A Scalable System for Real Time Video Transmission over Industrial Communication Networks

F. PÉREZ, M. MATEOS, P. IRIONDO AND D. ORIVE
Department of Automatic Control and Systems Engineering
University of the Basque Country
Alda. Urquijo, s/n, 48013 - Bilbao
SPAIN

Abstract: This paper presents the specifications for developing a system with the ability of transmitting video through an industrial network, without interfering with the high-priority data transmitted through that network. This can be achieved dynamically by adapting the video quality to the bandwidth available at each time instant. This system has been implemented and tested in an Industrial Ethernet network.

Key-Words: Scalability, Dynamic Control, RTP, RTCP, Real-Time, Industrial Networks

1 Introduction
The growing demand of data transmission in the present production and process control systems is generalizing the intensive use of communication networks in industrial environments.

This type of networks facilitates the following basic services: control, configuration and data storage. At first, the control service implies a critical information exchange among the different devices; therefore the networks should provide services of priority and interruptions. At second, an industrial network should allow the configuration of the present devices in the system through it. And finally, it should allow the data collection for the purpose of carry out the analysis or control of them [1].

Last years it is emphasizing the use of Industrial Ethernet networks in the factory automation [2]. There are already a number of standards used in the field level, including PROFInet in Siemens, EtherNet/IP in Rockwell, Modbus/TCP in Schneider Electric or SERCOS in Bosch Rexroth. This type of networks arose with the intention of comparing the telematic networks with the industrial communication networks. The main reason of the Ethernet success is the possibility that technology offers to use different types of communications protocols. The most important one is TCP/IP.

When Ethernet arose in 1973, the real number of devices that could be connected to a LAN network was very limited, and the performance was not efficient enough. In order to solve these difficulties “switching Ethernet” appeared in 1994 (IEEE 802.1D). The technology based on switches permitted obtains rates of 100 Mbps (Fast Ethernet). However the principal difficulty was in the determinism. Ethernet is not deterministic, and this implies that the use of Industrial Ethernet as a control network is impossible from a realistic way. Nevertheless, the improvements introduced last years on prioritization services and commutation techniques (IEEE 802.1Q) which permits delay times lower than 7 µs, are beginning to allow the use of Ethernet in control systems.

One of the main advantages that the use of Industrial Ethernet provides is that it allows to adding new functionalities to the network without varying its topology. Moreover, thinking about the common amount of control traffic among the devices on an industrial network, these functionalities don’t damage the network main services either. Considering that the traffic in a typical network with more than 30 controllers will occupy, in the worst case, 19% of the total 100Mbps network bandwidth, we have more than 80% of it to add new functionalities [3].

The platform analyzed in this paper is able to introduce those new functionalities to the network, without damaging the information of priority process.

2 Platform Architecture
The platform can be divided in four large blocks that mark clearly the development of it. The four blocks mentioned are described here:

• Cell controllers, supervision system: It is a system that allows to access directly to the device process...
and control data using a server or a cell controller (without needing of direct access).

• Video compression and codification system: It is a system responsible of codifying the signals that arrives from the video cameras allocated next to the industrial devices and compressing it depending of the amount of information requested by the system, so that the video data never complicates the transmission of control data. This system consists of the video camera (or video cameras) and the video server equipment.

• Real time video information transmission system: Is the responsible of transmitting all the information through the network, both control devices and video devices. It must be capable of giving all the real time features to the transmission. In the same way, it must interact with the other three modules. It consists of video server equipment, the intermediate equipment and the platform networks.

• Representation and remote control system: It is the responsible of visualizing to the operator the video received through the transmission. In the same way, it must provide a graphic interface that permits a simple interaction with the platform. It is situated on remote equipment.

3 Real-Time Transmission System

This system looks after collecting the data packages from two of the systems (the control data from the cell controllers supervision system and the video data from the video compression and codification system), and adapting them to the network so that they can be transmitted with real-time features. It also will indicate the level of compression required in each moment to the compression and codification system, so that the priority control data won’t be damaged by this video transmission. Finally, it interacts with the representation and remote control system to facilitate the video transmission on the remote equipment and to collect the control data originated in the system.

3.1 Real time

The first and clear objective is the design and development of a system capable to transmit information with real time features through an industrial network.

In this environment is possible to talk about two types of real time: strict real time for deterministic systems and soft real time for guaranteed continuity. [5]

In deterministic systems the loss of some time limit of execution can produce serious consequences due to the need of immediacy in the answers requested by the own system.

In soft real time systems, like the video transmission system, only a sensation of continuity in the destiny should be guaranteed. That is, it must be assured a maximum delay between the frames and a viewing with constant period faithful to the reality in the destiny.

3.2 Scalability

The image and video transmission must be scalable. That is to say, it should every moment adapt to the traffic needs on the network, reducing its volume when the control information is majority and enlarging it when it isn’t. In terms of video, this means that it must reduce the quality when the traffic in the network is already important, and it must enlarge it when possible. When the transmission doesn’t reach minimum quality values it should be rejected, although that will be function of the video compression and codification system, which will be the responsible of setting that minimum quality values. With regard to the remainder data types, the system that generates them is the one that will set the quality limits.

4 Design and Implementation

To implement the system two applications are needed, a sender application located on the video server equipment and a receiver application on the intermediate equipment (Fig. 1). Modular architecture is the same one in both applications (Fig. 2).
These applications are the ones that bear all the functionalities offered by the system. The previous modular architecture will be understood like a not physical but logical architecture.

4.1 Transmission module

The real time transmission module is responsible of letting the data arrive from an extreme of the Industrial Ethernet network to the other, maintaining the real time features. To get it, the module uses the RTP protocol (Real Time Protocol) [6]. The transmission module interacts with the control module in order to adapt the transmission speed every moment.

4.2 Control module

With this transmission control module is possible to controlling if the transmission is really a real time transmission. It controls that the packages are transmitted periodically, adapts the speed to the needs of the network, has flow control and congestion capacities, etc. It also takes charge of the transmission scalability, that is to say, that the video traffic is always adapted to the network traffic and it provides a way to transport the control information by the operator from the remote equipment. It uses the RTCP protocol (Real Time Control Protocol) to provide system with all the functionalities previously described.

4.3 RTP and RTCP Protocols [6] [7] [8]

RTP (Real Time Transport Protocol) is a protocol that offers extreme to extreme transport services for data with real time features. Examples of this kind of data are video and interactive audio. These services include identification of the type of load, packages sequence numeration, packages temporization and delivery monitorization. The applications generally use RTP over UDP to make use of its services of multiplexation and checksum. Nevertheless, RTP can also work over any another network or of transport layer. It’s important to know that RTP is capable of bearing multicast, but only if this can be borne by the lower layers.

The RTCP packages use the control channel. The RTCP control protocol is based on the periodic transmission of control packages to all the participants on a session, using the same mechanism of distribution as for the data packages. Its main function is to provide refreshing on the quality of the data distribution (which would correspond with the flow and congestion function of other transport protocols) by delivering sender and receiver report packages (SR and RR). Besides this type of packages, SDES, BYE and APP packages are also included. SDES packages take charge of identifying the participants on the session, BYE packages are sent to finalize the participation and APP packages allow to adding new functionalities to the system.
In running, the transmission speed and the size of the frames should be always adapted to the network available bandwidth. To perform it, the receiver sends control messages to the sender periodically, indicating the available bandwidth. The sender decides, processing that bandwidth information, the transmission rate and the size of the packages. These values are indicated to the video compression and codification system by the control interface.

4.4 Interfaces
To share information between the modules of one application a mechanism based on common “buffers” is performed. These buffers allow to joining real time transmission information. All the systems that will use the buffers have access to it, although it needs synchronism methods to avoid the simultaneous access to spaces of memory which can cause problems [10].

5 Troubles and Solutions
When the TCP/IP protocols were designed, they were thought to use for a reliable data transmission with minimum temporary restrictions (or even with no restriction). Nevertheless, the multimedia traffic requires to complying some features, so it needs the implementation of services that assure the fulfilment of them.

5.1 Packages loss
As it has been commented previously, RTP protocol is used for video transmission. Nevertheless, this protocol requires an underneath transport protocol. In this case it is used UDP, which is a non connection oriented protocol. Due to the nature of that protocol, which doesn’t guarantee reliability, and due to possible congestions situations in the network, it is possible that losses in the sent packages can appear. These losses are detected in the destiny (receiver) thanks to the sequence number field in the data packages (Fig. 3).

Actually, it is impossible to correct those losses, since the own nature of the used protocol impedes it. Nevertheless, the video transmission allows the assumption of an acceptable losses percentage. That percentage must be usually lower than 10% of the total packages. These losses will only mean a degradation of the quality. Therefore, although the losses are not rectified, they are controlled so that the video quality always overcomes an acceptable minimum threshold.

5.2 Synchronism Loss
Due to the use of non connection oriented transport protocol again, transmission synchronism losses can appear. The packages can arrive to the receiver with an order different to the one on the sender, since UDP does not preserve the sequence of information that the application provides.

In this case, the detection of the synchrony loss is also carried out on the receiver using the sequence number. Also, using the timestamp field besides the sequence number is possible to correct the jitter.

To correct this trouble is necessary an increase in the latency time. The latency time shouldn’t exceed a value adopted in the definition of the objectives of the transmission.

5.3 Rate Dynamic Control
The main objective of the system is to obtain a scalable transmission and, therefore, to adapting it dynamically to the network traffic every time. To get it, the video compression and codification system must be capable to vary its rate attending to the needs required by the transmission system. The own transmission system should also vary its transmission rate. When the traffic through the network is congested, the rate will decrease, but when the network is discharged, the rate will increase.

To carry out that control, the receiver detects every time the losses produced during an interval of transmission to according of the transmission rate and the number of participants in the session. With this information it generates the RR control packages which are sent periodically, filling the different fields correctly in the package.

The sender commits to analyzing the received RR packages and to obtaining the fraction lost of data packages during the last transmission interval. In this way, it calculates the available bandwidth and adapts its transmission speed and the video generation speed of the codification and compression system.

To carry out the network congestion analysis, a filter is used. This filter allows smoothing out the value of the lost packages and in that way avoids abrupt oscillations of the QoS. The filter used is the following one:

\[
Fract'lost = (1 – Filter) Fract'lost_{ant} + Filtro Fractlost
\]
Where \( \text{Fract'lost} \) is the resultant fraction of losses smoothed out when applying the filter to the fraction of losses \( \text{Fractlost} \). The value of the filter will vary between 0 and 1. The values close to 1 imply a greater influence of the fraction of losses in the last interval. However, the values close to 0 imply a greater influence of the losses accumulated during the previous intervals.

With this fraction of losses smoothed out, the sender calculates the network congestion and takes charge of the adaptation of the transmission dynamically. To do it, the value obtained is verified and, if it exceeds an upper threshold (10% of the total video traffic), it’s understood that the network is congested, so it’s requested a quality decrease. But if the value does not reach a minimum threshold (5%) a quality increase is requested, since the network is discharged.

These quality variations would usually imply variations both in the size of the frames (therefore modifications in the video codification and compression method) and in the sequenciation rate of the images collected by the camera. In the prototype system only sequenciation rate and transmission rate modifications have been defined. But this is only provisional.

The increments are carried out in additive way, increasing one package per second when an increment is needed. Nevertheless, the decrease will be able to carry out using additives or multiplying methods. The additives methods consist in a variation of one package per second when decrementing, just like when incrementing. Nevertheless, using a multiplying method, the variation varies in function of the rate just then. That variation will be 1/6 of the rate in that moment; so, if there is a 30 packages per second rate the decrement will be 5 packages per second, while if the rate is 20 packages per second, the decrease will be only 4 packages by second (it’s used round up).

6 Tests and Results
To verify performance of the transmission system, some transmission tests have been carried out in a 1Mbps simulated network. The tests have consisted on the transmission of fixed size video packages through a network loaded with identical traffic with priority. For the traffic with priority transmission it has been used sender applications similar to those developed for this system but avoiding the transmission control.

In the example (Fig. 5) an additive quality increment method has been used, increasing the transmission rate in one package per second each time when that has been needed. The decrement used has been an additive method, decreasing one package per second when it has been requested. In each one of these tests the value of the filter has been changed (with values of 0.1, 0.3, 0.5 and 0.7).

![Figure 5: Test results](image-url)
In the graphics (Fig. 5) show the results obtained for these tests. The continuous line represents the transmission and sequenzation rate in Kbps, while the discontinuous line represents the percentage of losses during the last interval of transmission (the total losses, before passing through the previously defined filter). The horizontal axis represents the control packages sent (considering the delivery of a package in each transmission interval).

It is observed that the best behaviour is given for a value of 0.3 in the filter. This same value has been found as the best for other tests too. With this value the system reaches the maximum transmission rate very fast, is very stable in spite of the service quality fluctuations and causes fewer losses than with the other values.

7 Conclusions
With the suggested system a suitable use of the capacities offered by an industrial network is achieved. It offers a value added service to communications system available at any over-measured network, like happen in Industrial Ethernet networks.

The transmission system proposed is used to transmit video only, since it is considered very useful to give the possibility to access to live images of an industrial process from a remote device without needing to approach that device. Nevertheless, other kind of information could be used.

It should keep in mind that the system configuration will change significantly in function of the network type, the number of machines present in the network and the type of them. Therefore, the system capacity will be adapted to the network capacity that should be analyzed previously to the installation.

For the architecture defined to the execution of the previously cited tests, it has been verified that the filter with more steady results, and therefore, with a better dynamic answer to the variations in the system is the filter with value 0.3. With this value, the system is capable of reaching the maximum transmission rate in a short time (in comparison with the remainder filters) and its behaviour when abrupt quality variations appear more softly.

About the methods used to carry out the variations in the transmission quality, the results obtained in the tests show that the use of a multiplying decrease method is more useful than the use of an additive method. With the first one the instability times are shortened, and therefore, the number of lost packages lost is reduced.

8 Acknowledgment
This work has been supported in one part by MCYT&FEDER under project DPI 2002-2399. And in other part by ETORTEK 2005, AmI GUNE.

References: