Real-Time with Ethernet

R. Messerschmidt
Otto-v.-Guericke University
Institute for Ergonomics, Manufacturing Systems and Automation
Center Distributed Systems
Universitätsplatz 2, D-39106 Magdeburg, Germany
ralf.messerschmidt@mb.uni-magdeburg.de

Abstract

After a general real-time consideration and classification the opportunities of Ethernet concerning real-time will be discussed, especially the IAONA Real-Time classes will be viewed and achievements of Ethernet components and features are considered in general.

Real-time types

Although the term real-time is often used in the field of Automation and Communication, an academic definition of real-time or even a numeric specifications for real-time behavior can not be given in general but only in close context with a special application of a distributed control. Is a system in all circumstances able to react to all occurring events correctly and timely then it is real-time capable [1]. If a communication system meets all the timely requirements of a certain application it is – related to this application- real-time capable.

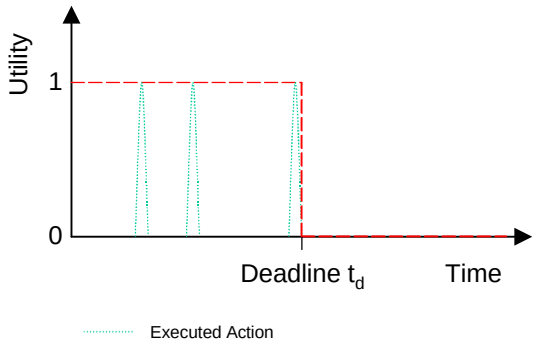

Figure 1: Time/Utility function with an execution deadline (timeliness)

This might mean that an action has to be completed until a maximum time line – the deadline (see Figure 1). This behavior can be described with timeliness.
But often this means too, that actions have to happen on a certain pre given moment (see Figure 2). This behavior can be described with simultaneousness [2].
According to Douglas Jensen´s Time/Utility Function Model [3] of real-time the first case means that the utility of completing an action is 1 until the deadline. The other case means the utility is only 1 in a small window of time. Time/Utility Functions are a more general formulation of time constraints. They express the utility of executing (and completing) an action as a function of when the action is executed (and completed).

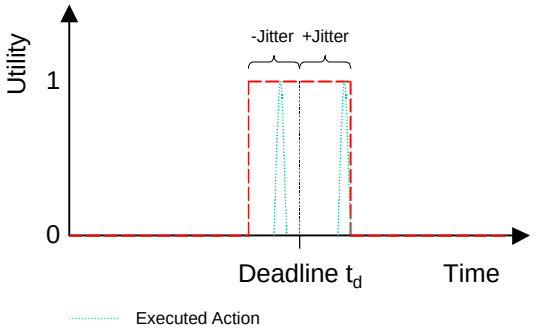

Figure 2: Time/Utility function with execution window (simultaneousness)

The utility values express the relative importance of an action.

Real-time types and Ethernet

In the first case of above described time constraints Ethernet TCP or UDP/IP is an appropriate protocol.
The second above described case can generally not be guaranteed with Ethernet TCP or UDP/IP, variations in the transmission duration (so called jitter) can be

caused by unpredictable delays in buffer queues that are in general too large for those applications. However, this can be overcome by combining guaranteed maximum transmission times (worst case studies) and an appropriate exact time synchronization mechanism. Such synchronization algorithms that will provide a synchronization exactness far beyond one millisecond currently are developed by the IEEE working group 1588.
IEEE 1588 synchronization will provide a technology to synchronize the internal clocks of distributed controls. So commands sent between nodes can be equipped with a time stamp of the desired execution time, thus actions can be executed simultaneously or coordinated on a common system time base. So not the communication is synchronized in general (even if this could also be realized by this technology) but the execution time directly is synchronized.
Synchronized communication is usually realized by time slot technologies, where the right to communicate is allocated cyclic to each node (e.g. by a master or a round passed token) so that nodes of a net can never communicate at the same time and the time difference between sending a command or data by a station and the receipt (and so also the execution) by an other station can be calculated or is pre given. So the nodes work synchronous without a shared common time base.

IAONA Real-Time Classes

The Industrial Automation Open Network Alliance IAONA pursues the aim of establishing Ethernet as the standard application in every industrial environment at an international level. Sense of this is to realize a general, interfaceless communication through all levels of an enterprise.
Concerning Ethernet TCP/IP the IAONA Joint Technical Working Group Hard Real-Time has defined 4 real-time classes that define real-time demands including the aspect of development needs.

Class 1 covers properties that are still available with standard Ethernet products. Class 2 covers products conform to today's standards that are optimized for real-time demands. Class 3 are products with new added functions realized in software combined with standard hardware (of course also this class is down compatible to the standards, but the new added functions can only be used by devices which are prepared for these added functions). Class 4 products have the added new functionality of class 3 but realized in hardware.

Stack optimization

On the shelf stacks today are optimized for TCP not for UDP. So according to IAONA Class 2 an optimized UDP/IP stack, that mostly are a component of the applied real-time operating system, would improve the time behavior for most industrial Ethernet products, since real-time data are usually transmitted using UDP.
Investigations of typical operating systems showed that stacks, as used today, have relatively high throughput times and the fluctuating of throughput times are around factor five. Transmission times on the wire are comparably short on Ethernet so most potential for improvement is in optimizing the processing time in stacks. So for instance a "zero-copy-stack" copies new data only once in a buffer and all the protocol layers of the Ethernet TCP/IP stack access the data at that buffer and not as usual where the data are copied in different buffers for each protocol layer.

Switches

Most industrial Ethernet systems require a switched network with components working with fullduplex to prevent collisions that otherwise – in the case of a shared medium – can occur. In a shared medium end devices must compete for access to the medium. The well know technology for that in Ethernet is CSMA/CD. Switching technology combined with fullduplex avoid the contention for access to the transmission medium, because each end device has one link to the switch for transmitting data and one link for receiving data.
In opposite to an hub - the traditional device for connecting devices over Ethernet - a switch does not mirror data coming in on one port to the outputs of all other ports. A switch knows the addresses of the devices connected on its ports and delivers data only to the addressed target device. Thus a switch is able to provide several connections at the same time. But what happens if , at the same time, data packets from several ports are addressed to the same destination? In that case the data packets are stored in the output queue of the port connected to the target device and will be transmitted to the target device one by one.
Also it has to be considered that he time to transmit packets depends on the packet size so the larger differences in packet sizes the larger the differences in transmission times.

In addition to control data, which requires real-time communication capability, additional data with different load profiles and characteristics must use the network. For example, visualization data, software

updates, e-mail traffic, office applications, and Internet data traffic. For this reason the network must be meticulously designed, to include segmenting those parts of the network where real-time behavior is necessary.
The terminal devices that require real-time behavior should be linked over as few switches as possible. Inevitably, the more switches between two terminal devices, the higher the “worst case” throughput and queue time. With backbones or other instances where there are no factors limiting real-time performance, the individual segments are commonly connected in a ring structure.
In addition, the interface between a real-time segment and the rest of the network must be precisely controlled. Since the data traffic from the general network can adopt any load profile, it must be monitored and restricted when entering a real-time segment. To prevent the real-time segment being overloaded, the amount of data traffic entering this segment must be limited. An effective way to achieve this is to configure the inter-segment link to 10Mbit/s, while all devices on the real-time segment communicate at 100Mbit/s. Further segmentation, as well as access control, can be accomplished by the use of routers and firewalls.

A recent innovation in Ethernet, standardized by the 802.1p working group, is a prioritization mechanism. An additional field, known as a tag, is added to the Ethernet frame. The tag contains information about the priority of the data.
Some Ethernet switches already support this function, Each transmission port has separate queues for the supported priority levels. Data packets in a higher priority queue are always transmitted before those in a lower priority queue.

In systems with strict time constraints the communication rise is usually known so worst case considerations can be done.
By that way systems with the above described execution deadline can be realized.

Ethernet with time slot technology

There are also technologies that provide an already described time slot technology over Ethernet. Since in each time slot only one device can send data there is no need for using switches to prevent from collisions. The use of hubs even improve the time behavior of those systems since hubs are faster than switches, because they do not process the data of the packets and so need less time than switches that have to process incoming data packets to find out its destinations.

As already mentioned above with this technology applications with simultaneousness action execution demands can be realized.

References

[1] F. Furrer: Ethernet-TCP/IP für die Industrieautomation: Grundlagen und Praxis. Huthig Verlag, Heidelberg, 2. Auflage, 2000.

[2] G. Gruhler: Feldbusse und Gerätekommunikationssysteme. Steinbeis- Transferzentrum Automatisierung (STA), Reutlingen, 5. neu gestaltete und erweiterte Auflage, 2000.

[3] E. D. Jensen: Overview of Fundamental Real-Time Concepts and Terms, 2002.