TIMICO

Real-time Ethernet

Christian Fischer Pedersen
Assistant Professor
cfp@eng.au.dk

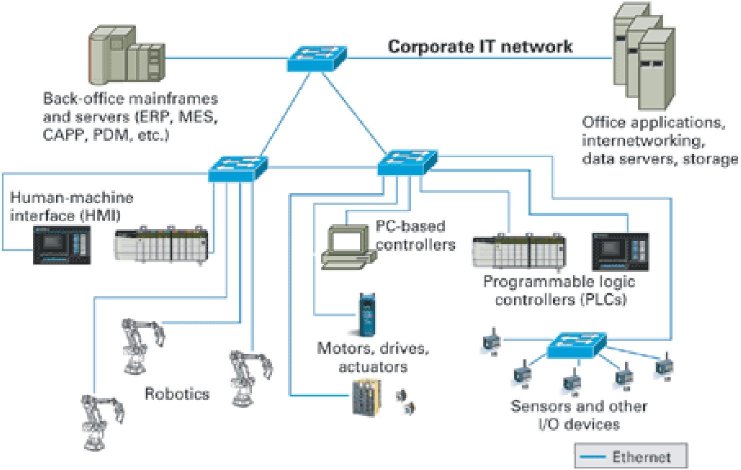
Section of Electrical and Computer Engineering
Department of Engineering
Aarhus University

November 3, 2014

Outline

Outline

Ethernet moves to the plant floor



Typical network requirements

Plant networks may span entire production facilities

- Connecting various nodes: office servers ↔ sensors/actuators

Control networks typically span plant floor

- Connecting specialized nodes

Device networks typically span parts of plant floor

- Connecting sensors and actuators

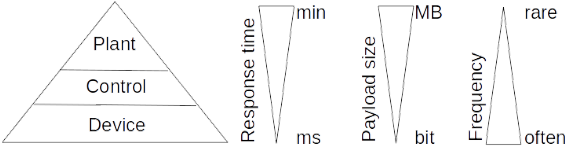


Figure : Typical network requirements at plant, control, and device levels

Ethernet convergence

Convergence in hardware and software standards

- More companies to support development
- All benefit from all
- Unifies and simplifies the wealth of field-buses
- Ubiquity reduces R&D, production, training, and maintenance costs
- Open standards may progress initiatives in open source projects and small/medium sized companies
- Get all the benefits of prior Ethernet R&D efforts for free

The OSI reference model

About the model

OSI (Open Systems Interconnection) model (ISO/IEC 7498-1)

- ISO: International Standards Organization
- IEC: International Electrotechnical Commission

Not a network architecture per se

- Comprises seven abstraction layers for communication
- Does not precisely specify services and protocols in each layer
- Just tells **in general** what each layer should do
- A layer serves the layer above it

Quite similar to the TCP/IP reference model

- Not a network architecture per se either

The OSI reference model

Ethernet's place in the model

	Layer	Data unit	Name	Function
Host layers	7	Data	Application	Meaning of data
	6		Presentation	Data representation and en/de-cryption. Machine de/inde-pendent data conversion
	5		Session	Interhost communication
	4	Segments	Transport	End to end connection, reliability and flow control
Media layers	3	Packet or datagram	Network	Determination of data paths within the network via logical addressing
	2	Frame	Data link	Data transmission via physical addressing, source, destination, and checksum
	1	Bit	Physical	Media (wire/fiber), signal connection, voltage levels, and binary transmission

Ethernet frame

Ethernet, IEEE 802.3

The OSI reference model

Communication in the model

- A layer serves the layer above it, i.e. layer L_n serves L_{n+1}
- Service Data Units (SDU) are passed down from L_{n+1} to L_n
- L_n encapsulates L_{n+1} SDUs into protocol data units (PDU)
- A L_{n+1} SDU is embedded in the payload of a L_n PDU
- SDU is data sent by a user of the services of a given layer, and is transmitted semantically unchanged to a peer service user
- L_n entity peers interact by the L_n protocol by transmitting PDUs

Ethernet frame structure

- Payload includes all other protocol headers and footers
- Payloads less than 46 octets are padded
- 802.1Q for VLAN and quality of service prioritization
- **Video (5:23): Renepick, How to build an Ethernet Frame, 2013**

Preamble	Start of frame delimiter	MAC destination	MAC source	802.1Q tag (optional)	Ethertype (Ethernet II) or length (IEEE 802.3)	Payload	Frame check sequence (32-bit CRC)	Interframe gap
7 octets	1 octet	6 octets	6 octets	(4 octets)	2 octets	42[note 2]–1500 octets	4 octets	12 octets
		64–1522 octets						
72–1530 octets								
84–1542 octets								

Figure : 802.3 Ethernet frame structure (Wikipedia).

Protocol overhead and efficiency

Ethernet protocol efficiency $= \frac{\text{Payload size}}{\text{Frame size}}$

- Good (nearly best) case: $\frac{1500\text{octets}}{1542\text{octets}} = 97.28\%$
- Bad (nearly worst) case: $\frac{46\text{octets}}{1542\text{octets}} = 2.98\%$

Other protocol headers in Ethernet payload, e.g.

- $\approx$ 8 (LLC) + 20 (IP) + 20 (TCP) = 48 octets

E.g. add 48 header octets to send 10 yields 58 octets ($>$ 46)

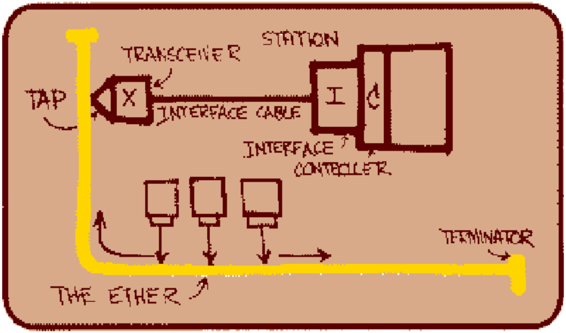
- Efficiency from payload perspective $\frac{10\text{octets}}{1542\text{octets}} = 0.65\%$

Ethernet timeline

- 1973: First versions of Ethernet developed at Xerox PARC
- 1975: Patent on Ethernet was filed by Xerox
- 1976: Ethernet deployed at PARC
- 1980: DIX (Digi. Equip. Corp/Intel/Xerox) std. v.1: 10 $\frac{\text{Mbit}}{\text{s}}$
- 1980: IEEE began R&D on LAN std. 802
  Different specs were promoted, e.g. CSMA/CD by DIX, Token Ring by IBM, Token Bus by GM
- 1985: IEEE std. 802.3 CSMA/CD was published. HDX
- 1990: In the decade, Ethernet became prevalent LAN tech.
- 1998: IEEE 802.3x introduced FDX
- 2011: Ethernet ubiquitous in home/industry: 0.1 and 1 $\frac{\text{Gbit}}{\text{s}}$

Ethernet layout anno 1976

Robert Metcalfe presents Ethernet to the National Computer Conference in June 1976



Video (9:14): The History of Ethernet, R. Metcalfe, 2007

Ethernet physical media examples

The original commercially available variant of Ethernet – no longer in use

10BASE5 (ThickNet)

- Coaxial cable with 9.5 mm diameter and 50Ω impedance
- Throughput of **10** Mbit/s
- **Base**band digital transmission
- Maximum segment length of **5**00 m
- Maximum 100 nodes per segment

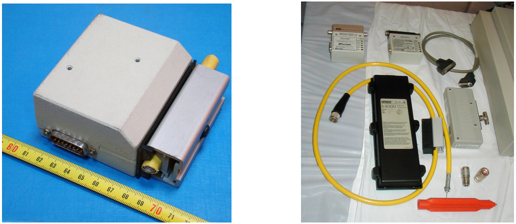


Figure : Transceivers, cables and tapping tool (Wikipedia)

Ethernet physical media examples

Other media examples

10BASE2 (ThinNet)

- Coaxial
- As 10BASE5 but cheaper and easier to install

10BASE-T (UTP)

- Full-duplex **U**nshielded **T**wisted **P**air
- Revision IEEE 802.3i completely new physical layer
- Also: 100Base-T/TX and 1000Base-T

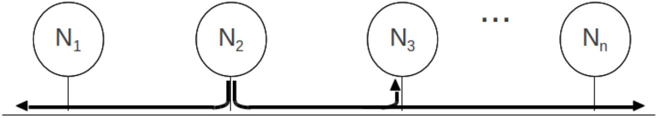
10BASE-FL

- Full-duplex fiber-optic cable
- Also: 100Base-FX & 1000BASE-SX/LX

Outline

Single segment Ethernet

- Originally single segment and HDX (< ca. 1990)
- CSMA/CD for sharing bandwidth on shared medium
- Speed is function of no. of simultaneously active nodes
- Collisions reduce throughput
- Packets are received by all. NIC filters per address
- During transmission: 1 node/time exclusive right to medium
- Damage to shared medium affects entire segment



CSMA/CD - Carrier Sense

Carrier Sense Multiple Access with Collision Detection

- Network interface card (NIC) listens for traffic on bus
- If traffic exist: Do not transmit
- If no traffic for a certain period (inter-frame gap): Transmit

CSMA/CD - Multiple Access

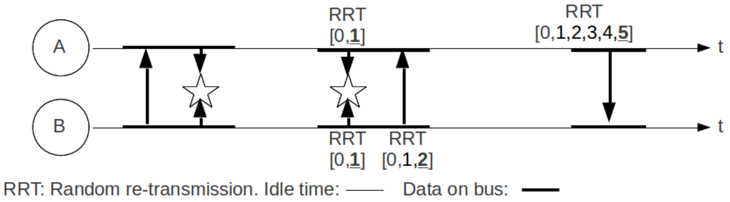
Carrier Sense Multiple Access with Collision Detection

- Multiple nodes on shared medium
- Collisions are highly likely
- All packets are passed by all nodes
- NIC filters: only addressee nodes receive data

CSMA/CD - Collision Detection

Carrier Sense Multiple Access with Collision Detection

- Collision occur if nodes send data simultaneously
- Nodes monitor own transmission
  - If node detects collision (high voltage signal)
  - Transmission stops and jam-signal broadcasted
  - Nodes discard packets and stops transmitting



CSMA/CD - Collision Detection

Carrier Sense Multiple Access with Collision Detection

Truncated binary exponential back-off algorithm

- To schedule retransmissions after detected collision
- Retransmit after a random number of slot-times:
  Max theoretical roundtrip time: $t_{\text{slot}} = t_{\text{data out}} + t_{\text{jam back}}$
- Back-off after $c \in [1; c_{\text{max}}]$ collisions: $t_{\text{back-off}} = r \cdot t_{\text{slot}}$
- Where
  - $r \sim \mathcal{U}[0, ..., 2^c - 1]$ where $\mathcal{U}$ is the uniform PMF
  - $t_{\text{slot}} = 51.2\mu$s at 10 Mbit/s
  - $t_{\text{slot}} = 5.12\mu$s at 100 Mbit/s
- Truncated means
  - Ceiling on retransmission timeout
  - IEEE 802.3 CSMA/CD standard: $c_{\text{max}} = 10$ (IEEE, 2008 [3])
  - Transmission canceled after 16 successive attempts and back-off interval reset to initial value

Medium monopolization by back-off algorithm

- Two nodes A and B have a number of packets to transmit
- At some point a collision occur and, e.g. B randomly selects 0 and A selects 1 in [0;1]
- Thus B will be successful in first retransmission
- B still has more packets to send and tries immediately to send again. By chance this collides with A
- A doubles the back-off interval whereas B uses the initial one. Thus B has higher chance of transmitting successfully
- This may repeat and B may monopolize the medium

Summary of single segment Ethernet

Pros in regard to real-time operations

- Low cost, flexible, and ubiquitous hardware

Cons in regard to real-time operations

- CSMA/CD w/ random back-off algorithm
- Nondeterministic medium access timing makes it hard to guarantee deadlines
- No packet prioritization means that time-critical packets may yield for non-critical packets

Perspectives

- Methods to counter the cons exist

Outline

Motivation and background

Single segment Ethernet

Multi segment Ethernet

Single segment real-time Ethernet

Multi segment real-time Ethernet

References

Main functionalities of a switch

Micro-segmentation

- Separate collision domain for each network segment
- Only two nodes coexist in a collision domain
- Nodes have dedicated point-to-point bandwidth
- Full duplex without collisions

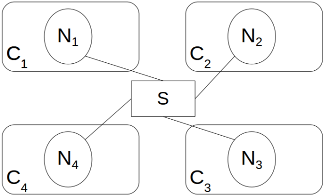
Figure : Industrial Ethernet switches from Molex

Micro-segmentation by switch

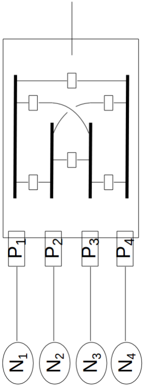
Single segment vs. switched Ethernet

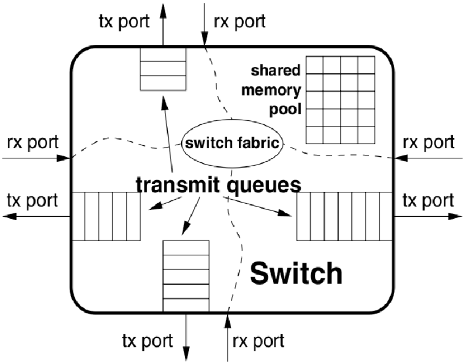
- One collision domain, C, for multiple nodes, N
- Collisions are possible
- Non-deterministic

- One collision domain, C, per node, N, due to switch, S
- Collision free
- Deterministic



Conceptual switch internals





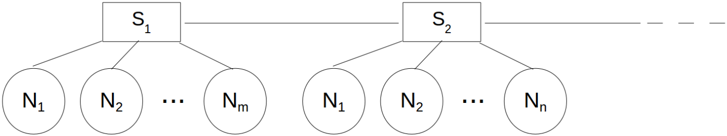
Switched Ethernet

Fully switched means full-duplex Ethernet

- Collision-free: Only 2 nodes per collision domain
- Net speed is x bit/s $\Rightarrow$ node-to-node is 2x bit/s

Cabling

- Twisted-pair or fiber optic
- Separate transmit (Tx) and receive (Rx) conductors



Switching methods on port basis

Store and forward

- Entire packets buffered and checksum verified
- Corrupted packets dropped

Cut through

- Reads only up to destination address before forwarding
- Forwards corrupted packets as checksum is not verified
- Higher speed than store and forward

Adaptive switching

- If error rates low: Cut-through mode
- If error rates high: Store-and-forward mode

Characteristics of switched traffic

Corruption

- Bit errors by noise or interference
- Drops and retransmits packets in store and forward

Latency

- Communication delay in media, e.g. wires, switches, routers
- Delay may build up even if throughput is approx. constant
- `Ping` command measures round-trip (NOOP) latency

Packet delay variation (PDV)

- Low PDV: RT system may deal with latency by time offsets
- High PDV: Unpredictable and hard to guarantee deadlines

Latency and packet loss example

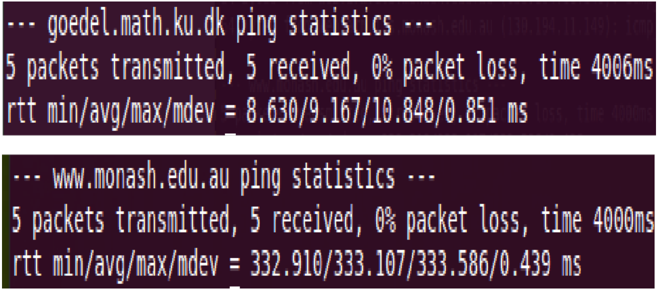


Figure : `Ping` statistics

Typical switch data

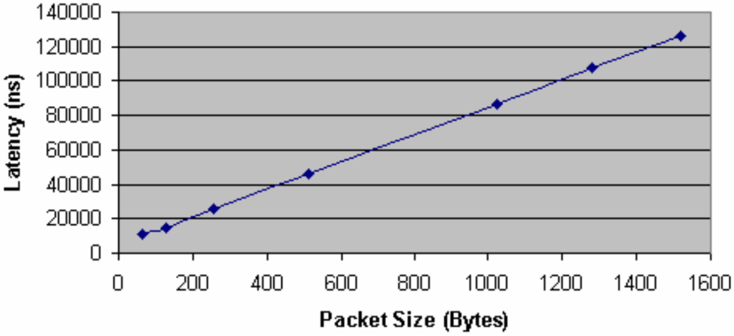
Port to port communication

- Less than 50 μs
- Depends on packet size

Packet delay variation

- Approximately 100 ns

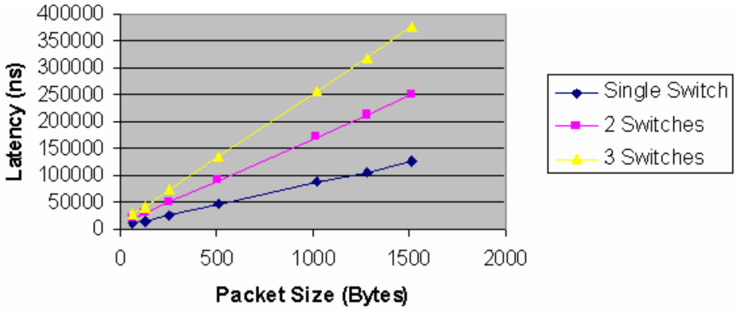
Latency: Single 100Mbit/s switch



- Min: (packet size, latency) = (64 Bytes, 10840 ns)
- Max: (packet size, latency) = (1518 Bytes, 126240 ns)

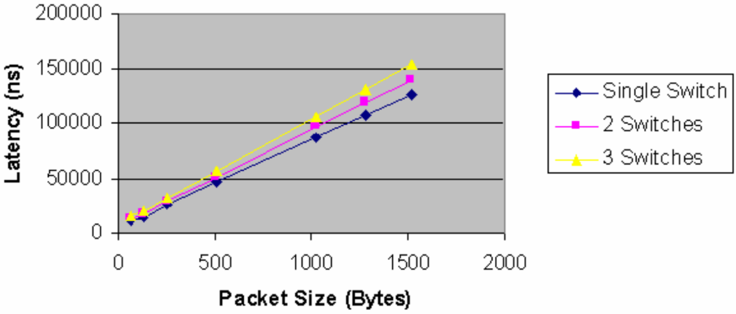
(Coley, 2004 [1])

Latency: Multiple 100Mbit/s switches



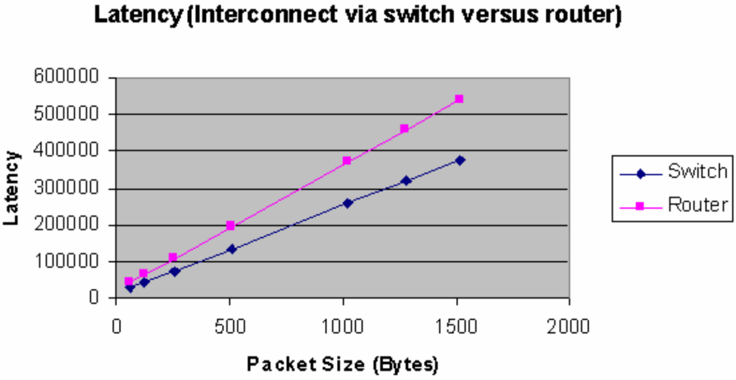
(Coley, 2004 [1])

Latency: Multiple Gbit/s switches



(Coley, 2004 [1])

Latency: Connecting two switches



- Routers read the address for proper routing

(Coley, 2004 [1])

Nagle's algorithm

Increases bandwidth at the expense of latency

- Combines outgoing packets with small payloads
- Decreases overhead/pay-load ratio
- Good for normal Ethernet operations

Poor compliance with real-time systems

- Increases latency and packet delay variations

Disable / enable the algorithm

- Disable it: TCP_NODELAY socket option
- E.g. Unix nc6 (netcat): Swiss army knife for TCP, UDP, ...
- "Client": nc6 –disable-nagle -l -p 1234
- "Server": nc6 127.0.0.1 1234

Summary of multi segment Ethernet

Pros in regard to real-time operations

- Fully switched gives full duplex
- No collisions: only two nodes in same collision domain
- Network segmentation via logical overlay virtual LAN
- Low cost, flexible, and ubiquitous hardware

Cons in regard to real-time operations

- Deadlines hard to guarantee
- Arrival times of packets hard to predict
- Ethernet frames vary in size
- Unknowns: Route, latency, packet delay var., corruption ratio

Perspectives

- Methods to counter the cons exist

Outline

Single segment real-time Ethernet approaches

Main challenge

- Collision avoidance on shared medium

Common approaches

- Token based
- Time slot based
- Statistics based

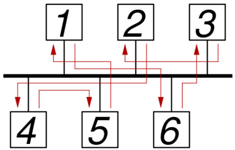
See (Moraes, 2010 [4]) for a survey on real-time in CSMA networks

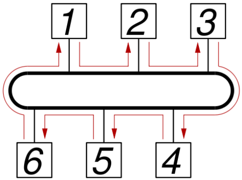
Single segment real-time Ethernet

Token based

Token represents permission to transmit

- Token circulated according to strategy
- Token management uses time and bandwidth
- Only one node may transmit at any time
- Limited performance vs. switched networks





Single segment real-time Ethernet

Time slot based

Time slot represents permission to transmit

- Time is divided according to strategy (TDMA)
- Strike a balance on time-slot lengths
  - Long: Long idle periods. Short: Frequent data partitioning
- Time management uses time and bandwidth
- Limited performance vs. switched networks

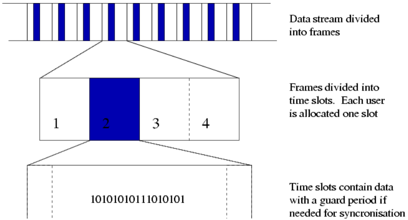


Figure : Stream into frames and time slots with data+guard (Wikipedia)

Single segment real-time Ethernet

Statistical based

- Keep network traffic low to bound probability of collisions
- Statistical guarantee to nodes for transmission time and bandwidth
- The overall network utilization decreases with stronger statistical guarantees
- Basically avoidance of invoking the back-off algorithm

Outline

Motivation and background

Single segment Ethernet

Multi segment Ethernet

Single segment real-time Ethernet

Multi segment real-time Ethernet

References

Multi segment real-time Ethernet approaches

Main challenge

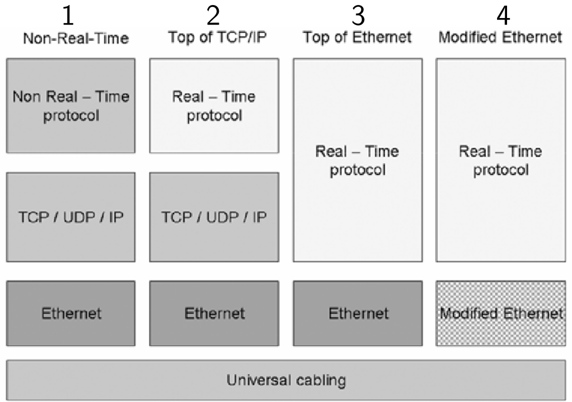
- Bandwidth control and scheduling, e.g. in switches

Common approaches

- ProfinetIRT
- EtherCAT
- Ethernet Powerlink
- Sercos III
- EtherNet/IP

Classification of multi segment RTE approaches

By protocol stacks

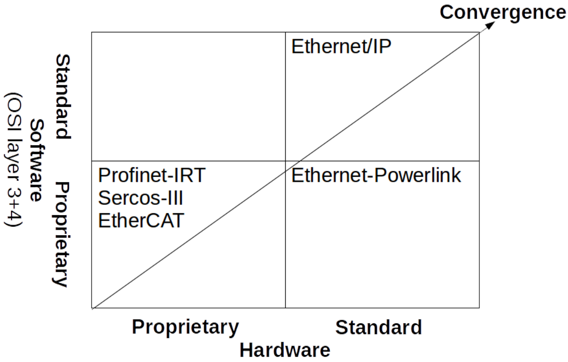


(Felser, 2005 [2])

1. No RTE overlay
2. Ethernet/IP
3. Ethernet Powerlink
4. Profinet v3, Sercos-III, EtherCAT

Classification of multi segment RTE approaches

By hardware and software compliance



Classification of multi segment RTE approaches

By bandwidth control methods

Separate slave polling

- Ethernet Powelink, ProfiNet IRT, Sercos-III

Combined slave polling

- EtherCAT

Time based

- Ethernet/IP (based entirely on standard Ethernet)

Ethernet Powerlink, ProfiNet IRT, and Sercos III

Separate slave polling

Per cycle

- RT traffic: Master polls each slave with an Ethernet telegram
- Non-RT traffic: Remaining cycle time

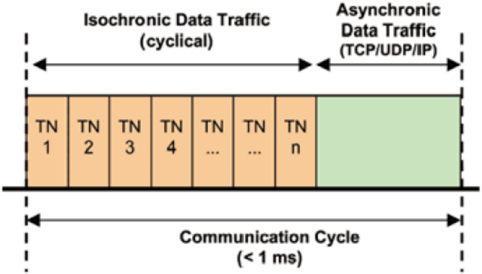


Figure : Time slot mechanism

Ethernet Powerlink, ProfiNet IRT, and Sercos III

Separate slave polling

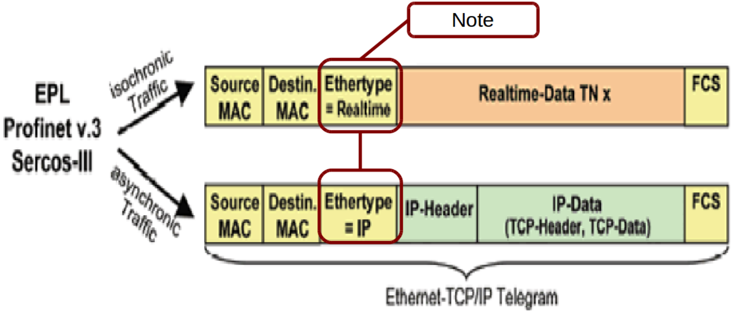


Figure : Telegram structuring

Ethertype

- ID of protocol encapsulated in the Ethernet frame payload

EtherCAT

Combined slave polling

Per cycle

- RT traffic: Master polls slaves with an Ethernet telegram shared by all
- Non-RT traffic: Space reserved as a sub-telegram in the shared telegram

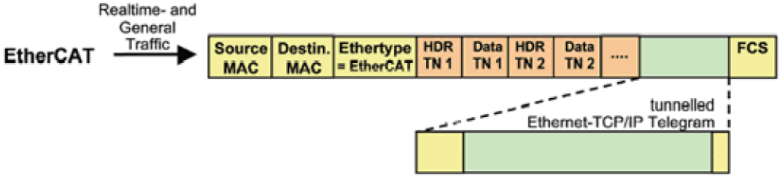


Figure : Telegram structuring

EtherCAT

Combined slave polling

- Designed for small payloads per node
- One Ethernet frame holds more sub-frames
  - Encoded and decoded by special hardware
  - Slaves read and write in same sub-frame on the fly
- Each with payloads for nodes in the network
  - Reduces overhead/payload ratio
- Synchronization via distributed clock
  - Very low packet delay variations

EtherNet/IP

Time based

Ethernet/IP

- Based entirely on standard Ethernet
- Synchronized clocks via Precision Time Protocol
- Control commands must be received by slaves in due time
- RT delivery by master by UDP
- QoS prioritization of traffic (802.1Q)

EtherNet/IP

Time based

Preamble	Start of frame delimiter	MAC destination	MAC source	802.1Q tag (optional)	Ethertype or length	Payload	Frame check sequence (32-bit CRC)	Interframe gap
7 octets of 10101010	1 octet of 10101011	6 octets	6 octets	(4 octets)	2 octets	46–1500 octets	4 octets	12 octets

Figure : QoS prioiritization in Ethernet frame; switches prioritize traffic

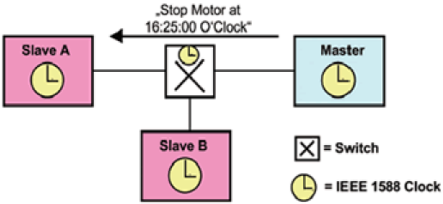


Figure : Precision Time Protocol for clock synchronization

Comparison of approaches in real-time setup

- Setup: 100 axles to be controlled synchronously
- Response time: Cycle time
- Jitter: Variations in response time
- Conclusion: More proprietary → Faster and less jitter

Organization	Response time (for 100 axles)	Jitter	Data rate
Ethernet/IP CIPSync ODVA	≈1ms	<1ms	100Mbit/s
Ethernet Powerlink EPSG	<1ms	<1ms	100Mbit/s
Profinet-IRT PNO	<1ms	<1ms	100Mbit/s
Sercos-III IGS	<0.5ms	<0.1ms	100Mbit/s
EtherCAT ETG	≈0.1ms	<0.1ms	100Mbit/s

Comparison of approaches in real-time setup

- Setup: Field bus vs. Ethernet
- No. of slave pollings measured in no. of packets/ms
- Performance criteria: Theoretical packet throughput (actual throughput lower due to processing and latency)
- Conclusion: New Ethernet based systems perform well

	Transfer rate	Packet size *with 8 bytes of user data*	Theoretical packe throughput
Profibus DP	12 Mbit/s	19 bytes	79 packets/ms
Sercos-II	16 Mbit/s	14 bytes	142 packets/ms
Ethernet	100 Mbit/s	72 bytes	174 packets/ms

Outline

Further reading

Online material

- Industrial Ethernet University
- Industrial Ethernet Book
- IEEE Standards Association
- ODVA (Open DeviceNet Vendors Association)
- Ethernet Powerlink Standardization Group
- EtherCAT Technology Group
- Fieldbus Foundation
- Modbus Organization
- Profibus and Profinet International
- Sercos – The Automation Bus

Further reading

Papers and books

- Snader, J.C., "Effective TCP/IP Programming", Addison Wesley, 2000
- Lammermann, S., "Ethernet as a real-time technology", Univ. of Telecomm., Leipzig, 2008
- Marx, U. and RTnet development team, "RTnet project", Univ. of Hannover, Germany, 2011

References I

[1] Coley, K. (2004). Using intelligent ethernet switches to enhance data packet delivery speed in Ethernet/IP networks. In *CIP Networks Conference*.

[2] Felser, M. (2005). Real-time ethernet – industry prospective. In *Proceedings of the IEEE*, Volume 93.

[3] IEEE (2008). IEEE 802.3-2008, part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications. Rev.: 3rd printing 22 June 2010. URL: http://standards.ieee.org/about/get/802/802.3.html. pp. 1-315.

[4] Moraes, R. (2010). Survey of real-time communication in CSMA-based networks. In *Network Protocols and Algorithms*, Volume 2.