~DRAFT~

CHAPTER 5: Link layer

Contents

1 Context of Chapter

The data-link, or simply, link layer is the second layer in the OSI model and performs the final stage of encapsulation before passing down packets to the physical layer to be converted into bits. Of the layers it is the lowest that is fully implemented and controlled inside routers, and as such is where hardware meets software. It takes a network layer datagram and encapsulates it into a link layer frame by adding a header with source and destination addresses and a footer (or trailer) with a value for error checking, and prepares data for the physical layer. The error checking mechanism is for errors occurring in the physical layer. Figure 1 shows the encapsulation into a link layer frame relative to the encapsulation at the network and transport layers. A data-link protocol specifies the structure of the frame, as well as a channel access protocol that specifies the rules by which a frame is transmitted onto the link. We will see that some features of the link layer overlap with the network layer and their interaction enables a number of different functions.

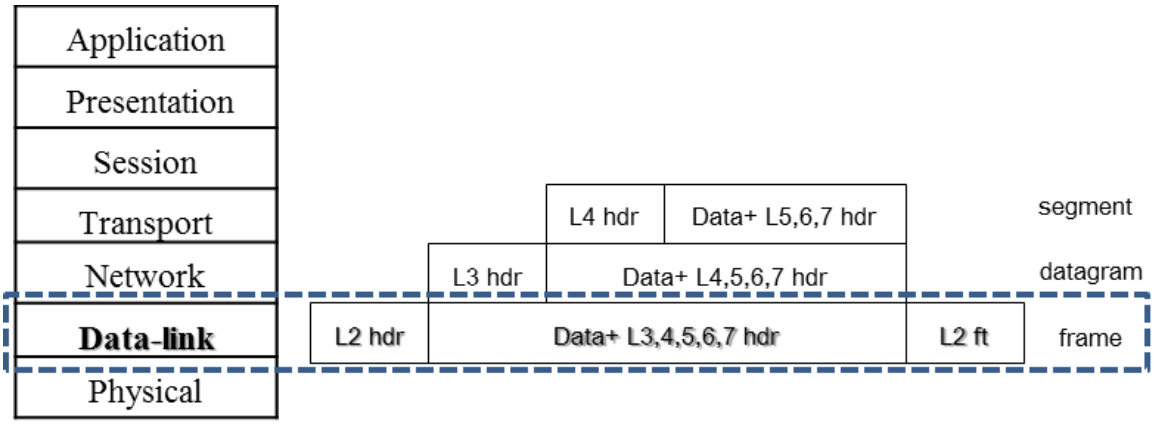

FIGURE 1: DATA ENCAPSULATION AT THE LINK LAYER

2 Introduction

To introduce the data-link layer, we must define what a link is. A *link* is a communication channel that connects adjacent *nodes* along the communication path from source to destination host. In order to move a datagram from source to destination node (any host, router, switch or WiFi access point that runs a link layer protocol), datagrams must be moved over each of the individual *links* along the path by framing them into data-link layer frames. This is the main purpose of the link layer protocol in a device, operating within a local area network (LAN) or wide area network (WAN). In order to move datagrams effectively over physical links, the link layer protocol performs the functions of local

delivery to adjacent nodes, addressing, and media arbitration or access control. The link layer also prepares a packet for the physical layer and may perform detection and correction of errors occurring in the physical layer. In many standards, the functionality is divided into two sublayers: the Link Layer Control (LLC) and Medium Access Control (MAC) sublayers.

Link layer protocols are typically implemented in a router's network interface card(s) (NIC), also known as the network adapter. The NIC contains the electronic circuitry required to communicate using a wired connection (e.g. Ethernet) or a wireless connection (e.g. WiFi). The software components of the link layer implement higher level link layer functionality such as assembling link layer addressing information and activating the controller hardware. On the receiving side, link layer software responds to controller event notifications such as frame arrival, handles error conditions and passes a datagram up to the network layer.

3 Objectives of Chapter/module

By the end of this Chapter you should be able to:

- Note the services provided and functions performed by the data-link layer
- Name and be able to explain the functioning of different error detection techniques in the link layer
- Name and describe some error correction techniques
- Explain link layer control in wired networks
- Know functions of the MAC sublayer
- Name medium access features in wireless networks such as WiFi and explain how three main classes of medium access work, their similarities and differences
- Explain briefly how ALOHA works and its evolution to slotted ALOHA
- Know what CSMA is and be able to name two different variants of CSMA and their application areas
- Understand aspects of Ethernet
- Be able to name different Ethernet media and identify the speed and cable technology from a media type name
- Be able to explain how ARP is used in Ethernet networks in different topologies within subnetworks, and what is required for interconnecting subnets.
- Know how Virtual LANs work and name the different kinds
- Understand the motivation and application for MPLS

4 Services of the link layer

A link layer protocol may offer the following services:

- Framing – encapsulation of datagrams into link layer frames
- Link access – performed by the medium access control (MAC) protocol, which specifies the rules by which a frame gains access to a link when other frames are competing for the same medium
- Reliable delivery – not present in all link layer protocols, a reliable delivery service guarantees to move each network layer datagram across the link without error.
- Error detection and correction – physical layer bit errors are introduced by various factors such as signal attenuation and electromagnetic noise. The link layer protocol in a receiving node performs mathematical calculations to detect such errors.

5 Error detection

For a receiver to detect transmission errors the sender must add redundant data (extra bits) to the outgoing frame as an error detection code. When the destination node receives a frame with an error detection code, the link layer protocol at the receiving end recalculates the error code. If the received error detection code matches the computed error detection code, the frame is considered valid. Link layer error detection mechanisms may not eliminate all errors but this is a first line of defence. We now detail some kinds of error detection.

a. Parity check

A single parity bit is the simplest form of error detection, which is able to detect a single bit error in a frame, or detect an error if an odd number of bit flips occurred. A parity bit is a redundant bit added to the data payload. A scheme may employ either even or odd parity, in some cases selectable by the network configurator. In an even parity scheme, the sender simply includes one additional bit, choosing its value (0 or 1) such that the total number of 1s in the original data bits plus a parity bit is even. For odd parity schemes, the parity bit value is chosen such that there is an odd number of 1s. The receiver then counts the number of 1s in the received bits. If an odd number of 1-valued bits are found with an even parity scheme, the receiver knows that some odd number of bit errors have occurred. If an even number of 1-valued bits is calculated in an even parity check scheme, an even number of bit errors may have occurred but the parity check is unable to detect this. The converse is

true for odd parity schemes. This feature may be a particular problem for *bursty* errors i.e. errors that are clustered together so the chances of multiple bits being corrupted within a single packet are high. Figure 2 illustrates the parity check of a packet in an even parity scheme, where two bit errors occurred within the channel. At the receiver end 0 bit errors were detected because the parity check still found an even bit-sum.

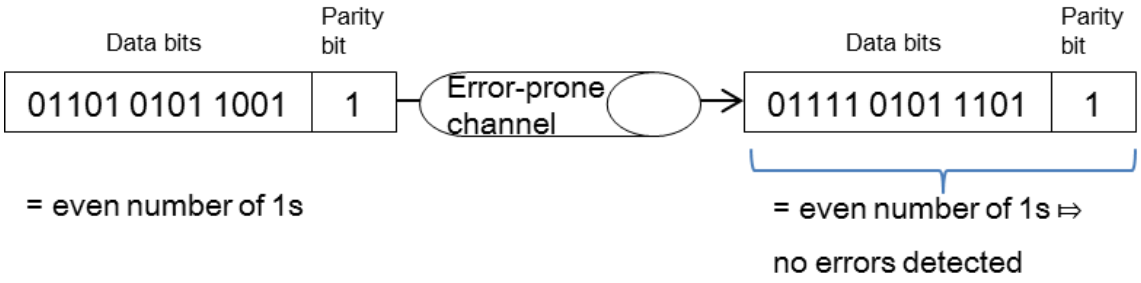

FIGURE 2: EVEN BIT PARITY CHECK AT RECEIVER END

The single bit parity check may be extended to a 2-dimensional version, which is somewhat more reliable but it is not common in practical systems. Instead, higher level error detection may be employed instead.

b. Cyclic Redundancy Check

Cyclic Redundancy Check (CRC) codes are also known as polynomial codes, the reason for which will become clear shortly. They are one kind of *checksumming* error detection. The idea is that a bit frame D consisting of k data bits is represented as coefficients of a polynomial expansion with exponents 0 to $k-1$. As an example a bit string such as 11011 is represented as the polynomial $x^4 + x^3 + x + 1$. There is no x^2 term because the bit in that position has value 0. To produce the CRC code, the XOR (logical exclusive OR) is performed on the bits. The sender and receiver must agree beforehand on a *generator polynomial* $G(x)$, which must be smaller than the number of bits in the frame. The checksum can be computed using the generator polynomial. If it divides without a remainder, it is concluded that no error has occurred.

This CRC method can detect single bit errors, two isolated single bit errors, odd number bit errors (provided the polynomial contains $x + 1$) and burst errors of length less than or equal to the degree of the polynomial. Other burst errors may be detected with varying probability.

6 Error Correction

Error correction may be forward error correction (FEC) where the receiver both detects and corrects errors by running an error-correcting code on board, and backwards error correction where the receiver instead requests the sender to retransmit the data. To correct errors by FEC, a receiver must identify which bit has been corrupted during transmission and reconstruct the original information accurately. FEC is used widely in wireless environments that can tolerate occasional errors but where retransmission may cause unacceptable delays, while retransmissions are more common when the connection is fast and can tolerate retransmissions, such as in fibre optic connections.

Of fundamental importance to error correction is the *Shannon capacity* of a channel (transmission medium, which can be wired or wireless). The Shannon capacity or Shannon limit is the maximum information rate at which reliable communication is possible over a channel depending on its SINR, which is the ratio of wanted signal power (S) to the power of interference (I) + noise (N). The SINR is related to the probability of bit errors. Shannon's theorem states that capacity

$$C = \log_2(1 + SINR) \quad (1)$$

for a channel with additive white Gaussian noise (that is, it is assumed that interference and noise follow a Gaussian or normal distribution and may simply be added on top of the desired signal). The code rate must be smaller than the channel capacity to ensure that reliable communication is possible. The theorem says that error correcting codes are possible or can exist for a given channel capacity but does not postulate the actual error correcting codes.

Hamming code is a single bit error detection mechanism that uses bit redundancy. The bits are arranged such that different incorrect bits produce different error results and the corrupt bit can be identified and its value corrected. Consider a frame consisting of k data bits and r check bits, with the resulting $n = k + r$ bit unit being called a *codeword*. The Hamming distance is the number of bits by which two codewords differ. To detect b-bit errors, requires $b + 1$ Hamming distance between a pair of codewords. We can now see that codewords with single parity have a minimum distance of two (calculated as 1 parity bit + 1), so in order to correct b-bit errors requires a distance of $2b + 1$. r redundant bits can provide 2^r combinations of information. In a $k + r$ bit codeword, there is a possibility that the r bits themselves may get corrupted. So to correct single bit errors, the number of r bits used must inform about $k + r$ bit locations plus a bit indicating no error information, i.e. $k + r + 1$. This implies the inequality

$$2^r \geq k + r + 1 \quad (2)$$

must hold, which can be used to calculate the number of required redundant bits.

7 MAC

The MAC sublayer controls how a node gains access to the transmission medium and permission to transmit a frame. The sublayer is responsible for moving packets from one network interface card to another across the shared channel. For point-to-point links that have a single sender on one end of the link and a single receiver at the other end of the link, the link access protocol is simple (or non-existent) - the sender can send a frame whenever the link is idle. The more interesting case is when multiple nodes share a single broadcast link - the so-called multiple access problem. A MAC broadcast protocol should ideally be decentralised so that there is no single node acting as a master controlling transmission opportunity and presenting a single point of failure. However, there are many useful centralised MAC protocols. A MAC protocol should also be simple and easy to implement so that it is inexpensive and does not consume much computational power. If a single node alone has a data rate of R bits/s and there are M nodes sharing a channel, the average throughput per node is R/M bits/s.

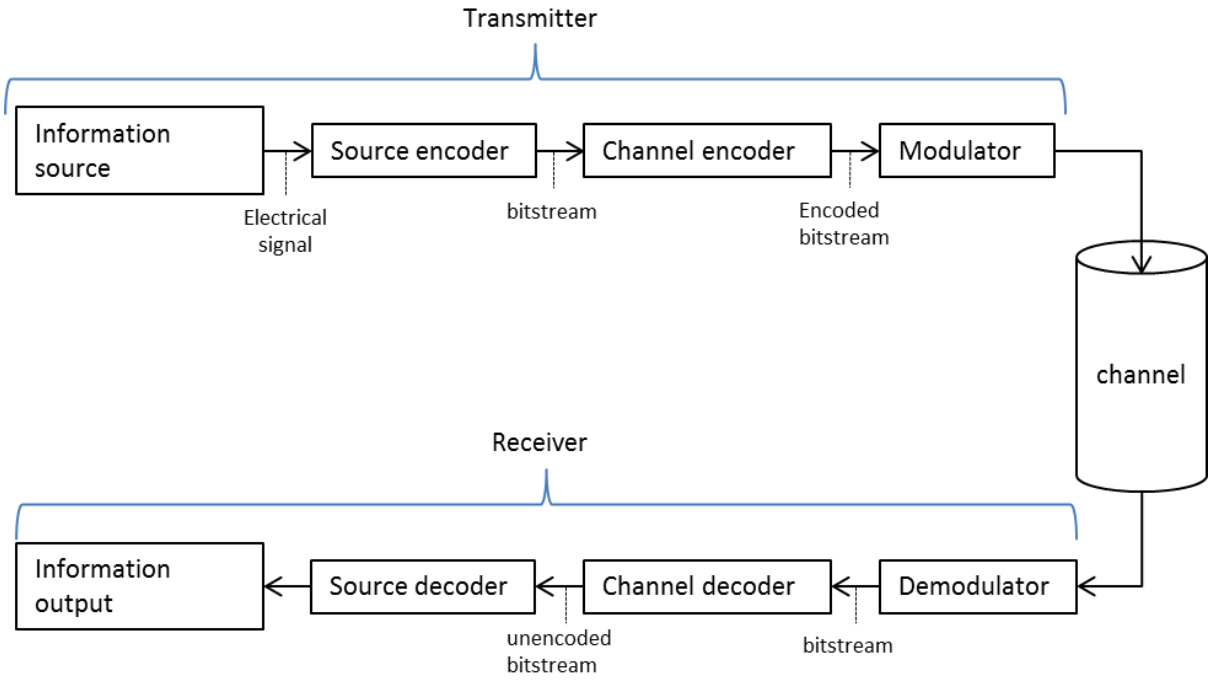

FIGURE 3: THE GENERALISED COMMUNICATION SYSTEM

At this point it may be useful to remind ourselves of the big picture of a communications system, and the various steps required in getting a data stream from sender to a receiver. The blocks in the top half of Figure 3 are inside the sender and the blocks in the bottom half represent elements in a receiver node.

The *source encoder* converts the electrical signal from the information source or transducer to a binary stream of bits. The *channel encoder* must then convert the bit stream to encoded information in the form of a codeword. This is where redundant bits such as parity and CRC bits for error detection and correction, discussed in Sections 5 and 6, are added. The modulator takes the codeword and

converts it to a continuous electrical signal by *modulating* it onto a *carrier signal* so that it can be transmitted through the channel. A carrier signal has a specific centre frequency and bandwidth around the centre frequency and the majority of the signal power should lie within the channel bandwidth. Amplitude modulation is where the amplitude of the signal power of the carrier is varied according to the bit sequence to be transmitted. For example, a 0 could be transmitted at -A amplitude and a 1 at A amplitude. In frequency modulation such as frequency shift keying (FSK) and phase shift keying (PSK) the signal is varied by frequency. A 0 could be transmitted at centre frequency + offset and a 1 could be transmitted at centre frequency – offset. Related to modulation is multiplexing, an important part of the link layer. Multiplexing in this context refers to how different signals may be modulated onto a common carrier signal or transmission medium. Understanding these elements of a communication system is important to understand the functioning of MAC protocols as these are at the interface of the physical layer (the channel) and the data link layer.

According to IEEE 802.3-2002 section 4.1.4, the functions required of a MAC are:

- receive/transmit normal frames
- retransmission and back-off in half-duplex connections
- append/check FCS (frame check sequence)
- inter-frame gap enforcement
- discard corrupted frames
- prepend or remove preamble, start frame delimiter, and padding bits
- half-duplex compatibility: append/remove MAC address

a. Fixed assigned MAC

Frequency Division Multiple Access

Frequency Division Multiple Access (FDMA) use frequency division multiplexing. That is, the total available bandwidth is partitioned into separate channels of equal channel bandwidth and each data stream is assigned a certain channel for transmission at a given time. An example of FDMA systems were the first-generation (1G) mobile phone systems, where each phone call was assigned to a specific uplink frequency channel, and another downlink frequency channel. A related technique is wavelength division multiple access (WDMA), based on wavelength division multiplexing (WDM). In WDMA different data streams get assigned to different coloured fibres in an optical fibre bundle.

Time Division Multiple Access

In Time Division Multiple Access (TDMA) different data streams are assigned different constant-length time slots. 2G cellular systems were based on a combination of TDMA and FDMA where each frequency channel was divided into eight time slots, seven used for phone calls and one for signalling data.

Orthogonal Frequency Division Multiple Access

In Orthogonal Frequency Division Multiple Access (OFDMA), using principles of Orthogonal Frequency Division Multiplexing (OFDM), the signal band is divided into many subcarriers that are orthogonal to each another and subsets of subcarriers are assigned to individual users. This orthogonality is achieved by having the carrier spacing equal to the reciprocal of the symbol (codeword) period. The data to be transmitted on an OFDM signal is spread across the carriers of the signal, each carrier taking part of the payload and thus reducing the data rate taken by each carrier. Each data stream is then modulated onto a different subcarrier/time slot block. OFDM is actually a two-dimensional modulation scheme since it modulates over both frequency and time. OFDMA is used for the downlink and OFDM for the uplink in LTE networks. We will discuss OFDM in more details in the chapter on wireless networks.

Code Division Multiple Access (CDMA)

Frequency hopping spread spectrum (FHSS) is a type of CDMA whereby the carrier is rapidly switched among many different frequencies in a pseudo-random order that is known to both receiver and transmitter, enabling the receiver successfully to distinguish and decode different data streams. Direct sequence spread spectrum (DSSS) is a code division multiplexing scheme where the message signal is used to modulate a bit sequence known as a pseudo-noise code. The pseudo-noise code consists of a radio pulse that is much shorter in duration than the original message signal and so has a larger bandwidth. The message signal scrambles and spreads the pieces of data resulting in a bandwidth size nearly identical to that of the pseudo-noise sequence. Both DSSS and FHSS are part of the IEEE 802.11 WiFi standard.

b. Demand assigned MAC

Demand assigned MAC protocols assign communication channels based on the capacity required by nodes on the network (the demand). Channels are allocated to nodes for a specified amount of time, which may vary from a fixed-time slot to the time it takes to transmit a data packet.

Polling

A prevalent centralised demand assignment scheme is *polling*. In this method a central device denoted the master controls the channel access to the slave devices by checking with each slave in a specific predetermined order whether it has data to transmit. If the polled node (the slave) has data to transmit it informs the controller, which then allocates the (or a) channel to the ready node for a time slot. The slave node then uses the full data rate to transmit. If the polled slave node has no data to transmit it declines the controller's request and the controller simply moves on to query the next node on the network. Polling procedures may be configured so that higher priority nodes are polled more often or are always polled first.

The main disadvantage of polling is the considerable overhead that is required by the large number of messages the master generates to query the slave nodes and also that nodes must wait their turn before transmitting even if the message is important (unless extra measures are employed to handle different priority classes). The latter can be remedied by implementing a hybrid system where transmission occurs mainly by polling but some high priority devices may interrupt the process in an unsolicited manner when there is a change of state or it has a message to transmit. This kind of operation is common in telemetry systems (Distributed Network Protocol v.3 or IEC 61850-7-2 based networks). There is also a polled system defined in IEEE 802.11, called Point Coordination Function.

Reservation-based schemes

In reservation-based medium access, time slots (called mini slots because they are usually smaller than data time slots) are set aside for carrying reservation messages. When a station has data to send, it requests a data slot by sending a reservation message to the master in a reservation mini slot. On receiving the reservation request, the master computes a transmission schedule and announces the schedule to the slaves.

Polling or solicited communication is a taking turn protocol. Another taking turns protocol is the token-passing. Here there is no master node but a certain special frame known as a token is exchanged among the nodes in some fixed order. A node holding a token is allowed to transmit data up to a certain limit and must pass on the token once it has no more data to send or has reached the limit for the current round. If a glitch causes node not to forward the token some recovery procedure must be invoked to get the token back in circulation. IEEE 802.5's token-ring protocol is an example. IEEE 802.4's token bus protocol implements token-passing over a "virtual ring" on a coaxial cable.

c. Random Access MAC

In random access MAC protocols nodes attempt to transmit data whenever they have data to send. Each node sends at the full rate of the channel. If nodes experience a collision, they will all back off

and wait a pseudo-random time period before reattempting transmission. Since each node is likely to choose a different delay period collisions can be overcome. However, a few or several attempts may be necessary before all nodes get a chance to transmit their data. Some examples are now detailed.

ALOHA

The ALOHA protocol is one of the oldest MAC protocols. The principles of ALOHA are 1) A node sends data when it has data to send without checking first, 2) If a node receives any data from another station during that time, there has been a message collision. All transmitting nodes must try resending "later". Nodes resend after waiting for the frame time duration or immediately with a certain probability.

Note that the first step implies that Pure ALOHA does not check whether the channel is busy before transmitting. Since collisions can occur and data may have to be sent again, ALOHA cannot use 100% of the capacity of the communications channel. How long a station waits until it transmits, and the likelihood of collisions occurring are interrelated, and both affect how efficiently the channel can be used.

Slotted ALOHA

Slotted ALOHA was the next iteration of ALOHA, introducing discrete timeslots and an increase in the maximum throughput. A node can only transmit at the beginning of a timeslot, thus reducing collisions while frames are in transit. Only transmission attempts within 1 frame time and not 2 consecutive frame times need to be considered, since collisions can only occur during each time slot

Carrier Sense Multiple Access (CSMA)

The CSMA protocol requires that nodes verify that there is no other traffic currently active on a shared transmission medium before transmitting, i.e. nodes *sense* to verify there are no *carrier* signals currently on the medium before transmitting. CSMA with collision detection (CSMA/CD) adds the feature of terminating transmission as soon as a collision is detected to basic CSMA, shortening the time required before a retry can be attempted. CSMA with collision avoidance (CSMA/CA) requires that transmission be deferred for a random interval if the medium is sensed busy before transmission. The IEEE 802.11 WLAN standard's Distributed Coordination Function (the counterpoint to Point Coordination Function) is based on CSMA/CA. Many other variations on CSMA exist.

8 Ethernet and switched Local Area Networks

In general, each NIC of a host or router has a link layer address used for link layer protocols, whether it is an Ethernet interface, or WiFi card or other. Link layer *switches*, however, do *not* have link layer

addresses associated with their interfaces that connect to hosts and routers because a link layer switch must carry datagrams between hosts and routers transparently, that is, without the host or router having to address the frame explicitly to the intervening switch. This feature is an important feature for implementing certain protocols, as we will see.

a. Link layer addressing

A link layer address is variously called a LAN address, a physical address, or a MAC address. MAC addresses are 6-byte (48-bit) addresses typically expressed in hexadecimal notation, used to get link layer frames from one interface to another physically connected interface on the same network. When an adapter wants to send a frame to some destination adapter, the sending adapter inserts the destination adapter's MAC address into the frame and then sends the frame into the LAN. The receiving adapter then checks whether the destination MAC address in the frame matches its own MAC address. If it does match, the adapter extracts the enclosed datagram and passes the datagram up the protocol stack. If not, the adapter discards the frame without passing the enclosed datagram up. If the frame is to be broadcast to all adapters on the network, the MAC broadcast address is used. For LANs with 6-byte addresses (e.g. Ethernet and 802.11), the broadcast address is a string of 48 consecutive 1s (that is, FF-FF-FF-FF-FF-FF in hexadecimal notation). MAC address allocation is managed by the IEEE such that every single network adapter has a different MAC address. In addition to being unique, MAC addresses are generally permanent although it is now possible to change an adapter's MAC address via software.

Address Resolution Protocol

Address Resolution Protocol (ARP) maps IP addresses of hosts and router interfaces on the same subnet to MAC addresses (e.g. Ethernet addresses). As such, it sits at the interface of the network and link layers. ARP client and server processes operate on all computers using IP over Ethernet, typically implemented as part of the NIC's driver. However, it is a generic protocol functioning by finding a mapping between ordered pairs (P, A) and arbitrary physical addresses, where P is a network-layer protocol and A is an address of this protocol P. At the time of ARP's development, different protocols were used in the network layer.

Each host or router has an ARP table with the IP addresses, MAC addresses and time to live of the current address mapping. ARP is self-learning and so nodes create their own ARP tables without requiring the intervention of a network administrator. An ARP request includes a valid mapping between the network layer protocol address and the MAC address of the request initiator, in addition

to the network address looked for, so one entry for the initiator is created in the ARP cache when the request is received.

ARP operates in request/response transmission pairs on the local network, as illustrated in Figure 4. The source node transmits a broadcast request message (to FF-FF-FF-FF-FF-FF) containing information about the intended destination's IP address. The destination then responds to the source with the hardware address of the destination. It does this unicast. This process is illustrated in Figure 4.

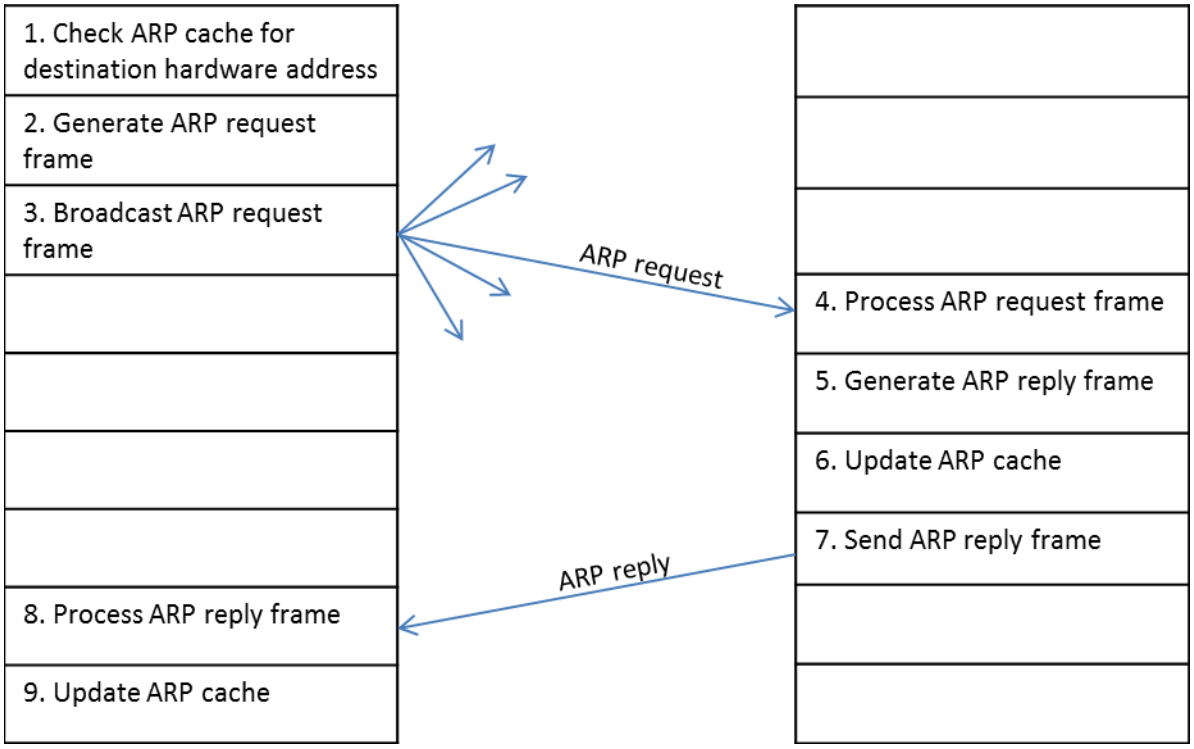

FIGURE 4: ARP TRANSACTION PROCESS

Hardware type		Protocol type
Hardware address length	Protocol address length	Operation code
Sender hardware address (octets 0-3)		
Sender hardware address (octets 4-5)		Sender IP address (octets 0-1)
Sender IP address (octets 2-3)		Target hardware address (octets 0-1)
Target hardware address (octets 2-5)		
Target IP address (octets 0-3)		

FIGURE 5: FORMAT OF AN ARP MESSAGE

There are four types of ARP messages in the ARP protocol, identified by four values in the "operation code" field of an ARP message, as seen in the second row of the second column in Figure 5. The types of message are:

1. ARP request
2. ARP reply
3. Return ARP request
4. Return ARP reply

The terms source and destination apply to the same devices throughout the transaction since they refer to the direction the payload data is travelling. However, there are two different protocol messages sent in ARP, one from the source to the destination and one from the destination to the source. For each ARP message, the sender is the one that is transmitting a message at the time and the target is the one receiving it. The identity of the sender and target change for each message but source and destination are constant according to the payload data to be sent (not the control message).

The format of an ARP message is shown in Figure 5, and request and response variations are shown in Figure 7. ARP Request and ARP Reply messages have the same packet format but with different operation field values. An ARP Request packet also differs from a subsequent reply by the missing hardware address of the destination, so that it is easy to create a reply to a request. When receiving a request packet, in which the desired station finds its network layer (e.g. IP) address, the following steps are completed:

- The MAC address of the network adapter is inserted in the field Hardware Destination Address.
- The two address fields for the sender and the destination are swapped.
- The Operation field takes value 2 to mark the message as an ARP Reply.
- Finally, the reply packet is sent.

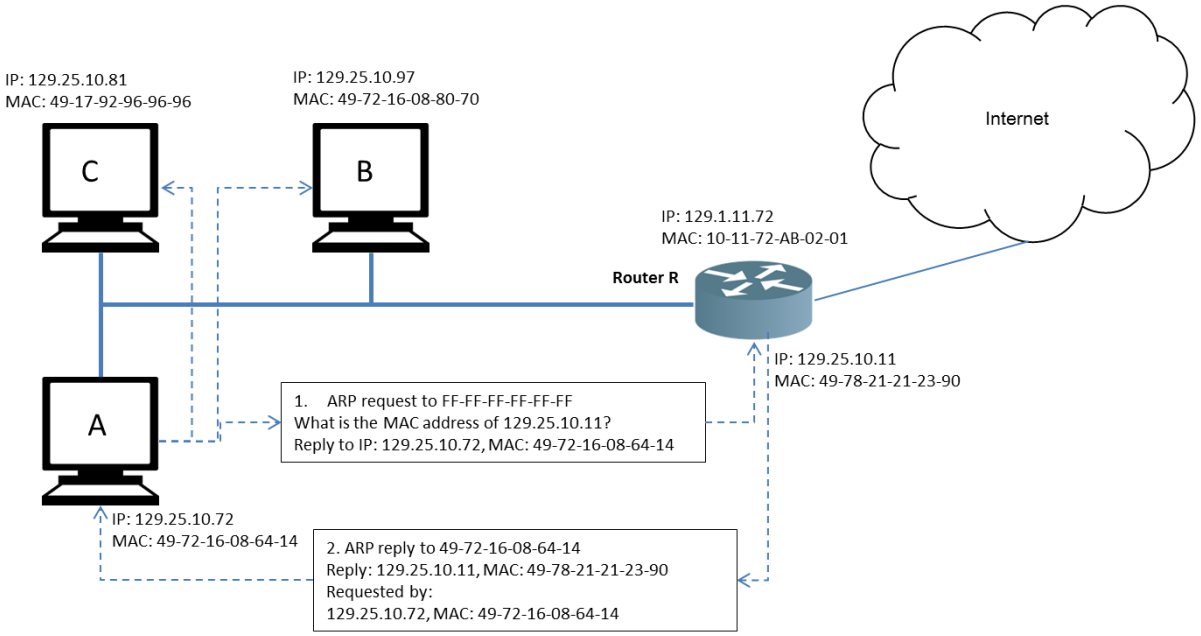

FIGURE 6: ILLUSTRATION OF AN ARP TRANSACTION IN A LAN

Hardware type		Protocol type
Hardware address length	Protocol address length	Operation code 1
Sender hardware address (octets 0-3) 49-72-16-08		
Sender hardware address (octets 4-5) 64-14		Sender IP address (octets 0-1) 129.25
Sender IP address (octets 2-3) 10.72		Target hardware address (octets 0-1) 00-00
Target hardware address (octets 2-5) 00-00-00-00		
Target IP address (octets 0-3) 129.25.10.11		

ARP request to FF-FF-FF-FF-FF-FF

Hardware type		Protocol type
Hardware address length	Protocol address length	Operation code 2
Sender hardware address (octets 0-3) 49-72-16-08		
Sender hardware address (octets 4-5) 64-14		Sender IP address (octets 0-1) 129.25
Sender IP address (octets 2-3) 10.72		Target hardware address (octets 0-1) 49-78
Target hardware address (octets 2-5) 21-21-23-90		
Target IP address (octets 0-3) 129.25.10.11		

ARP reply to 49-72-16-08-64-14

FIGURE 7: REQUEST AND REPONSE ARP MESSAGE VARIANTS

An example is illustrated in Figure 6, with the message formats of the request and reply shown below in Figure 7. In the example of Figure 6, computer A wishes to send a data packet to router R. In order to do this it needs the link layer address and the IP address to be able to tell its data-link layer for which computer the packet is destined. This is the purpose of request broadcast message 1. The intended destination router R recognises its IP address in the target (i.e. destination) address specified in the ARP message, as illustrated in Figure 7. This enables it to send a reply (message 2) to computer A which includes its MAC address. With this information computer A informs its data-link layer of the MAC address and can send the payload packet to the correct destination.

Gratuitous ARP is used when a node has selected an IP address and wishes to check that no other node is using the same IP address. It can also be used to force a common view of the node's IP address across the subnet (e.g. after the IP address has changed). This is common when an interface is first

configured, as the node attempts to clear out any stale caches that might be present on other hosts. To do this the node sends an ARP request for itself.

The *arp* command can be used to output the ARP table (ARP cache) of a computer or to manipulate the ARP table such as creating permanent entries or deleting entries. The option *-a* with the *arp* command is used to view the ARP table, while *-d* <*computer*> can be used to delete the entry for *computer* and the option *-s* <*computer*> <*layer-2-address*> is used to add *computer layer-2-address* manually to the cache.

Sending a datagram to destination outside of the subnet

ARP performs address resolution to enable data transmission within a single subnet but to send a packet across a router into another subnet an extra step is required. Proxy ARP is the name given when a node responds to an ARP request on behalf of another node. This is commonly used to redirect traffic sent to one IP address on another system, such as we now describe.

Suppose we have subnet 1, which has hosts with interfaces in the IP address range 111.111.111/24 and subnet 2 has the network address 222.222.222/24. Host 111.111.111.111 in subnet 1 must send a packet to host 222.222.222.222 on subnet 2 so it creates a datagram with the destination address of 222.222.222.222. The packet must pass through the first hop router on the path to the destination in subnet 2. Therefore host 111.111.111.111 must send the packet with the destination link layer address of the router adapter interface facing subnet 1, which it acquires by ARP since the router is attached to its subnet. Once the packet reaches the router the router uses its forwarding tables (constructed by the network layer) to forward the packet via its correct interface to subnet 2. The router can then use ARP in subnet 2 to obtain the correct MAC address for the actual intended destination host and send the packet to the intended destination using the MAC address of the ultimate destination.

b. Ethernet

The most widely used network connection for personal computers is an Ethernet connection. Ethernet is really a standard for computer network technologies that describes both hardware and communication protocols, being both a link layer and physical layer technology. It was initially developed at Xerox PARC in the 1970's for sharing printers as a way to limit costs. It was intended to be used in a bus topology, with computers attached to segments of coaxial copper cable limited to 500 m in length. The original work on Ethernet used 75 Ohm coaxial cable, and operated at 3 Mbps. By the mid-1990's the standard speed had increased to 100 Mbps. At present speeds up to 40 Gbps are available (albeit for specialist networks) with 10 Gbps Ethernet (10GBASE-T) having been standardised back in 2007.

Nowadays Ethernet is mainly used in a switch-based star topology using store-and-forward packet switching. Earlier on hubs were used instead of switches, where the hosts and routers directly connected to a hub with twisted-pair copper wire. A hub is a physical layer device that acts on individual bits rather than frames. When a bit enters one interface of a hub it boosts the signal's power and broadcasts it on all its other interfaces. In the hub-based start topology Ethernet collision control was required since a hub operates only at the physical layer, so MAC protocols were developed for the purpose of collision detection and avoidance. If using switches instead for Ethernet LAN, there are no collisions because switches operate up to the link layer and forwards frames selectively based on the MAC address, and are usually used in full duplex connections so there is no need for a MAC protocol. Carrier Sense Multiple Access protocol with Collision Detection (CSMA/CD) is used to access a segment for transmission when a half-duplex device or connection is used.

A side note on half and full duplexing: Full-duplex transmission is achieved by setting switch interfaces, router ports, and host NICs to full duplex, provided the host network card and the switch port are capable of operating in full duplex mode. A dedicated switch port is required for each full duplex node. Micro-segmentation, where each network device has its own dedicated segment to the switch, ensures that full duplex will work properly. The network device has its own dedicated segment so there are no collisions in full duplex transmission. With full duplex transmission, the device can send and receive at the same time, effectively doubling the amount of bandwidth between nodes and ensuring that no collisions occur.

Most modern computers have an Ethernet-enabled NIC built into the motherboard, such as the one shown in Figure 8. RJ.45 is a common socket type used for Ethernet connections.

FIGURE 8: CISCO ETHERNET NIC WITH FOUR RJ.45 SOCKETS

Ethernet LANs may be implemented using a variety of media. Examples are:

- 10BASE-5 (10B5) Low loss coaxial cable (also known as "thick" Ethernet)
- 10B2 Low cost coaxial cable (also known as "thin" Ethernet)
- 10BT Low cost twisted pair copper cable (also known as Unshielded Twisted Pair (UTP), Category-5)

- 10BF Fibre optic cable
- 100BASE-T Low cost twisted pair copper cable (also known as Unshielded Twisted Pair (UTP), Category-5)
- 100BF Fibre Fast Ethernet
- 1000BT Low cost twisted pair copper cable (also known as Unshielded Twisted Pair (UTP), Category-5)
- 1000BF Fibre Gigabit Ethernet
- 10000BT Category 6 (Unshielded Twisted Pair, a.k.a Cat-6)
- 10000BT Fibre 10 Gigabit Ethernet

The number at the beginning of the standard specifies the speed. "BASE" means that baseband Ethernet only is carried on the physical medium. The final part refers to the physical medium, T for twisted pair and F for fibre. Some design rules that should be observed when using different cable technologies for Ethernet installations are given in Table 1. If the distance or maximum number of systems per segment are exceeded performance is likely to be compromised with higher bit error rates and more delay-prone network.

TABLE 1: NETWORK DESIGN RULES FOR DIFFERENT TYPES OF ETHERNET CABLE TECHNOLOGY

Segment type	**Max Number of systems per cable segment**	**Max Distance of a cable segment**
10B5 (Thick Coax)	100	500 m
10B2 (Thin Coax)	30	185 m
10BT (Twisted Pair)	2	100 m
10BFL (Fibre Optic)	2	2000 m

Ethernet has also evolved to provide network Operations and Management (OAM) functions. These standards allow network operators to manage faults on the network and to validate the performance of the service.

c. Virtual LANs

VLANs allow a network manager to segment a LAN into different broadcast domains logically so that computers on the VLAN do not physically have to be located together. VLANs allow multiple virtual local area networks to be defined over a single physical LAN infrastructure and it is possible to have hosts that are connected together on the same physical LAN not being allowed to communicate directly. This enables the network administrator to organise a network according to requirements without needing the physical LAN to mirror the logical connection requirements of the organisation. The purpose of a VLAN is to remove the limitation of physically switched LANs with all devices

automatically connected to each other. Each port on a switch can be configured into a specific VLAN, and then the switch will only allow devices that are configured into the same VLAN to communicate. In this way VLANs solve problems arising in LANs:

- Lack of traffic isolation. In physical LANs broadcast messages will be sent to all devices on the network. Limiting the scope of such broadcast traffic would improve LAN performance and introduce improved security and privacy.
- Inefficient use of switches if grouping and isolation were done physically and not through VLANs
- User management with organisational changes and movements of people. Rewiring every time organisational changes occur that require restructuring of LANs would be expensive and time-consuming. With VLANs this can be avoided.
- Increased broadcast domain size when the number of devices on a LAN grows large that makes inefficient use of resources, slows down the communication process and the hosts themselves

VLAN membership can be classified by port, MAC address, protocol type or IP Subnet Address. The 802.1Q draft standard defines Layer 1 and Layer 2 VLANs only. Protocol type based VLANs and higher layer VLANs are allowed, but not defined in this standard. As a result, these kinds of VLANs are proprietary.

In order to set up a VLAN switches must include configuration tables indicating which VLANs are accessible via which interfaces.

VLANs do not includes a mechanism for communication between VLANs so to do this layer 3 (network layer) devices must be introduced to the network. One way of interconnecting VLANs to allow communication between them is to connect a VLAN switch port to an external router and configure the port to belong to both (or all) the required VLANs. It is not always necessary to use a separate switch as there are some devices on the market that contain both a VLAN switch and a router. This is a type of VLAN created through an *access link*. It connects VLAN-unaware devices to the port of a VLAN-aware bridge.

Another more scalable method to interconnect VLANs is called *trunking* or tagging. A trunk is a single transmission channel between two points, each point being either the switching centre or the node. VLAN trunking allows a single network adapter to behave as n virtual network adapters for n VLAN network segments, carrying multiple VLANs through a single network link using a trunking protocol, where n has a theoretical upper limit of 4096 but is typically limited to 1000. Trunking requires that the network switch, network adapter, and the drivers for the operating system all support VLAN tagging in order for them to trunk. A VLAN tag is added to packets going through a trunk in order to identify the VLAN that the packet belongs to.

There are several different trunking technologies:

- **IEEE 802.1Q**: the IEEE industry open standard for VLAN frame tagging and the only technology that can be used to interoperate between devices from different vendors. In 802.1Q The VLAN frame tag is placed inside an Ethernet frame when it reaches a switch that is a member of a VLAN. If the switch has a trunk port, the Ethernet frame can be forwarded out the trunk link port.
- **Inter-Switch Link** (ISL): Cisco proprietary VLAN frame tagging method that encapsulates an Ethernet frame with an ISL header.
- **LAN Emulation** (LANE): LANE is used to communicate with multiple VLANs over Asynchronous Transfer Mode (ATM).
- **802.10** (FDDI): Protocol for sending VLAN information over FDDI.
- **VLAN Trunking Protocol** (VTP) is a proprietary CISCO protocol for propagating VLAN information to all the switches in a VTP domain. VTP advertisements can be sent over 802.1Q or ISL trunks. The information propagated is management domains, configuration revision number and currently known VLANs and their specific parameters.
- MAC-based approaches where the network manager specifies the set of MAC addresses that belong to each VLAN. In this case whenever a device attaches to a port, the port is connected into the appropriate VLAN based on the MAC address of the device.
- Network-layer approaches (e.g., IPv4, IPv6, or Appletalk)

d. Multiprotocol Label Switching

Multiprotocol Label Switching (MPLS), standardised by the IETF in RFC 3031, is a data-carrying technique generally considered to reside between traditional OSI data link and network layers. It incorporates Virtual Circuit techniques into a routed datagram network by applying fixed-length labels. It was designed to provide a unified data-carrying service for both circuit-based clients and packet-switching clients which provide a datagram service model. MPLS is used on top of IP addressing by selectively labelling datagrams and allowing routers to forward datagrams based on fixed-length labels instead of IP addresses when possible. It can be used to carry many different kinds of traffic, including IP packets, E1/T1, ATM, Frame Relay, SONET, and Ethernet frames but is also beginning to replace some of these access technologies, most notably ATM. MPLS is mainly used for traffic engineering purposes and to forward IP protocol data units and Virtual Private LAN Service Ethernet traffic. Generalised Multi-Protocol Label Switching (GMPLS) is a protocol that extends MPLS to manage further classes of interfaces and switching technologies other than packet interfaces

and switching. Examples are TDM, link layer switching, wavelength switching and fibre (port) switching.

9 Conclusion

In this Chapter we discussed the data link layer, starting with some services it provides. We saw that the link layer can be seen under the structure of two sublayers - the Link Layer Control sublayer and the MAC sublayer. In terms of control, we discussed link layer error detection and correction techniques. We saw that there are three categories of MAC protocols: fixed assigned, demand assigned and random access MAC. Fixed MAC includes the fundamental ways of dividing network resources in wireless networks i.e. FDMA, TDMA and CDMA. We mentioned the polling MAC technique and some examples of its use. The ALOHA and slotted ALOHA random access protocols were introduced, and the CSMA technique for gaining channel access. ARP was introduced and its functioning explained in some detail for use within a subnet. It was also mentioned what alterations are required to provide the same mapping function over different subnets. The Ethernet protocol was introduced. The concept of a VLAN was introduced and different ways of creating VLANs. Finally we briefly mentioned the MPLS protocol.

10 Review Questions

1. If all the links in the Internet were to provide reliable delivery service, would the TCP reliable delivery service be redundant? Why or why not?
2. What are some of the possible services that a link-layer protocol can offer to the network layer and mention which of these services have an implementation in TCP and IP?
3. Why would the token-ring protocol not work well if a LAN had a very large perimeter?
4. How big are the MAC address space, IPv4 address space and The IPv6 address spaces respectively?
5. What are some desirable features of a MAC protocol in a wireless network that is spread out over large distances?
6. If a sending device does not know the MAC address of the destination device, what protocol is used to find the MAC address of the receiving device?
7. Will a host always update its ARP cache upon receiving an ARP request?
8. Host A wants to send data to host B. Host B is on a different segment from host A. The two segments are connected to each other through a router. What will host B see as the source MAC address for all frames sent from host A?

9. In what situations do you think contention based MAC protocols are suitable and why?
10. How does CSMA/CD improve performance over CSMA only?
11. How does replacing a hub with a switch affect CSMA/CD behaviour in an Ethernet network?
12. What is an advantage of token passing protocol over CSMA/CD protocol?
13. Suppose that N switches supporting K VLAN groups are to be connected via a trunking protocol. How many ports are needed to connect the switches? Justify your answer.
14. When a new trunk link is configured on a switch, which VLANs are allowed over the link?
15. Using the information in this chapter, what is a way to improve network performance by increasing the bandwidth available to hosts and limiting the size of the broadcast domains?

Answers

1. No, it would not be redundant because although the reliable delivery service would ensure that datagrams sent over the link are received without errors, arrival in the correct order is not guaranteed. The IP datagrams may take different routes and arrive out of order and TCP is still required to ensure that the byte stream is delivered in the correct order.
2. Framing, done by both IP and TCP; reliable delivery and flow control, both provided by TCP; error detection, provided in TCP/IP and error correction are some examples.
3. When a node transmits a frame it must wait for the frame to propagate around the entire ring before it can release its token so if the perimeter is large the propagation time will be long.
4. 2^{48} MAC address, 2^{32} IPv4 address and 2^{128} IPv6 addresses
5. The protocol should be simple to implement, decentralised and require minimal overhead
6. ARP
7. No, the ARP cache is only updated if the ARP request is for its IP address.
8. Because host B is on a different segment that is separated by a router, the MAC address of all frames sent from host A will be the MAC address of the router. Anytime a frame passed through a router, a router rewrites the MAC address to the MAC address of the router's exit interface for the segment and then sends the frame to the local host. In this case, the router will change the source MAC address of the frame sent from host A with the MAC address of its interface connecting to the segment host B is on. Host B will see that the frame came from the MAC address of the router with the IP address of host A.
9. Contention-based protocols are suitable for bursty traffic under light to moderate load. These techniques are decentralised, simple and easy to implement.
10. In CSMA, a station monitors the channel before sending a packet. Whenever a collision is detected, it does not stop transmission leading to some time being wasted on backoffs and retransmissions. In CSMA/CD scheme, whenever a station detects a collision, it sends a signal to inform other stations that a collision has occurred, thus reducing time wastage and improving in performance.

11. Collisions are eliminated and CSMA/CD can be disabled.
12. The CSMA/CD is not a deterministic protocol. A packet may be delivered after many collisions leading to long variable delay. Some packets may not be delivered at all. On the other hand, token passing protocol is a deterministic approach, which allows a packet to be delivered within a known time frame. It also allows priority to be assigned to packets.
13. The N switches can be connected together so the first and last switches each use one port for trunking while the middle N-2 switches use two ports each. This makes the total number of ports 2+2(N-2)=2N-2.
14. By default, all VLANs are allowed on the trunk link and the network administrator must remove by hand each VLAN that should not be allowed to traverse the trunked link.
15. By creating and implementing VLANs in the switched network, it can be segmented into smaller broadcast domains at link layer. For hosts on different VLANs to communicate, a router or network layer switch will then be required.

11 Further reading:

1) J. Kurose and K. Ross, “Computer Networking: A Top-Down Approach (International) Chapter 3”, Pearson Education Ltd., Essex, sixth edition, 2013.