Ethernet

The LAN Killer

"Ethernet works in practice but not in theory."

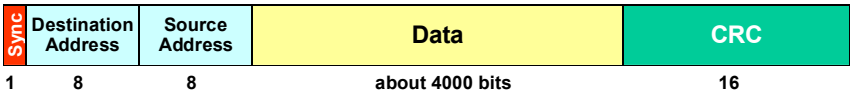
Robert Metcalfe

Yeah,...Robert Metcalfe was the inventor of Ethernet.

History (1)

- Late 1960s: **Aloha protocol University of Hawaii**
- Late 1972: **Robert Metcalfe developed first Ethernet system based on CSMA/CD**
  - **Xerox Palo Alto Research Center (PARC)**
  - **Exponental Backoff Algorithm was key to success (compared with Aloha)**
  - **2.94 Mbit/s**

Original Ethernet Frame



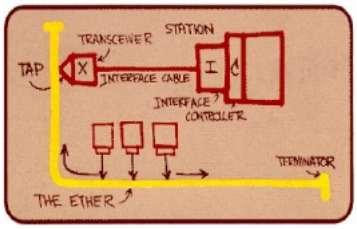
The Aloha protocol was fairly simple: send whenever you like, but wait for an acknowledgement. If there is no acknowledgement then a collision is assumed and the station has to retransmit after a random time. "Pure Aloha" achieved a maximum channel utilization of 18 percent. "Slotted Aloha" used a centralized clock and assigned transmission slots to each sender, hereby increasing the maximum utilization to about 37 percent. Robert Metcalfe perceived the problem: another backoff algorithm was needed but also "listen before talk". Metcalfe created Carrier Sense Multiple Access Collision Detection (CSMA/CD) and a truncated exponential backoff algorithm which allows a 100 percent load.

Robert Metcalfe's first Ethernet system used a transmission rate at 2.94 Mbit/s which was the system clock of the Xerox Alto workstations at that time. Originally, in 1972 Metcalfe called his system Alto Aloha Network, but one year later he renamed it into "Ethernet" in order to emphasize that this networking system could support any computer not just Altos – and of course to clarify the difference to traditional Aloha!

History (2)

- 1976: **Robert Metcalfe released the famous paper: "Ethernet: Distributed Packet Switching for Local Computer Networks"**

Original sketch



(C) Herbert Haas 2005/03/11

The press has often stated that Ethernet was invented on May 22, 1973, when Robert Metcalfe wrote a memo to his bosses stating the possibilities of Ethernet's potential, but Metcalfe claims Ethernet was actually invented very gradually over a period of several years. In 1976, Robert Metcalfe and David Boggs (Metcalfe's assistant) published a paper titled, "Ethernet: Distributed Packet-Switching For Local Computer Networks."

Metcalfe left Xerox in 1979 to promote the use of personal computers and local area networks (LANs). He successfully convinced Digital Equipment, Intel, and Xerox Corporations to work together to promote Ethernet as a standard. Now an international computer industry standard, Ethernet is the most widely installed LAN protocol.

History (2)

- 1978: **Patent for Ethernet-Repeater**
- 1980: **DEC, Intel, Xerox (DIX) published the 10 Mbit/s Ethernet standard**
  - **"Ethernet II" was latest release (DIX V2.0)**
- Feb 1980: **IEEE founded workgroup 802**
- 1985: **The LAN standard IEEE 802.3 had been released**

First Ethernet standard was entitled "The Ethernet, A Local Area Network: Data Link Layer and Physical Layer Specifications" and focused on thick coaxial cable only.

The IEEE Working Groups

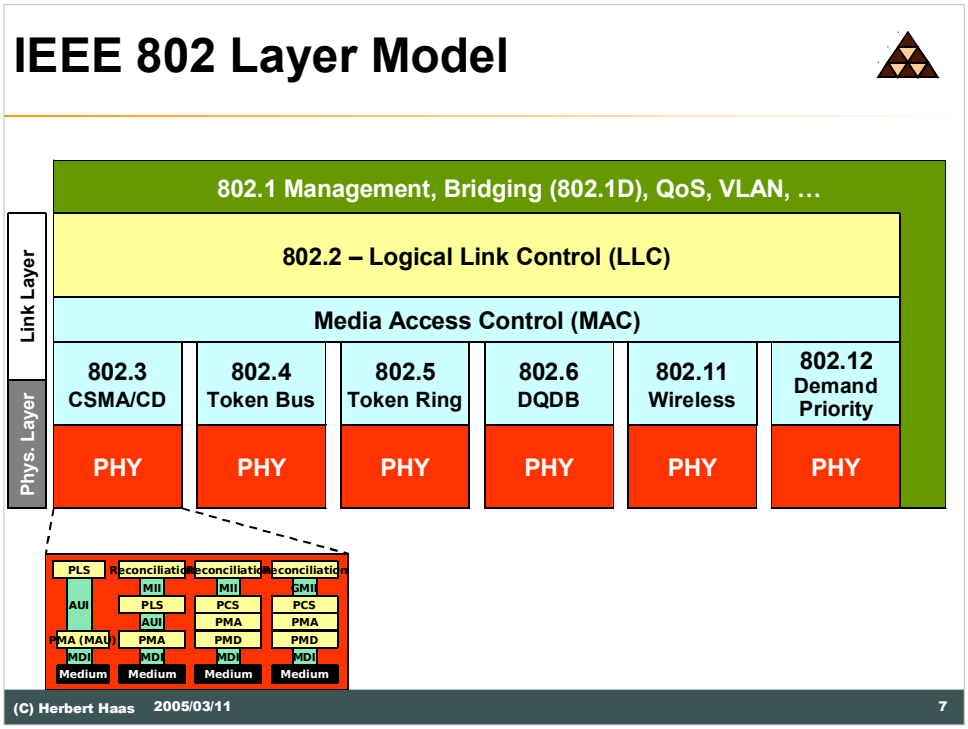
- 802.1 Higher Layer LAN Protocols
- 802.2 Logical Link Control
- 802.3 Ethernet
- 802.4 Token Bus
- 802.5 Token Ring
- 802.6 Metropolitan Area Network
- 802.7 Broadband TAG
- 802.8 Fiber Optic TAG
- 802.9 Isochronous LAN
- 802.10 Security
- 802.11 Wireless LAN
- 802.12 Demand Priority
- 802.13 Not Used — Superstition?
- 802.14 Cable Modem
- 802.15 Wireless Personal Area Network
- 802.16 Broadband Wireless Access
- 802.17 Resilient Packet Ring

On this slide you can see a summary of the most important IEEE standards so far. The Ethernet system is covered by the standards 802.1, 802.2 and 802.3.

The 802.1 describes management and optional functions inside the Ethernet technology like the Spanning-tree (SPT) process, Ethernet bridging, VLAN systems, etc.

The 802.2 standards describes the Logical Link Control (LLC) function, which is only used in 802.3 Ethernet systems, and that allows the use of Ethernet in connection-oriented or connection-less mode.

The 802.3 standard describes the physical layer of the Ethernet system plus the media access that is controlled by the CSMA/CD procedure.



The physical layer is responsible for the speed of the transmission currently there are four different speeds available, 10, 100, 1000, 10000 Mbit/s. In the graphic the physical interface structure of the 10, 100, 1000 Mbit/s systems is shown.

The interface function between the physical layer and the Ethernet data-link layer is performed by the CSMA/CD algorithm.

The Medium Access Control layer is responsible for addressing and it controls whether a data frame is picked up from the wire and is loaded into the buffer of the Ethernet card or not.

The Logical Link Control layer is responsible for the interface function between the Ethernet layer and higher layers on top of Ethernet plus the support of connection-less or connection-oriented mode.

The 802.1 Management cannot be seen as an separate Ethernet layer but it describes additional optional Ethernet functions like bridging, QoS, flow control, SPT, etc.

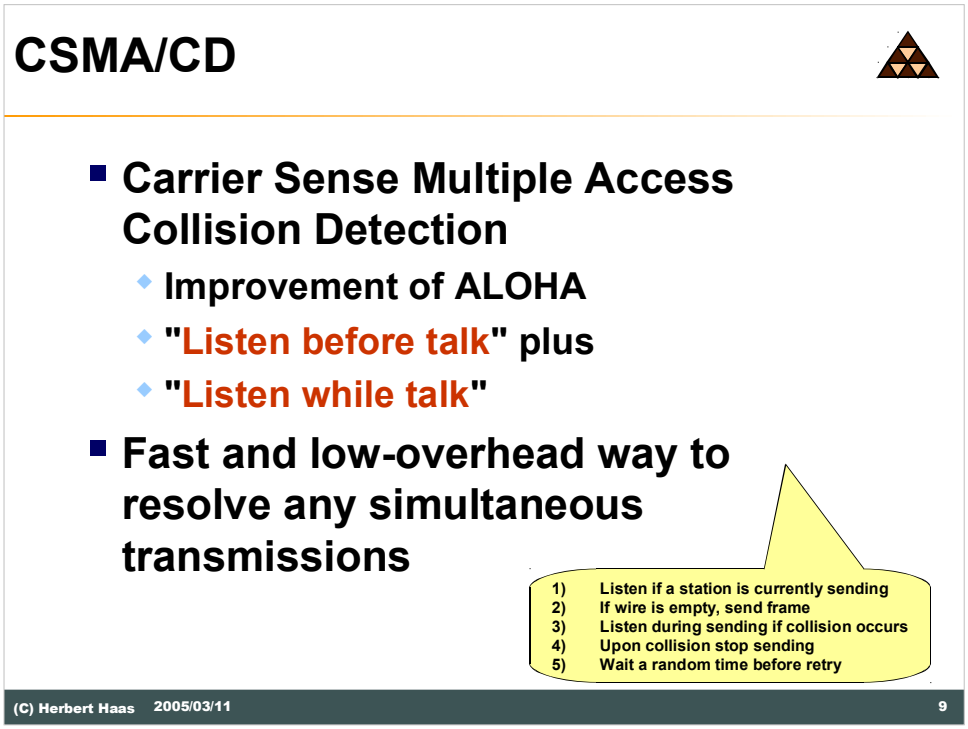
IEEE 802.3/Ethernet

- Since 1984 the IEEE also maintains the DIX Ethernet standard
- Both frame types are supported by "Ethernet NICs"
  - Network Interface Cards

Remember at the early days of Ethernet there were two competing organizations the IEEE committee responsible for the 802.X standards and the companies Digital, Intel and Xerox which where responsible for the Ethernet 2 DIX standard.

In the year 1984 the DIX committee disappeared and the IEEE took over the responsibility to maintain and adapt the DIX standard for new upcoming Ethernet technologies.

Today all Ethernet interface cards support both frame types the 802.3 and the ETH 2 frame.



Ethernet is a shared media technology, so a procedure had to be found to control the access onto the physical media. This procedure was called the Carrier Sense Multiple Access Collision Detection (CSMA/CD) circuit.

The way it works is quite simple, every stations that wants to send need to do a Carrier Sense to check if the media is already occupied or not.

If the media is available the station is allowed to perform an Media Access and may start sending data.

In the case that two stations almost at the same time access the media, a collision will happen. To recognize and resolve a collision is the task of the Collision Detect circuit.

Every station listens to its own data while sending. In the case of a collision the currently sending stations recognize the collision by the superimposition of the electrical waves on the wire. A jamming signal will be sent out to make sure all involved stations recognize the occurrence of an collision.

All stations involved in the collision stop sending and start a randomize timer. When the randomize timer expires the station may try to access the media again.

Slot Time

- Minimum frame length has to be defined in order to safely detect collisions
- Each frame sent must stay on wire for a RTT duration – at least
- This duration is called "slot time" and has been standardized to be 512 bit-times
  - 51,2 µs for 10 Mbit/s

(C) Herbert Haas 2005/03/11 10

There is a very basic Ethernet rule that says a collision must be detected while a station is transmitting data. Therefore a stations needs to keep on sending at least of the duration of the RTT of the Ethernet system. The maximum allowed RTT is standardized and is called the slot time. The slot time for 10Mbit/s Ethernet systems is set to 51,2 µs.

If collisions occur after expiration of the slot time we talk about “late collisions”, which may cause malfunctions in the network.

For example if a station transmits a frame and no collision was detected, the station assumes correct delivery of the frame. Now the station removes the frame from the transmit buffer, leaving no chance to retransmit the frame in the case of a late collision.

Slot Time Consequences

- So minimum frame length is 512 bits (64 bytes)
- With signal speed of 0.6c the RTT of 512 bit times allows a network diameter of
  - 2500 meters with 10 Mbit/s
  - 250 meters with 100 Mbit/s
  - 25 meters with 1000 Mbit/s (!)

NOTE: Only valid on shared media (!)

(C) Herbert Haas 2005/03/11 11

The minimum frame length in Ethernet systems is set to 64 byte or 512 bit. This minimum frame length plus the slot time in combination with the speed of electrical signals on a wire (~ 180.000 km/s) determines the maximum outspread of an Ethernet system.

Therefore we end up at a maximum outspread of 2500 meters for 10Mbit/s Ethernet systems. The maximum outspread of faster Ethernet systems is directly related to their shorter slot times, because of the higher speed.

These distance limitations must only be taken into account in shared media environments like Ethernet Bus and Hub systems. In more modern switched environments using full duplex communication these distance limitations can be neglected.

Exponential Backoff (1)

- Most important idea of Ethernet !
- Provides maximal utilization of bandwidth
  - After collision, set basic delay = 512 x slot time
  - Total delay = basic delay * rand
  - 0 <= rand < 2^k
    - k = min (number of transm. attempts, 10)
- Allows channel utilization

The retransmission in case of collisions is controlled by the exponential backoff algorithm.

The retransmission is delayed about a basic delay, which is set to 26,2 milliseconds for 10 Mbit/s Ethernet, times a random factor. The range out of which the randomize factor is selected is increasing with the number of retransmission attempts.

Repeated collisions indicate a busy medium, therefore the station tries to adjust to the medium load by progressively increasing the time delay between repeated retransmission attempts.

Exponential Backoff (2)

- After 16 successive collisions
  - Frame is discarded
  - Error message to higher layer
  - Next frame is processed, if any
- Truncated Backoff (k<=10)
  - 1024 potential "slots" for a station
  - Thus maximum 1024 stations allowed on half-duplex Ethernet

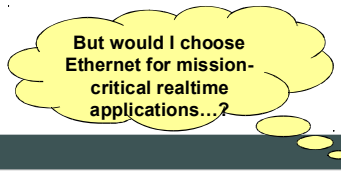
The retransmission of a frame is attempted up to a defined maximum number of retries typically known as the attempt limit. The attempt limit is set to a maximum of 16 retries by the standard.

After 16 retries the frame is discarded and a error message is sent to higher layers. Then the station continuous to process the next frame.

Due to the truncated backup algorithm a maximum of 1024 potential time slots for a station are available. So the maximum number of stations attached to half duplex Ethernet systems should not exceed 1024 stations.

Channel Capture

- Short-term unfairness on very high network loads
- Stations with lower collision counter tend to continue winning
- 10 times harder to occur on 100 Mbit/s Ethernet
- Rare phenomena, so no solution against it



In the case of very high network loads Ethernet tends to prefer stations with lower collision counters, because they try to access the media in shorter time intervals than stations with a higher collision counter.

This is a phenomena that was never solved in Ethernet systems, but can be disregarded for today's Ethernet networks, because most of them are switched networks where collisions play no or just a minor role.

Collision Detection

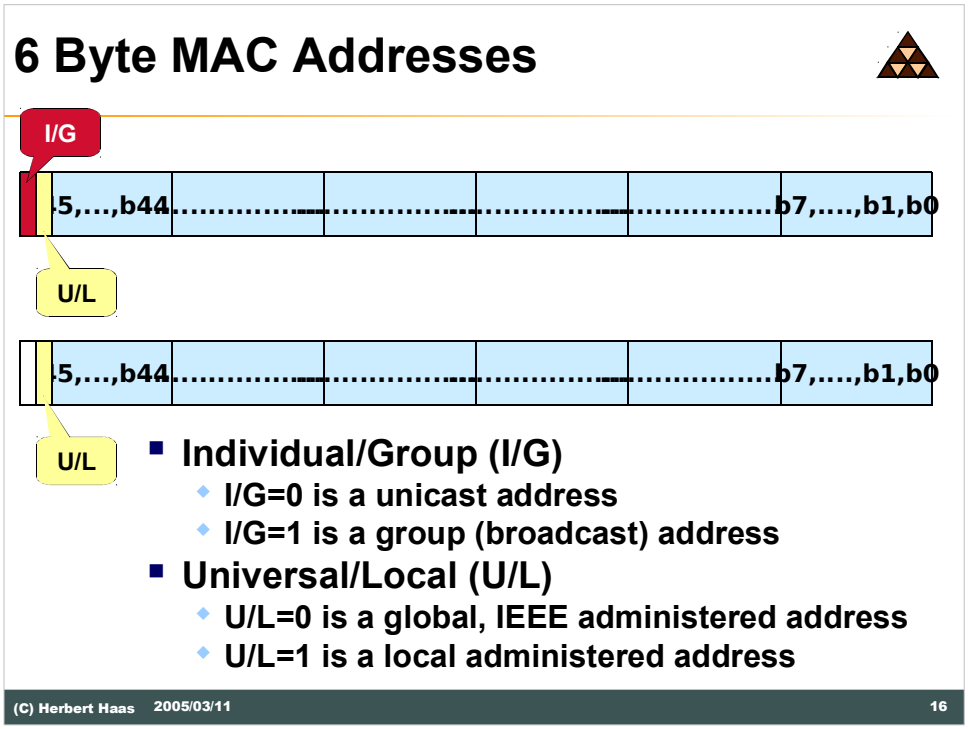
- 10Base2, 10Base5
  - Manchester with –40 mA DC level
  - "high" = 0 mA, "low" = –80 mA
- 10BaseT
  - Manchester with no DC offset
  - Collisions are detected by Hub who sends a "Jam" signal back
  - Similarily at 100BaseT and 1000BaseT

(C) Herbert Haas 2005/03/11 15

The method of collision detection is different for every physical layer.

In coaxial Ethernet, transceivers send their Manchester code using the DC offset method. A "high" value is nominally zero current; a "low" value is nominally -80 mA. This results in a DC component to the signal of -40 mA, which creates a voltage of -1 VDC (the transceiver sees a 25 ohm load from the two 50 ohm cables going "left and right" away from the transceiver). When two transceivers send at the same time, their currents add, increasing the DC component of the combined signal to -2 VDC Thus, we can detect collisions by looking for DC signals in excess of the maximum that could possibly be generated by a single transmitter.

In 10BASE-T, the Manchester code is sent symmetrically, with no DC offset. Collisions are detected in the repeater hub, which can observe when two or more devices are transmitting at the same time. Normally, the hub does not repeat a station's own signal back to the station on its receive cable pair. However, when a collision is noted, the hub does send a signal (the so-called "collision enforcement", or "jam") to the transmitting stations. The stations detect collisions by noting when they see a signal on their receive pair at the same time that they are transmitting on their transmit pair.

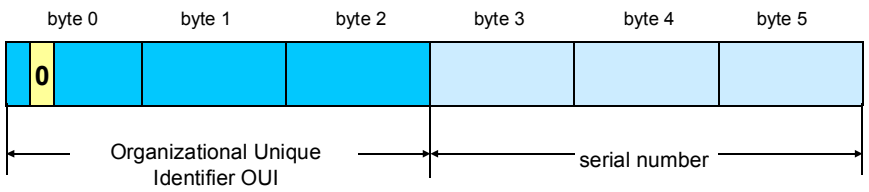


The addresses used in Ethernet systems are called MAC addresses. A MAC address is 6 bytes or 48 bits long and is typically written in hexadecimal notation. Each Ethernet network card has one burnt in MAC address. Network cards of some vendors even support the use of programmable local administered MAC addresses.

Ethernet is using a canonical address format, which defines the order how bits from the transmission buffer are put onto the medium. In Ethernet systems the least significant bit of each byte is put on the medium first followed by the more significant bits.

The first two bits of a MAC address on the **wire** have a special meaning. The first bit (I/G) specifies whether the MAC address is a unicast address (0) or a broadcast/multicast address (1). The second bit (U/L) specifies whether it's a global and unique MAC address, or a locally programmed and administered address.

MAC Address Structure



- **Each vendor of networking component can apply for an unique vendor code**
- **Administered by IEEE**

(C) Herbert Haas 2005/03/11 17

The MAC addresses are globally administered by the International Electrical and Electronic Engineering (IEEE) standardization organization.

Each vendor of networking components can apply for an globally unique vendor code. The vendor code costs 1000$ and occupies the first three bytes of the MAC address.

The remaining three bytes of the MAC address may be used by the vendor to address its components.

Ethernet Frames

- Due to different development branches, there are two different frame types
  - IEEE type: consists of MAC and LLC
  - DIX type: consists of a Type field
- Why using both?
  - Different applications have been defined for either IEEE or DIX

Due to the historical development of Ethernet there are two different types of Ethernet frames. The DIX type commonly called Ethernet 2 frame and the IEEE type known as 802.3 frame.

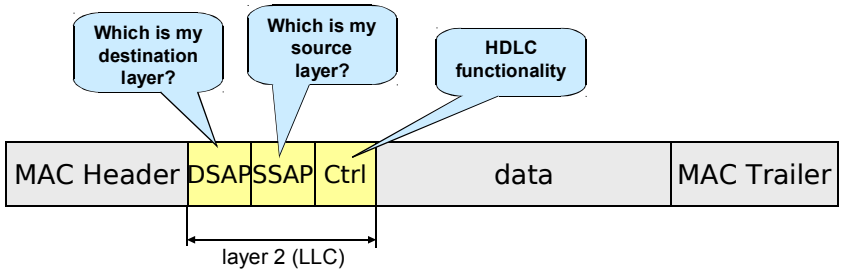
The IEEE frame type consists of a MAC part, an LLC (802.2) part and is using the Destination/Source Service Access Points (DSAP, SSAP) to interface with higher layers.

The DIX frame type consists of a MAC part and a Type field used to interface with higher layers.

Which frame type is used depends on the higher layer protocols e.g. for the transport of IP frames the DIX type is specified.

IEEE 802.2 (LLC)

- **Every IEEE LAN/MAN protocol carries the Logical Link Control header**
  - **HDLC heritage**



Basic frame format of every IEEE protocol

(C) Herbert Haas 2005/03/11 19

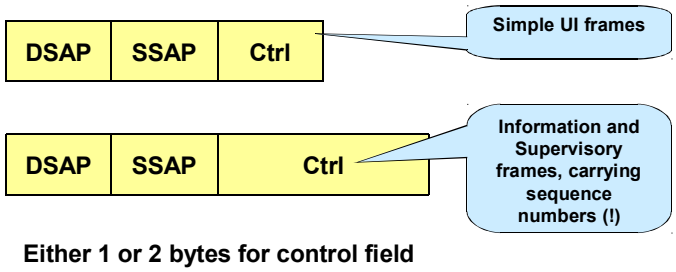
The LLC (802.2) is part of every basic frame format that is specified by the IEEE e.g. Token ring, Token bus, Ethernet, etc.

The DSAP and SSAP field are both eight bit in length and are used to address layer 3 processes. With the SSAP the layer 2-3 interface used at the source is specified, while the DSAP specifies the layer 2-3 interface at the destination. But typically it is very unlikely to use a SSAP value different from the DSAP value, because only layer 3 processes of the same kind are able to communicate with each other. So IP to IP communication would use a SSAP and DSAP value of 0 x 06.

The Control field inside the LLC can be used for connection-oriented or connection-less communication and the way it works is basically the same what HDLC does.

LLC Details

- **According sophisticated HDLC functionalities, 4 LLC classes defined**
  - **Class 1 is most important (UI, no ACKs)**



Either 1 or 2 bytes for control field

(C) Herbert Haas 2005/03/11 20

The LLC functionality is divided into four classes:

•Class 1- connection-less unacknowledged service

•Class 2- connection-oriented service

•Class 3- Class 1 plus connection-less acknowledged service

•Class 4- Class 2 plus connection-less acknowledged service

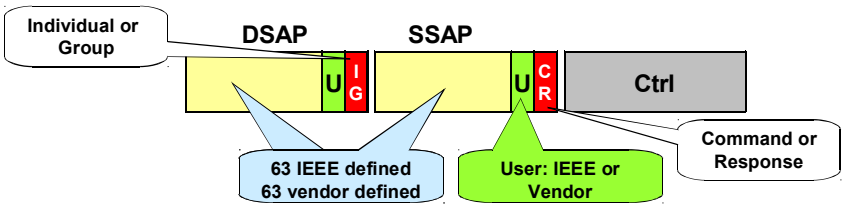
Class 1 offers best effort service only, while Class 2 works connection-oriented with error recovery and flow control support.

The most important service class is the Class 1 connection-less service, because the tasks of error recovery and flow control are typically performed by higher layer processes e.g. TCP.

Only protocols like Microsoft's Netbeui or IBM's SNA need Class 2 connection-oriented service, because error recovery and flow control is not supported by their protocol stacks.

SAP Identifiers

- 128 possible values for protocol identifiers
- Examples:
  - 0x42 … Spanning Tree Protocol 802.1d
  - 0xAA… SNAP
  - 0xE0… Novell
  - 0xF0… NetBios



(C) Herbert Haas 2005/03/11 21

The DSAP and the SSAP are both 8 bit in length. The least significant bit in the DSAP is reserved to indicate whether it's a individual or group access point. In the SSAP this bit is the command/response bit and is not used in Ethernet systems. The U bit is used to specify whether its an IEEE or vendor specific access point.

Hex E0 ..... Novell (U=0)
Hex Fy ..... reserved for IBM (U=0)
Hex F0 ..... Netbios (U=0)
Hex F4 ..... IBM LAN manager individual (U=0)
Hex F5 ..... IBM LAN manager group (U=0, I/G =1)
Hex F8 ..... remote program load (U=0)
Hex 04 ..... SNA path control individual (U=0)
Hex 05 ..... SNA path control group (U=0, I/G =1)

The range Hex 8y to 9C (with U=0) is reserved for free usage except y = xx1x (binary notation); U=1

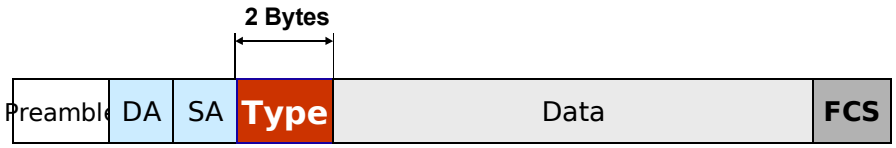
Hex 00 ..... Null SAP

A station with running LLC software always responds to a frame destined to the Null SAP a LLC Ping can be implemented.

Hex 03 ..... LLC sub-layer management (U=1)
Hex 06 ..... DoD IP (U=1)
Hex 42 ..... 802.1d Spanning Tree Protocol (U=1)
Hex AA ..... TCP/IP SNAP (U=1)
Hex FE ..... ISO Network Layer (U=1)

DIX Type field

- 2-bytes Type field to identify payload (protocols carried)
  - Most important: IP type 0x800
- No length field



"THE" Ethernet Frame

(C) Herbert Haas 2005/03/11 22

The Type field used by the DIX Eth2 frame format is 16 bit in length and allows therefore to address up 65 536 different layer 3 processes. The Type field only allows the addressing of the destination service access point. The indication of the source service access point is not supported by the DIX frame format. Typically only layer 3 processes of the same kind are able to communicate with each other.

Some Type field examples:

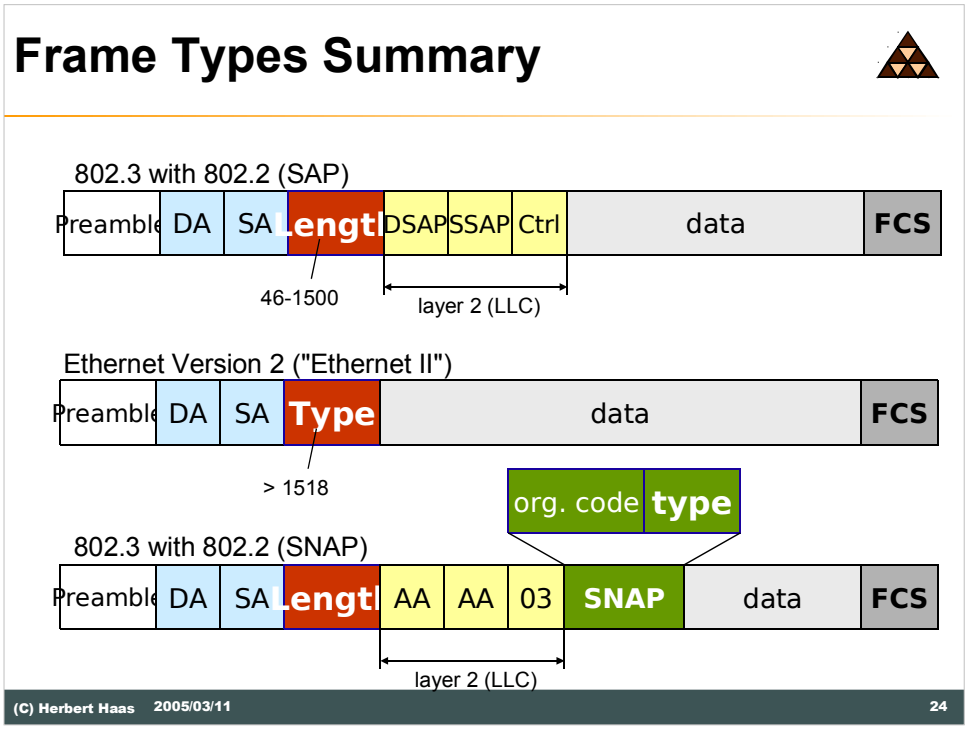
Hex 0800.....	IP
Hex 0806.....	ARP
Hex 8035.....	RARP
Hex 814C.....	SNMP
Hex 6001/2.....	DEC MOP
Hex 6004.....	DEC LAT
Hex 6007.....	DEC LAVC
Hex 8038.....	DEC Spanning Tree
Hex 8138.....	Novell

SNAP

- Demand for carrying type-field in 802.4, 802.5, 802.6, ... also !
- Subnetwork Access Protocol (SNAP) header introduced
  - If DSAP=SSAP=0xAA and Ctrl=0x03 then a 5 byte SNAP header follows
  - Containing 3 bytes organizational code plus 2 byte DIX type field

The IEEE had problems to address all necessary layer 3 processes, due to the short (8 bit) DSAP and SSAP fields in the IEEE header. So they introduced a new frame format which was called Subnetwork Access Protocol (SNAP). The SNAP format was simply importing the DIX Type field by the backdoor. This new header format was then also used for technologies like Token Ring, Token Bus, DQDB, etc.

In the SNAP format the DSAP and the SSAP is set to the hex value of AA. This indicates an five byte extension to the standard 802.2 header, which is made up of a three byte long field called Organization Unique Identifier (OUI) and the two byte Type field.



So we end up with three different frame formats used in Ethernet systems. The 802.3 without SNAP, the DIX Eth2 format and the 802.3 with SNAP.

The DIX Eth 2 frame format is mainly used for the data transport of protocols that have the functionality of error recovery and flow control implemented in their protocol stack e.g. IP.

The 802.3 without SNAP frame format is used for protocols that need the functions of error recovery and flow control on layer 2 e.g. Netbeui, SNA.

The 802.3 with SNAP frame format is used by vendors to implement proprietary protocols, for example Cisco´s CDP, VTP, CGMP, etc. protocols. For such purposes the OUI field is used to indicate the vendor and the type field value is chosen vendor specific.

PHY Variants

- 10Base2 (10 Mbit/s, 200 meters)
- 10Base5 (500 meters)
- 10BaseT (star-like cabling, hub needed)
- 10BaseF (fiber)
- 10Broad36 (broadband cable)
- 100BaseT
- 1000BaseT
- 1000BaseX

In this graphic an overview of the currently available physical layers of the Ethernet system is shown.

The 10Base5, 10broad36 and the 10Base2 Ethernet bus systems can be seen as historic and might only be found in existing elder installations.

The 10BaseT was the first Ethernet system that allowed to build up star shaped networks by the help of HUB´s and CAT 3 Unshielded Twisted Pair (UTP) cables. Also a fiber interface 10BaseF exits for the 10 Mbit/s Ethernet system, but is very rarely used because of the higher costs compared to copper interfaces.

The 100BaseT uses a cabling infrastructure of CAT 5 UTP cables and a 4B/5B coding scheme. This encoding scheme adds a fifth bit for every four bits of user data, to allow enough changes in the signal for synchronization purposes.

10BaseT as well as 100BaseT are only using the pins 1, 2, 3 and 6 from the eight pin RJ45 connector.

1000BaseT is a copper interfaces that allow the transport of Gigabit Ethernet on CAT 5e UTP cables by the use of all four pairs, 5 level PAM code and echo cancellation. 1000BaseT is backward compatible to 10BaseT and 100BaseT.

1000BaseX can be used in combination with fiber interfaces or shielded balanced copper cables with a 8B/10B coding.

Twisted Pair Cabling

- Category X cables
  - Cat 3 (Voice grade)
  - Cat 4
  - Cat 5
  - Cat 5e (1000BaseT, unshielded)
  - Cat 6
  - Cat 7
- Category depends on twisting cycles per length unit, isolation, and shielding

(C) Herbert Haas 2005/03/11 26

The cables types used in networking are divided in different categories which determine the capability of a cable e.g. max. frequency, impedance, attenuation, crosstalk, etc.

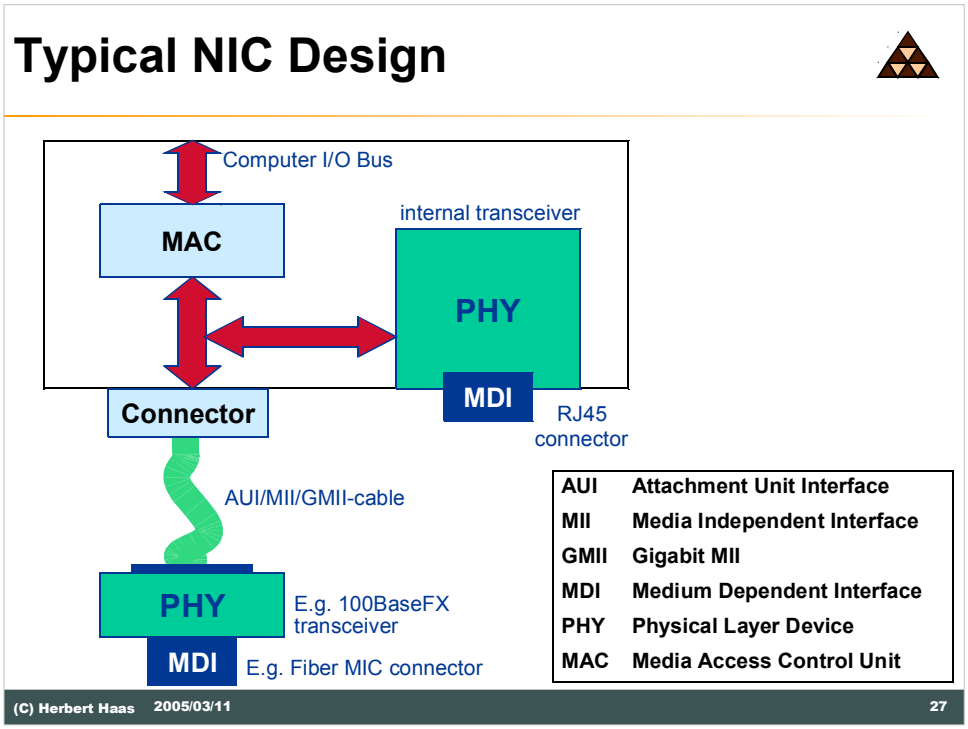
The CAT 3, 4, 5, 5e, 6 are specified by the T568-B standard published by the Electronic Industry Association and Telecommunications Industry Association (EIA/TIA).

CAT 7 cables are currently not covered by the standard but it is assumed that they will provide a bandwidth capacity of up to 400 MHz.

CAT 3……	16 Mhz
CAT 4……	20 MHz
CAT 5……	100 MHz
CAT 5e……..	100 MHZ
CAT 6……	250 MHZ

The Category 5e (CAT5e), or Enhanced Category 5, was ratified in 1999. It's an incremental improvement designed to enable cabling to support full-duplex Fast Ethernet operation and Gigabit Ethernet.

Like CAT5, CAT5e is a 100-MHz standard, but has stricter specifications for crosstalk, attenuation and return loss.



In this graphic we find a drawing about the principal design of a network interface card.

We find the MAC layer directly located on the Ethernet card which is responsible for the interaction between the physical and the Data-link layer. Then there is a physical interface directly located at the Ethernet card itself equipped with an RJ45 connector.

The AUI/MII/GMII connector represents a bus system for 10/100/1000 Ethernet systems used for media conversion with the help of an transceiver.

Summary

- Successful because simple
- Two frames: DIX (Ethernet2) and IEEE (802.3)
- Shared medium has consequences
  - Collisions ➔ Slot time ➔ Network diameter
  - Unpredictable, bad for realtime
- Increased data rate until today ➔ 10 GE already available (!)

Quiz

- What is a hub?
  List typical properties:
  - Half/full-duplex?
  - Different data rates?
  - Collision behavior?
- What is the canonical addressing format?
- What is a jam signal?
- What is 802.3u and 803.3z ?
- What is a runt? What is the opposite?