Robotics and Computer–Integrated Manufacturing 49 (2018) 309–316

Contents lists available at ScienceDirect

Robotics and Computer–Integrated Manufacturing

journal homepage: www.elsevier.com/locate/rcim

Hard real-time communication solution for mechatronic systems

Andrzej Przybył

Częstochowa University of Technology, Institute of Computational Intelligence, Al. Armii Krajowej 36, 42-200 Częstochowa, Poland

ARTICLE INFO

Keywords:
Real-time communication
Industrial distributed control systems
Computer numerical control

ABSTRACT

The paper proposes a method to build a highly efficient real-time communication solution for mechatronic systems. The method is based on the Ethernet physical layer (PHY) and on field programmable gate array (FPGA) technology and offers a better performance when compared to commercially available communication solutions. Although it is not directly compatible with the OSI/ISO model of TCP/IP protocol, vertical integration is done with a gateway. This provides simplicity and safety. Moreover, the use of the FPGA allows for integrating the communication solution with the user algorithm of particular distributed device inside a single chip. Therefore, the proposed solution is efficient and highly integrated.

1. Introduction

In real-time communication the most important goal is to deliver messages from their source to the destination in a deterministic time. Moreover, the delivery time must be also as short as possible to meet the requirements of the control system. Currently the industry is using more and more often real-time Ethernet (RTE) communication solutions because of their higher performance when compared to the existing fieldbuses. At present there is not one common defined RTE standard and many different and non-compatible solutions are used [6]. The research on the real-time communication over Ethernet is still ongoing [5].

It should be noted that there is a broad range of real-time communication systems which are based on the Ethernet physical layer (PHY) but are not compatible with higher layers of the OSI/ISO model of TCP/IP protocol. Such systems are commonly called the real-time Ethernet, while they are in fact the Ethernet PHY-based fieldbuses (let's call them EFbus). The EFbuses have typically very high real-time performance but they are not compatible with off-the-shelf Ethernet hardware and software [16,18,26].

One of the possible future scenarios in the development of the real-time communication systems is the existence of universal multi-protocol network devices. Such versatility is possible, especially when network devices are based on the FPGA and the Ethernet PHY devices. The second future scenario is the coexistence of different and not compatible systems side by side at the factory level [32,33], connected via gateways. Only time will tell which solution will be widely accepted.

The paper proposes a significant extension and an improvement of a general idea introduced in [18] to build a hard real-time Ethernet-PHY-based solution for industrial control systems. Primarily, an important modification of the previously used classic version of the time-division-multiple-access (TDMA) mechanism is presented, i.e. time-division-multiple-access with latency time compensation (TDMA-LTC). Secondly, an efficient frame fragmentation mechanism is proposed. In addition, extended and more detailed simulation and analysis of the timing dependencies as well as the experiment results are presented.

The real-time communication efficiency of the proposed method is better than in commercially available solutions. The method is particularly well suited for systems where real-time performance (i.e. the horizontal aspect) is the most important [8]. It could be used as a hard real-time backplane bus for a machine or a process control.

Despite the fact that the proposed solution is a type of EFbus and it is not compatible with the OSI/ISO model of TCP/IP protocol, the vertical integration is achieved by means of an application layer gateway. As indicated in papers [22,25,28], using the gateway is a good way to cope with more and more common security problems. Moreover, gateways enable the use of different (i.e. non-compatible) communication solutions in the factory for monitoring and control of technological processes in many production plants [15].

This paper is organized as follows: Section 2 presents the short analysis of commercially available Ethernet-based communication solutions and their characteristics. The idea of the proposed method is described in Section 3. In Section 4 the performance analysis is presented and finally conclusions are given in Section 5.

2. The analysis of the available real-time Ethernet-based communication solutions

Currently on the market there exist many different real-time industrial networks which are able to fulfill different requirements for differing application areas. However, their users want to increase func-

E-mail address: andrzej.przybyl@iisi.pcz.pl

http://dx.doi.org/10.1016/j.rcim.2017.08.001
Received 17 May 2016; Received in revised form 7 July 2017; Accepted 8 August 2017
Available online 17 August 2017

tionality and diagnostics available for the network. The examples are: shortening of a communication cycle to improve quality control, making it possible to close the controller loops over the network [4,11,14] efficient connection with hardware emulators [18], recording of the data of a rapidly changing phenomenon e.g. for identification of the processes, etc. This means that research on the improvement of the real-time communication solutions is still going on.

Current RTE systems are typically based on the high performance 100Mbit/s physical medium, which currently appears to be sufficient for most of the increasing demands. However, in some cases much of the theoretically available bandwidth is wasted because of the inefficient message exchange model. Increasing the network baud rates to the 1 Gb/s, i.e. migration from Fast Ethernet to Gigabit Ethernet, improves only a part of the network parameters [10,19] and at the same time generates higher production and maintenance costs of such a system [20]. The most important factors affecting the real-time network performance are: the network message exchange model (interaction model) and the method used for accessing the shared medium.

The other aspect is that most networks which operate on the basis of standardized Ethernet have a major drawback: they generate a large overhead for communication for small data packets. This overhead is related to the minimum frame length equal to 64 bytes plus 8 bytes of preamble and start frame delimiter (SFD) plus 12 bytes length of inter frame gap (IFG) as defined by the IEEE 802.3. A typical message in the distributed control systems has about 8 bytes or even less. Taking into account the above data, available bandwidth (i.e. 100Mbit/s for Fast Ethernet) is used very inefficiently and it is equal to about ten percent. As shown in [18], the limitation of the minimum frame length, as well as required IFG, can be omitted if the full-duplex mode is used. One of the examples is the VARAN Bus [26] in which the important increase in the network efficiency for small data packets is obtained. Another very interesting project, which is still under development and based on the use of short frames, is a VABS (Very High Performance Automation Bus System) [23]. As indicated in the last paper, conventional RTE networks are confined to certain performance limits and a new approach, which uses its own specified data link layer protocol offers a performance advance.

An effective method to handle short messages in standardized Ethernet is to use so-called summation frames and on-the-fly mechanism introduced by Beckhoff and used in EtherCAT [9] and Sercos III [7]. Another very efficient method for the treatment of short messages is used in Profinet IRT [17,24]. This method is called dynamic frame packing (DFP). The above networks use full-duplex communication, and short messages are not burdened with a high overhead. However, there are some restrictions related to the solutions that uses summation frame. They (e.g. EtherCAT and Profinet IRT with DFP) will only work in line topologies. Although, topologies other than line are allowed in EtherCAT, they are processed internal exactly as a line topology. This introduces an unfavorable delay in data transmission via a large quantity of network devices. Moreover, in systems with long linear paths the growth rate of time synchronization error is the major barrier to the scalability of systems [30].

Summing up, the analysis presented above shows that existing solutions do not fully exploit the possibilities offered by the Ethernet medium. It is possible to develop new, more effective methods of communication, tailored to the requirements of distributed real-time control systems.

3. The method to maximize the real-time performance of Ethernet-PHY-based distributed systems

To make the proposed solution highly efficient the specified data link layer protocol is used. This protocol is able to handle up to 256 distributed nodes, but it is mainly dedicated to the system with dozens of nodes and a short communication cycle. Fig. 1 shows the time-space diagram of the proposed real-time communication for an example network

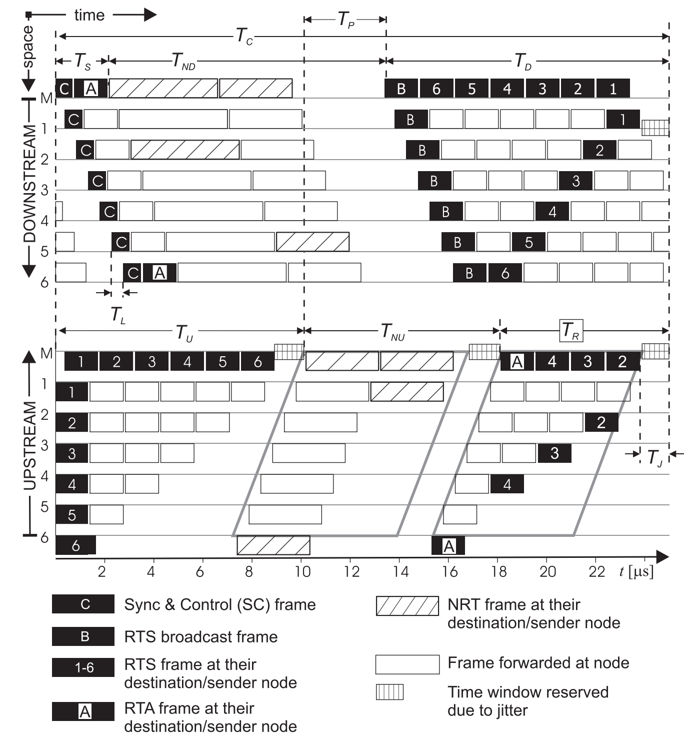

Fig. 1. The detailed view of the proposed real-time communication solution presented in the form of a time-space diagram for example of the six-nodes network.

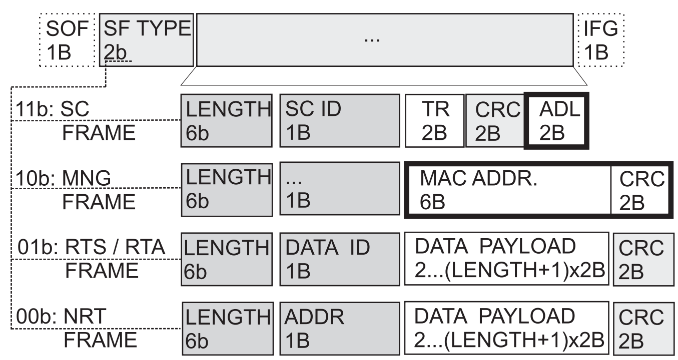

Fig. 2. The format of the short frames used in the proposed system.

with six nodes, connected in a line by daisy-chaining. In the proposed method the communication is organized in cycles with time period T_C. The communication cycle is in turn divided into several time windows, based on the TDMA method. These time windows are defined on the basis of requirements of a particular application. Some of them, i.e. T_U, T_P and T_D, are constant and they are set at the configuration stage, while the others are adjusted to best match the requirements of each successive communication cycle.

The use of the TDMA method requires that all distributed nodes are synchronized. In particular it is required to properly indicate the start of the communication cycle to allow all nodes to start sending their frames at the same time. The synchronization with jitter value T_J less than 1 µs is well known and worked out issue [3,13,29] and it is out of the scope of this paper.

In the proposed solution it was assumed that very short frames, designed in a way to minimize the communication overhead for small data packets, will be used (Fig. 2). As a result, it allows for a very effective communication with extremely small protocol overhead $m = 6$ bytes, which includes preamble, header, CRC16 and inter frame gap. This means that the proposed system is not compatible with the OSI/ISO