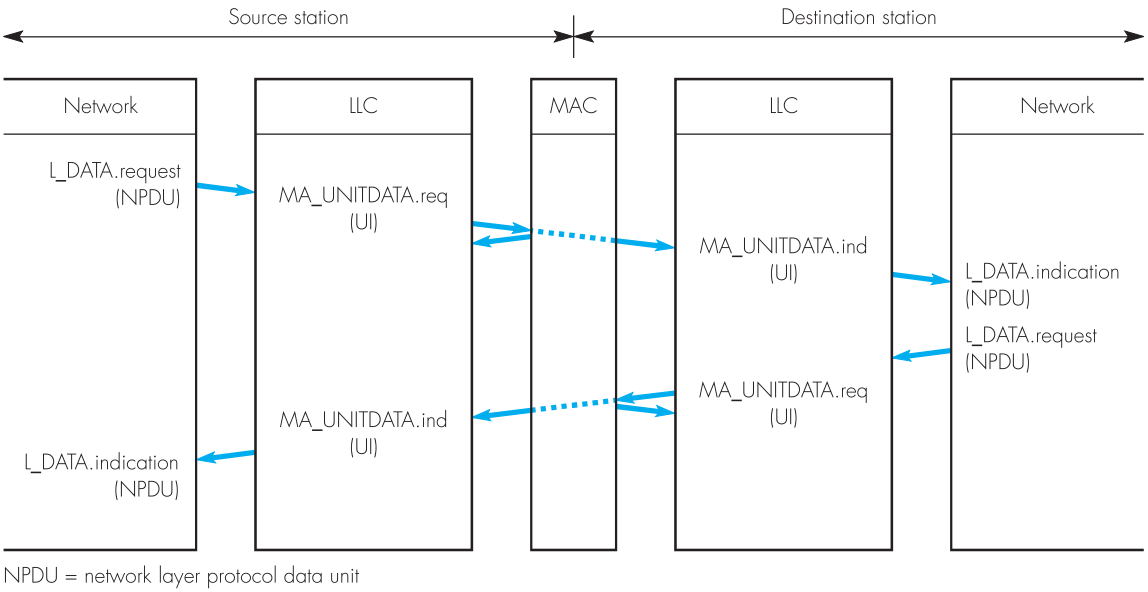


Figure 3.20 LLC/MAC sublayer interactions.

On receipt of the request, the MAC protocol entity creates a frame ready for transmission on the link. In the case of a CSMA/CD network, it creates a frame containing the preamble and SFD fields, the DA and SA fields, an I-field, and the computed FCS field. The complete frame is then transmitted bit serially onto the cable medium using the CSMA/CD MAC method.

A similar procedure is followed in the NIC of the destination station except that the corresponding fields in each PDU are read and interpreted by each layer. The user data field in each PDU is then passed up to the layer/sublayer immediately above together with the appropriate address parameters.

3.6.4 Network layer

The most popular protocol stack used within LANs is **Novell NetWare**. Hence it is common for all the stations connected to a site LAN to communicate using this stack. The network layer protocol associated with this is a connectionless protocol known as the **internet packet exchange** (**IPX**) protocol and, as its name implies, it can route and relay packets over the total LAN. There is no LLC sublayer associated with the stack and hence the IPX protocol communicates directly with the MAC sublayer.

The protocol stack used within the Internet is TCP/IP and the network layer protocol associated with this stack is a connectionless protocol known as the Internet protocol (IP). Hence, as we showed earlier in Figure 8.9, it is common for server machines such as e-mail servers that need to communicate

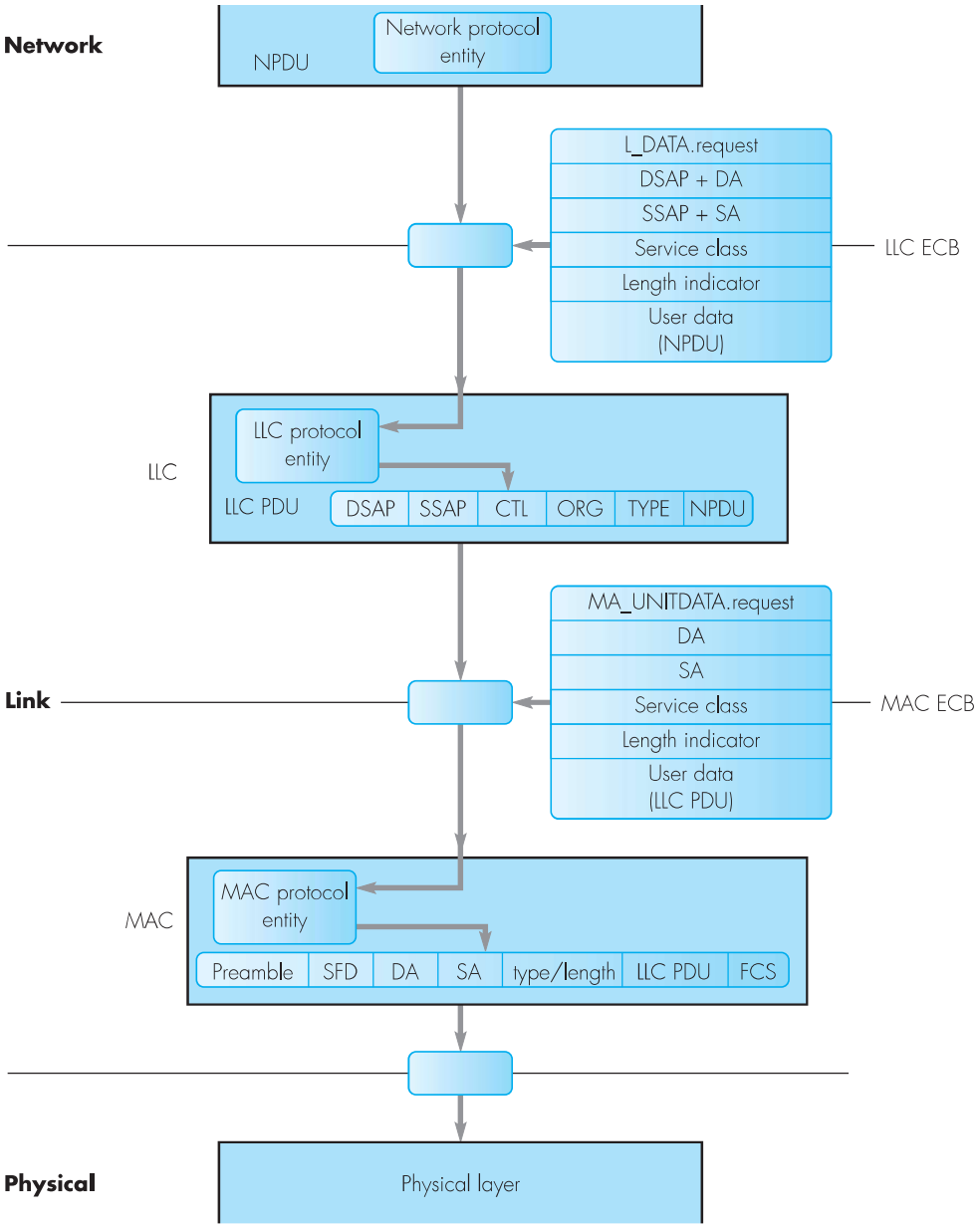


Figure 3.21 Interlayer primitives and parameters.