

Digital Video Transmission and the FDDI Token Ring Protocol

B. Cousin
ENSERB - LaBRI*
351 Cours de la Libération
F-33405 Talence cedex
FRANCE

fax : 33 - 56.37.20.23
tel : 33 - 56.84.65.29
e-mail : cb@tirukali.greco-prog.fr

Abstract

This paper proves that the temporal constraints for digital video transmission can be met by the synchronous transmission mode of the FDDI token ring protocol. First, we introduce the FDDI protocol. Second, we establish the temporal constraints for real time image transmission. Then we propose a technique based on an optimal use of emitter and receiver image buffers, and an optimal constituting of transmission units. Finally, we verify that the temporal constraints are fulfilled by the FDDI protocol and the proposed technique. This technique requires only an optimal allocation of the exact bandwidth needed by the image transmission. Moreover, image sample blocking enables larger TTRT to be used, and thus, reduces the overhead induced by the token rotation. We also prove that the proposed technique produces a minimal and constant delay equal to $2 * TTRT$ plus the physical response time of the network, in spite of the aperiodic delivery of the image samples due to the access method of the FDDI protocol.

1. Introduction

Our research group is interested in digital video transmission. Given the very large bandwidth required by digital video, only high speed networks can be considered to achieve the image transmission throughput. We have chosen to focus our study on the FDDI token ring protocol because it is the earliest representative of standardized high speed local area networks, and because its component sets are widely available now, enabling low cost and easy experimentation [FDDI 87, FDDI 88, FDDI 89, AMD 89].

FDDI uses fundamentally a **technique** of asynchronous transmission (that is to say, the delay in transmission is variable), but this protocol stipulates two **modes** of transmission : the asynchronous mode and the synchronous mode. The synchronous mode guarantees a station a pre-allocated bandwidth and the right to transmit with an average periodicity equal to a value negotiated among all the stations. This periodicity is referred to as the Target Token Rotation Time (TTRT). Furthermore, the protocol guarantees a maximum rotation time of the token that cannot surpass $2 * TTRT$. The synchronous mode is used for those applications whose bandwidth and response time limits are predictable, permitting them to be preallocated. The asynchronous mode is used for those applications whose bandwidth requirements are less predictable (bursty or potentially unlimited) or whose response time requirements are less critical. Asynchronous bandwidth is instantaneously allocated from the pool of remaining ring bandwidths that are unallocated or unused.

* Unité de recherche n° 1304 associée au CNRS

At first glance, it seems easy to transmit digital voice or video by means of the synchronous mode of the FDDI protocol. But let us recall that, the digital image traffic is relatively unpredictable, first because a compression process can be required to enable high definition images to be transmitted on a 100 Mbit/s network, and second, in order to have a low frequency of access to the medium to reduce the overhead induced by the token rotation, it is necessary to block together and send simultaneously several images or several parts of different images. Moreover FDDI protocol does not propose isochronous transmission : the token ring access method can lead to variations in the transmission delay of images due to variation of the network load.

An upward-compatible enhancement, called FDDI-II [FDDI 9b, Ross 90, Teener 90], adds a circuit switched (isochronous) service to the existing packet service of the basic FDDI. FDDI-II employs a cycle structure to control the multiplexing of packet data and circuit switched data in the same ring. This structure repeats on the ring every 125 microseconds. The allocation of isochronous channels is allowed with a variable granularity. Supported channels sizes include 8, 16, 32 and 64 Kbit/s plus any multiple of 64 Kbit/s up to 6.144 Mbit/s, providing a maximum of 98.304 Mbit/s isochronous payload. In addition, the aggregate of any or all of the isochronous channels may be used as one circuit service, satisfying the needs of heavy load applications. The remaining and unallocated bandwidth is dedicated to the packet service.

We have chosen to use the synchronous mode of transmission of the FDDI protocol and not the isochronous mode of FDDI-II. In first place because FDDI-II network is not yet available, and in second place, because we do not suppose that the sample traffic occurs in precise amounts on a time basis. This is due to the compression process which is required to enable high definition image to be transmitted on a 100 Mbit/s network. So our sample traffic is relatively unpredictable, and the use of circuit switched service would need overallocation of isochronous channels.

At the emitter, to enable the synchronous allocation of the right mean bandwidth required by the video transmission to be fulfilled, our intention is to carefully undertake of the constituting of each synchronous transmission unit. If the token is late, the transmission unit carries all the image samples produced from the instant of the previous token arriving at the instant of the expected token arriving (TTRT after the arriving of the previous token). If the token is early, the transmission unit carries all the remaining samples, because the control of the quantity of data transmitted at each capture of the token does not have to be managed at the level of the FDDI sender, the normal throughput of the emitter of images naturally assuming this role. The buffering induced by this technique can be usefully exploited to block several samples in one unit of transmission enabling the transmission to be improved, first because the overhead induced by the structure of the transmission unit (starting and ending delimiters, addresses, frame control, and so on) is spread over a large number of samples, second because the token rotation overhead is reduced with its frequency. This way of doing the constituting, enables the emitting delay to be minimized and needs only an optimal allocation of the exact bandwidth required by the image transmission.

At the receiver, in order to allow the transmission of periodic information with asynchronous techniques, it is necessary to supply sufficient buffer space in the memory at the level of the receiver to accommodate the inevitable variations in transmission delay. These memory buffers imply a systematic delay inhospitable to interactive applications, but these memory buffers can be usefully exploited to detect and then to correct loss, corruption, and duplication in all or part of the images of the video, and to synchronize the flow of data to correct the inherent drift problem due to the presence of distinct clocks in the sender and in the receiver.

In short, the problems of data corruption, of synchronization, and of adaptation of the receiver to the flow of data, all oblige both the sending and the receiving stations to buffer part of the images. Buffering at the level of the sender allows us to send large frames onto the network

and thus enables the transmission to be optimized. The sender buffer is also needed to adapt the flow of the image samples to the token delay. Buffering at the level of the receiver allows us to manage the aperiodic delivery of the image samples due to the previous phenomena. In fact, the accumulation of these different bufferings is quite acceptable physiologically; a delay of several images occasions a delay of less than a tenth of a second, a negligible delay in human terms. We recall that this delay is applicable to all the images, and thus only the starting of the video is affected by it. Nevertheless, this buffering does demand a great deal of memory.

2. FDDI

In a former work, we have developed a model of the FDDI protocol [Cousin 91]. Accordingly to the standard statements, we have formally proved that the time between successive token arrivals at any given station has an upper limit of twice the Target Token Rotation Time (TTRT) negotiated by the stations during the initialization phase of the protocol minus the token delay measured at the given station during the previous rotation. This limit is tantamount to proving that the token rotates quickly enough to satisfy the standard statement : "The protocol guarantees an average response time (TRT : Token Rotation Timer) not greater than the TTRT, and a maximum TRT not greater than twice TTRT". In [Johnson 87] Johnson proves a similar but weaker result to assure that the token rotates quickly enough to prevent initiation of recovery unless there is failure of a physical resource or unless the network management entity within a station initiates the recovery process. A formal proof of the two properties can be found in [Sevcik 87], but the study is applied on a lightly modified and improved FDDI protocol.

We will let N denote the number of stations in the network, assuming one network connection per station. The stations around the ring will be numbered from 0 to $N-1$ in a clockwise manner.

The protocol FDDI uses an access method called Token Ring [Johnson 86, Ross 89]. FDDI stations are connected in a ring where a token passes round. Any station that wants to transmit over the ring has to capture the token. At the end of its transmission, that station must release the token. Every station, therefore, obtains the right to use the medium turn by turn.

We denote $T_j(I)[k]$ the moment where the token is received at the station k during its I^{th} rotation. So, $T_j(I+1)[k]-T_j(I)[k]$ is the duration of the I^{th} rotation of the token, duration measured at the station k .

During the initialization phase of the FDDI protocol, all the stations connected to the ring negotiate the value of TTRT. The TTRT chosen is the smallest. TTRT has to lie between the two values, $TTRT_{\min}$ and $TTRT_{\max}$ ($TTRT_{\max} \geq TTRT \geq TTRT_{\min}$). The value $TTRT_{\min}$ corresponds to the minimum time for the management and rotation of the token. A TTRT value less than $TTRT_{\min}$ would not even allow the token to reach all of the stations, and thus such a time is unacceptable. A TTRT value greater than $TTRT_{\max}$ is conceivable without major problems except that it creates a partition of the medium that is somewhat prejudicial to equal access because a station holding the token could very well keep it a very long time. Furthermore, a TTRT value greater than $TTRT_{\max}$ slows the detection of errors and in addition the reconfiguration of the ring.

Let us note that the smaller the value of TTRT, then the more important becomes the amount of time dedicated to the management of access to the medium. In fact, the number of rotations of the token per unit of time is inversely proportional to TTRT. Thus the token consumes a great part of the bandwidth. We have every interest in sustaining as great as possible a TTRT within the limits of foreseeable applications [Dykeman 88].

To provide the required service towards the token rotation duration the standard specifies a timer in each station called TRT (Token Rotation Timer). It is used to control scheduling during normal operation and to detect and recover from serious error situations. Whenever the TRT expires, it is reinitialized to the TTRT value and the variable "Late_Ct" is set. If the token arrives at the station before the expiration of the timer (early), it is reinitialized to the value TTRT and the variable "Late_Ct" is reset.

Although in the standard the FDDI timing control is assured by both the value of the TRT and the Late_Ct variable, in order to simplify the timing model we use only one parameter. We denote by $TRT(I)[k]$ the value of the timer at the station k during the I^{th} token rotation. It can be recursively defined by the token rotation time T_j :

(TRT definition) :

$$\forall k, TRT(0)[k] = T_j(1)[k] - T_j(0)[k], \text{ and } \forall I \neq 0, \\ \text{if } TRT(I-1)[k] \leq TTRT \text{ then } TRT(I)[k] = T_j(I+1)[k] - T_j(I)[k], \\ \text{else } TRT(I)[k] = (TRT(I-1)[k] - TTRT) + T_j(I+1)[k] - T_j(I)[k].$$

At the first rotation the TRT value is initialized. Afterwards if the token is early, that is to say if the previous rotation respects the negotiated periodicity TTRT, the next TRT is equal to the token rotation duration. Otherwise, if the token is late, the next TRT is equal to the sum of the token rotation duration and the token delay. This sum enables the delay to be taken into account from one rotation to the next, and thus, enforces the periodicity.

Each station has another timer called by the standard : THT (Token Holding Timer). It contains the maximum duration during which the stations can transmit in asynchronous mode. It is set at each early token reception with the token gain. We denote by $THT(I)[k]$ the value of the timer at the station k during the I^{th} token rotation :

(THT definition) :

$$\forall I, \text{ if } TRT(I)[k] \leq TTRT \text{ then } THT(I)[k] = TTRT - TRT(I)[k] \text{ else } THT(I)[k] = 0.$$

Similarly, $\delta_j(I)[k]$ denotes the delay of a late token :

(δ_j definition) :

$$\forall I, \text{ if } TRT(I)[k] > TTRT \text{ then } \delta_j(I)[k] = TRT(I)[k] - TTRT \text{ else } \delta_j(I)[k] = 0.$$

We denote $T_s(I)[k]$ the synchronous transmission duration during the I^{th} token rotation. Using the previous notations, the maximum token rotation duration can be described by :

(Token rotation duration property) :

$$\forall I, \forall k, TRT(I)[k] \leq 2 * TTRT.$$

So now, we have to show that the synchronous transmission mode of the FDDI protocol enables real time image transmission to be carried out.

3. Images

3.1 Temporal constraints

Both transmission of digital images and digital voice have temporal constraints that we do not ordinarily encounter in conventional data transfer. These temporal constraints associate samples. A sample is that portion of a signal that is digitalized. For example, a sample could be a group of bits, one byte of coded sound, or a line of an image.

The set of the samples makes a sequence $\{e_i\}$. So we can associate to each sample its running number in the sequence. We indicate the moment of production of the sample by the emitter with the notation T_e . Likewise, we use the notation T_v (visualization) to indicate the moment

when the sample can be displayed on the visual equipment. T_e and T_v are strictly increasing functions.

The preservation of the quality of the video during transmission requires that two constraints must be satisfied. First constraint : the delay after the emission of the video must be humanly tolerable, virtually instantaneous. We refer to the time that one must wait to see the first image of a video as T_{max} . This time is critical if the user intervenes in the unfolding of the video; that is, if the video is in any sense interactive. Second constraint: the images should appear on the screen of the receiver at the same speed relative to one another as they are produced by the emitter. If these two constraints are satisfied, then the video is received with temporal integrity. Two relations suffice to express these constraints:

$$\text{Tolerable delay constraint} \quad : \forall i, T_v(i) < T_{max} + T_e(i).$$

$$\text{Temporal integrity constraint} \quad : \forall i, \forall j, T_v(j) - T_v(i) = T_e(j) - T_e(i).$$

These temporal constraints exist only if the video should be visualized on its arrival at the receiver (in real time, no less!). These constraints do not exist if the video is broadcast in deferred time (for example, if it is pre-recorded for later broadcast), and if it is thus consequently stored on its arrival at the receiver. In such a case, the transmission of the video can simply be treated as the transmission of a large file.

3.2 Transmission

In fact, in as much as they are located on distinct sites, the receiver of images is completely independent of the sender of images, and it is thus difficult to respect the two previous constraints. Several phenomena intervene in the variation of transmission delay : the access method, the image compression and decompression processes, the blocking of samples, the clock drift, etc.

Conventionally, the clock of the receiver of images is slaved to the clock of the emitter by means of a synchronization included in the signal. Since the conventional methods of transmission use **isochronous** (circuit switched) technique, the intervals of time between samples are preserved during their transmission. The synchronization of the receiver with the sender of the images is therefore easily achieved. It suffices to slave the receiver's clock to the flow of the received images. Only a constant delay is added (the propagation delay).

The most current techniques of transmission now use **asynchronous** transmission technique. With this technique, the delay in the transmission of the samples varies: it depends on the access method to the medium, on the resolution of collisions, on the load of the network, etc. Consequently the time separating two samples at their reception may differ from the time separating them at their production. We can no longer count on slaving the clock of the receiver directly on the flow of received images.

The technique of asynchronous transmission permits a better use of support than does isochronous transmission because sporadic flows can be compensated for. Asynchronous transmission works well with dynamic allocation methods of the bandwidth between different links as a function of the load. Nevertheless, in order to be efficient, the overhead introduced by this dynamic management must be compensated for by a better allocation of the traffic. In contrast, isochronous techniques can use a method of static allocation that requires little or no management overhead.

We denote T_r the moment of reception of a sample. We denote by T_t the response delay of a sample over the network. These moments are described by the relation :
(*Tr definition*) :

$$\forall i, T_r(i) = T_e(i) + T_t(i).$$

3.3 Blocking

The blocking of several samples in one unit of transmission enables the transmission to be improved. The overhead induced by the structure of the transmission unit (starting and ending delimiters, addresses, frame control, and so on) is spread over a large number of samples. As all the samples blocked in the same transmission unit are sent and received at the same moment, this technique produces variation in the transmission delay. Consequently the time separating two samples at their delivery may differ from the time separating them at their production.

The response delay T_t of local area networks like FDDI consists of the access delay T_a , the transmission delay T_d , and the propagation delay T_p . The propagation delay depends on the propagation speed and the length of the media. The propagation delay can be regarded as constant. The transmission delay depends on the data rate and the length of the transmission unit. The access delay depends on both the load and the access method used by the protocol. Access delays fluctuate in most LAN. They are related by the relation :

(*T_t definition*) :

$$\forall i, T_t(i) = T_a(i) + T_d(i) + T_p.$$

As the samples, the transmission units makes a strictly increasing sequence $\{S_i\}$. So we can associate to each unit is number in the sequence. We denote $I(i)$ the number of the transmission unit associated to the sample of number i . We denote $\text{deb}(I)$ the number of the first sample of the unit I . And we denote $\text{fin}(I)$ the number of the last sample of the unit I .

If two consecutive samples do not belong to the same transmission unit then they belong to two distinct but consecutive units.

(*Consecutive units definition*) :

$$\forall i, \text{if } I(i+1) \neq I(i) \text{ then } I(i+1) = I(i) + 1.$$

Previously, we noted that all the samples blocked in the same transmission unit are sent and received at the same moment, because the samples blocked in a same transmission unit are available at the receiver when the transmission unit is entirely received.

(*Unit receiving moment definition*) :

$$\forall i, \forall j, \text{if } I(i) = I(j) \text{ then } \text{Tr}(i) = \text{Tr}(j).$$

3.4 The usable synchronous bandwidth

Let S (and R respectively) the station number of the sender (the receiver) of samples. If we use the FDDI protocol to send the images, a station begins to send when it receives the token. So $T_j(I)[S]$ is also the moment where the transmission units associated with the I^{th} token rotation are sent. As FDDI protocol uses token ring as the access method, the moments T_e , T_j , and T_a are related by :

(*Access method definition*)

$$\forall i, T_j(I(i))[S] = T_e(i) + T_a(i).$$

To use the synchronous transmission mode of the FDDI protocol to send a video, first, we need to know the average synchronous throughput $D_s[k]$ required by each station k to transmit the images in real time, to ensure that we always maintain the following relation :

(*Synchronous bandwidth definition*) :

$$\sum_{k \in [n, n+N]} D_s[k] \leq D.$$

That is to say, the sum of throughput sent should be lower than the effective throughput D of the network. This avoids overallocation of the medium. The effective throughput is obtained

by starting from the nominal throughput minus the throughput used to manage the network, essentially the packaging of the frames and the management of the token. Network management has the responsibility for maintaining this statement. Every station requesting to transmit in synchronous mode calls the network management for a reservation of the average throughput required ([FDDI 88] and [FDDI 87]).

Secondly, once the TTRT is fixed, once we know the average synchronous throughput $Ds[k]$ required by each station k to transmit the images in real time, to achieve correct protocol operation we have to maintain the previously established relation :

(*Synchronous transmission duration definition*) :

$$\forall I, \forall n, \sum_{k \in [n, n+N[} Ts(I)[k] \leq TTRT.$$

To maintain this relation for any rotation, the evident solution is to limit the duration of the synchronous transmission to $Ts[k]$.

(*Maximum synchronous transmission definition*) :

$$\forall k, Ts[k] = Ds[k].TTRT + D.$$

Unfortunately, the load on the network can make the moment at which the token arrives at a station vary greatly, remembering that a station must capture the token before it can transmit. This moment is remembered by the token rotation timer (TRT) local to each station. It may be early or late with respect to the negotiated TTRT period. Logically, in order to maintain the inequality of the synchronous bandwidth relation, each station k should have the right to transmit at most $Ds[k].TRT$ bits. This quantity is extremely difficult to manage because the TRT varies as a function of the load with each rotation of the token. ($0 < TRT < 2 * TTRT$). Moreover, the implementation of FDDI does not permit us to get the value of TRT in time. We risk, then, to exceed the duration $Ts(I)[k]$ attributed to each station k , and thus to violate the inequality of the synchronous transmission duration relation, if we do not adapt the length of a frame to the rotation time of the token.

However, if the token is early ($TRT < TTRT$), this indicates that the network is underloaded, and thus it is permissible to transmit $Ds[k].TTRT$ bits, but impossible. It is impossible because the image emitter has not yet produced enough samples in such a short time. Our proposition is to send all the produced samples when an early token arrives. In that case, we know that the synchronous bandwidth relation and the synchronous transmission duration relation are obviously enforced by the regular throughput of the image emitter.

Inversely, if the token is late ($TTRT < TRT$), then the FDDI operations ensure that the delay cannot surpass $2 * TTRT$, even if all of the stations transmit the entirety of their throughput synchronously. But if we want to respect the synchronous transmission duration relation, the stations are allowed to send at most $Ds[k].TTRT$ bits at each token rotation. The remaining $(TRT - TTRT).Ds[k]$ bits are not sent with the first arrived token, but the FDDI operation guarantees that the delay will not be cumulative, so the remaining bits will be sent with the next units. In fact, the protocol is self-regulating because the overload induced by the token diminishes if the time between two passes of the token grows. Furthermore, if one of the stations does not use the entire throughput allocated to a synchronous transmission, then the unused time will be recovered, first of all, to ensure other synchronous transmissions, in recovering the delay, and in re-establishing the negotiated frequency of rotation of the token; secondly and ultimately to authorize asynchronous transmissions.

So, we propose building each synchronous transmission unit in such a way (Figure 1). If we denote $J(I)$ the set of sample numbers carried by the transmission unit associated with the I^{th} token rotation, then the relation of correct constituting of the transmission unit property is defined by :

(Correct constituting of transmission unit definition) :

$$\forall I, \forall i \in J(I), T_j(I)[S] - TRT(I-1)[S] \leq T_e(i) < T_j(I)[S] - \delta_j(I-1)[S].$$

In conclusion, we propose to use the synchronous mode of the FDDI protocol to transmit images of a video. The mean transmission rate $D_s[k]$ necessary for the transmission of the video should be known, and the application requires the network manager to make an appropriate reservation for the duration of the video to guard against congestion of the media. The negotiation procedure for the TTRT could then be started, if required. The smaller the value of the required TTRT, the smaller the delay in transmission. However, we have already raised the idea that the efficiency of the FDDI protocol will be accordingly weakened. The calculations that we have undertaken indicate that the ideal value lies in the neighborhood of twenty milliseconds [Cousin 90]. Independently of the fact that the negotiated value of the TTRT should lie between TTRT_{min} and TTRT_{max} to ensure the proper global functioning of the network, our application can accommodate a large range of values for TTRT. If the token is late, the application should be able to transmit at most $L_s[k] = D_s[k].TTRT$ bits at each rotation of the token. If the token is early, the control of the quantity of data transmitted at each capture of the token does not have to be managed at the level of the FDDI sender, the normal throughput of the emitter of images naturally assuming this role.

3.5 Validation

First of all, we have to prove that the definition of the correct constituting of transmission units enables the maximum synchronous transmission duration requirement, and thus the synchronous bandwidth definition, to be fulfilled.

(Maximum synchronous transmission duration property) :

$$\forall k, \forall I, T_s(I)[k] < T_s[k].$$

Proof :

Assuming that $L_s(I)[k]$ is the number of bits of the transmission unit associated with the I^{th} token rotation, by definition of $T_s(I)[k]$:

$$T_s(I)[k] = L_s(I)[k] + D.$$

The number of bits of a transmission unit has an upper limit of the duration between the first and the last sample of the transmission unit multiplied by the effective throughput of the sample.

$$\Rightarrow T_s(I)[k] \leq (T_e(\text{fin}(I)) - T_e(\text{deb}(I))) \cdot D_s[k] + D.$$

Assuming the correct unit constituting definition :

$$\Rightarrow T_s(I)[k] \leq ((T_j(I)[S] - \delta_j(I-1)[S]) - (T_j(I)[S] - TRT(I-1)[S])) \cdot D_s[k] + D.$$

$$\Rightarrow T_s(I)[k] \leq (TRT(I-1)[S] - \delta_j(I-1)[S]) \cdot D_s[k] + D.$$

Two cases appear :

1. Either $TRT(I-1)[S] < TTRT$ then $\delta_j(I-1)[S] = 0$:

$$\Rightarrow T_s(I)[k] \leq TRT(I-1)[S] \cdot D_s[k] + D.$$

Which can have an upper limit, according to the assumption :

$$\Rightarrow T_s(I)[k] \leq TTRT \cdot D_s[k] + D.$$

According to the definition of maximum synchronous duration :

$$\Rightarrow T_s(I)[k] \leq T_s[k]. (\diamond)$$

2. Or $TRT(I-1)[S] > TTRT$ then by definition $\delta_j(I-1)[S] = TRT(I-1)[S] - TTRT$:

$$\Rightarrow T_s(I)[k] \leq TTRT \cdot D_s[k] + D.$$

By definition of $T_s[k]$:

$$\Rightarrow T_s(I)[k] \leq T_s[k]. (\diamond)$$

The temporal integrity constraint can be achieved, first, if the samples are buffered between the receiver and the image visualization equipment. The buffer has to be large enough to contain all the samples produced during $2 * TTRT$ duration. Secondly, we can prove that all the samples are received in time at the receiver (i.e. before being displayed).

(Correct timing visualization property) :

$$\forall i, Tv(i) > Tr(i).$$

To prove this property, we need to prove the property of correct reception. If the visualization moment of the first sample is delayed by twice the Target Token Rotation Time then the reception moment of the samples is limited by the visualization moment of the first and the last sample of the same transmission unit.

(Correct reception property) :

$$\forall i, Tv(0) = Tr(0) + 2 * TTRT \Rightarrow Tv(\text{fin}(I(i))) < Tr(i) + 2 * TTRT - \delta_j(I(i)-1)[S] \leq Tv(\text{deb}(I(i)+1))$$

Proof :

According to the correct constituting relation :

$$\forall I, \forall i \in J(I), T_j(I)[S] - TRT(I-1)[S] \leq Te(i) < T_j(I)[S] - \delta_j(I-1)[S].$$

$$\text{For } i = \text{fin}(I) : (1) \forall I, T_j(I)[S] - TRT(I-1)[S] \leq Te(\text{fin}(I)) < T_j(I)[S] - \delta_j(I-1)[S],$$

$$\text{and for } i = \text{deb}(I) : (2) \forall I', T_j(I')[S] - TRT(I'-1)[S] \leq Te(\text{deb}(I')) < T_j(I')[S] - \delta_j(I'-1)[S].$$

$$\text{According to the TRT definition : } \forall I, TRT(I)[S] = T_j(I+1)[S] - T_j(I)[S] + \delta_j(I-1)[S],$$

$$\text{which can be rewritten : } \forall I, T_j(I+1)[S] - TRT(I)[S] = T_j(I)[S] - \delta_j(I-1)[S].$$

Let $I' = I + 1$, then the relations (1) et (2) can be rewritten :

$$(3) \forall I, Te(\text{fin}(I)) < T_j(I)[S] - \delta_j(I-1)[S] \leq Te(\text{deb}(I+1)).$$

According to the Te definition :

$$\forall i, Tv(i) = Te(i) + Tv(0) - Te(0).$$

The relation (3) can be rewritten :

$$\Rightarrow \forall I, Tv(\text{fin}(I)) < T_j(I) - \delta_j(I-1)[S] + Tv(0) - Te(0) \leq Tv(\text{deb}(I+1)).$$

From the assumption about the visualization moment of the first sample :

$$Tv(0) = Tr(0) + 2 * TTRT.$$

$$\Rightarrow \forall I, Tv(\text{fin}(I)) < T_j(I) - \delta_j(I-1)[S] + Tr(0) + 2 * TTRT - Te(0) \leq Tv(\text{deb}(I+1)).$$

According to the Tr definition, the Tt definition and the access method definition :

$$\forall i, Tr(i) = T_j(I(i))[S] + Td(i) + Tp.$$

$$\Rightarrow \forall i, Tv(\text{fin}(I(i))) < T_j(I(i))[S] - \delta_j(I(i)-1)[S] + T_j(I(0))[S] + Td(0) + Tp + 2 * TTRT - Te(0) \leq Tv(\text{deb}(I(i)+1)).$$

Assuming that the transmission delay is constant for a fixed data rate : $\forall i, Td(i) = Td$.

$$\Rightarrow \forall i, Tv(\text{fin}(I(i))) < Tr(I(i)) - \delta_j(I(i)-1)[S] + T_j(I(0))[S] + 2 * TTRT - Te(0) \leq Tv(\text{deb}(I(i)+1))$$

From the assumption of the sending moment of the first sample : $Te(0) = T_j(I(0))[S]$,

$$\Rightarrow \forall i, Tv(\text{fin}(I(i))) < Tr(I(i)) - \delta_j(I(i)-1)[S] + 2 * TTRT \leq Tv(\text{deb}(I(i)+1)). (\diamond)$$

Then, we prove the correct timing visualization property :

$$\forall i, Tr(i) < Tv(i).$$

Proof :

Recurrent demonstration :

1. For $i=0$, the relation is obvious because, according to the visualization moment of the first sample assumption ($Tv(0) = Tr(0) + 2 * TTRT$) then : $Tr(0) < Tv(0)$. (\diamond)

2. Assuming that the recurrent assumption is true for $i \in [0, n]$, two cases appear :

2.1 Either the samples n and $n+1$ belong to the same transmission unit : " $I(n) = I(n+1)$ ".

Then, according to the Tv definition :

$$Tv(n+1) = Tv(n) + Tv(n+1) - Tv(n).$$

According to the second temporal constraint :

$$Tv(n+1) = Tv(n) + Te(n+1) - Te(n).$$

According to the strictly increasing function Te :

$$Tv(n+1) > Tv(n).$$

According to the recurrent assumption :

$$Tv(n+1) > Tr(n).$$

According to the initial assumption : if $I(n)=I(n+1)$ then $Tr(n)=Tr(n+1)$.

$$Tv(n+1) > Tr(n+1). (\diamond)$$

2.2 Either the samples n and $n+1$ do not belong to the same transmission unit : " $I(n) \neq I(n+1)$ ".

Then, we know that " $I(n+1)=I(n)+1$ ", because the samples and the transmission units are numbered in an strict increasing manner.

According to the Tr definition, the Tt definition and the access method definition :

$$Tr(n+1) = Tj(I(n+1))[S] + Td(n+1) + Tp. \text{ Which can be rewritten :}$$

$$Tr(n+1) = Tj(I(n))[S] + Tj(I(n+1))[S] - Tj(I(n))[S] + Td(n+1) + Tp.$$

According to the following assumption : $\forall i, Td(i)=Td$.

$$Tr(n+1) = Tr(n) + Tj(I(n+1))[S] - Tj(I(n))[S].$$

From the token rotation duration property, the TRT definition and the δ_j definition :

$$Tr(n+1) \leq Tr(I(n)) + 2 * TTRT - \delta_j(I(n)-1)[S].$$

According to the correct reception property :

$$\forall i, Tv(\text{fin}(I(i))) < Tr(I(i)) + 2 * TTRT - \delta_j(I(i)-1)[S] \leq Tv(\text{deb}(I(i)+1)). \text{ Then :}$$

$$Tr(n+1) < Tv(\text{deb}(I(n)+1)).$$

According to the definition of the deb and fin functions : if $I(n) \neq I(n+1)$ then $\text{fin}(I(n))=n$ and $\text{deb}(I(n)+1) = n+1$.

$$\text{We deduce that : } Tr(n+1) < Tv(n+1). (\diamond)$$

The Tolerable delay constraint introduces a delay T_{\max} , which can be deduced from the previous definitions and relations :

(T_{\max} lower limit property) :

$$T_{\max} > Tt(0) + 2 * TTRT.$$

Proof :

According to the Tolerable delay constraint :

$$\forall i, T_{\max} > Tv(i) - Te(i).$$

Which can be rewritten :

$$\forall i, T_{\max} > Tv(0) + (Tv(i) - Tv(0) - Te(i) + Te(0)) - Te(0).$$

According to the temporal integrity constraint :

$$T_{\max} > Tv(0) - Te(0).$$

According to the visualization moment of the first sample :

$$T_{\max} > Tr(0) + 2 * TTRT - Te(0).$$

According to the Tt definition :

$$T_{\max} > Tt(0) + 2 * TTRT. (\diamond)$$

Accordingly, at the price of a delay due to the buffering equal to $2 * TTRT$ plus the physical response time of the network, we prove that it is useless to request a rotation time equal to half of the delay required by the application, as that would let it over-determine the maximum rotation time guaranteed to be more than $2 * TTRT$.

4. Conclusion

In view of this study, we can observe the image transmission has to resolve two problems, namely the preservation of the synchronization between images, and the minimization of the buffer length.

Yet the ring topology in the FDDI protocol both necessitates and allows controlled access to the medium, and favors the management of a method of access favorable to the transmission of images by the creation of two modes of transmission : the synchronous and the asynchronous modes. The synchronous mode of transmission guarantees a station an average throughput and the right to transmit with a periodicity, on the average, equal to a value --the TTRT -- negotiated among all the stations. Moreover, this mode guarantees that the maximum rotation time cannot exceed $2 \cdot \text{TTRT}$.

Nevertheless the service provided by the FDDI synchronous transmission mode is not sufficient to satisfy the temporal constraints required by the image transmission. To succeed in satisfying these constraints, we propose to carefully undertake the constituting of each transmission unit. If the token is late, the transmission unit carries all the image samples produced from the instant of the previous token arriving at the instant of the expected token arriving. If the token is early, the transmission unit carries all the remaining samples. To deal with the asynchronous transmission technique and the aperiodic delivery of the samples, we propose to buffer samples at the level of the sender and at the level of the receiver.

First, we prove that this proposed way of constituting and buffering enables the temporal constraints to be satisfied. Second, we prove that the overall delay of any sample is equal to $2 \cdot \text{TTRT}$ plus the physical response time of the network. We prove that it is possible to allocate only the exact mean throughput to achieved image transmission in real time. This allocation optimizes the use of the medium bandwidth. Moreover, sample blocking enables our application to be adapted to a large range of TTRT, enabling an efficient image transmission process.

We observe that in order to allow the transmission of periodic information with asynchronous techniques, it is necessary to transmit temporal information explicitly; it is further necessary to supply sufficient buffer space in memory at the level of the sender to deal with the token delay and at the level of the receiver to accommodate the inevitable variations in transmission delay. These memory buffers imply a systematic delay inhospitable to interactive applications. In short, the great throughput required by the transmission of images obliges one to use a great quantity of rapid access memory. Yet the memory buffers necessary at the level of the receiver in order to allow the use of an asynchronous technique of transmission can be usefully exploited to detect and then to correct loss, corruption, and duplication in all or part of the images of the video.

A prototype is under development which will put into practice what we prove. This prototype consists of three subsystems : the video subsystem, the communication subsystem, and the computer interface. The communication subsystem uses the AMD SuperNet components for the FDDI network. Each subsystem is linked to the others by a very high speed bus. This architecture enables both the local and distant image distribution not to be restricted by the bus throughput of the computer. Specific components enabling fast image processing (real time compression, image analysis,...) will be included in the prototype in the near future. This prototype will enable us to study the network management, principally the synchronous allocation bandwidth process which is not largely developed in the current FDDI network proposed.

References

- [AMD 89] AMD, "The SuperNet family for FDDI", Advanced Micro Devices technical manual, July 1989.
- [Cochennec 85] J.Y.Cochennec, P.Adam, T.Houdoin, "Asynchronous time-division Networks : Terminal Synchronization for Video and Sound Signals", GLOBECOM'85, 1985.

- [Cousin 90] B.Cousin, L.Kamoun, "ATM, HSLAN and Images", 3rd IEEE COMSOC International MultiMedia Workshop (Multimedia'90), Bordeaux - November 1990.
- [Cousin 91] B.Cousin, R.Castanet, "Evaluation du protocole FDDI pour la diffusion d'images en temps réel", colloque francophone sur l'ingénierie des protocoles (CFIP'91). Pau - France, 16-20 Septembre 1991.
- [Dykeman 88] D.Dykeman, W.Bux, "Analysis and Tuning of the FDDI Media Access Control Protocol", IEEE Journal on Selected Areas in Communications, vol SAC6 n°6, p997-1010, July 1988.
- [FDDI 87] "FDDI Token Ring Media Access Control (MAC)", ANSI X3.139, 1987.
- [FDDI 88] "FDDI Token Ring Physical Layer Protocol (PHY)", ANSI X3.148, 1988.
- [FDDI 89] "FDDI Token Ring Physical Layer Medium Dependent (PMD)", Draft Proposed ANSI X3.166, March 1989.
- [FDDI 9a] "FDDI Token Ring Station Management (SMT)", Draft Proposed ANSI X3T9.5, May 1990.
- [FDDI 9b] "FDDI Token Ring Hybrid Ring Control (HRC)", Draft Proposed ANSI X3.186, 11 May 1990.
- [Iyer 85] V.Iyer, S.P.Joshi "FDDI's 100 Mbit/s protocol improves on 802.5 spec's 4 Mbit/s limit", EDN Electronics Data Network, p151-160, May 1985.
- [Johnson 86] M.J.Johnson, "Reliability Mechanisms of the FDDI High Bandwidth Token Ring Protocol", Computer Networks and ISDN Systems n°11, North Holland, 1986.
- [Johnson 87] M.J.Johnson, "Proof that Timing Requirements of the FDDI Token Ring Protocol are Satisfied", IEEE transactions on communications vol COM35 n°6, June 1987.
- [Lam 78] S.S.Lam, "A New Measure for Characterizing Data Traffic", IEEE transaction on communications, vol COM 24 n°1, January 1978.
- [Ross 89] J.E.Ross, "An Overview of FDDI : The Fiber Distributed Data Interface", IEEE Journal on Selected Areas in Communications, vol SAC7 n°7, p1043-1051, September 1989.
- [Ross 90] J.E.Ross, J.R.Hamstra, R.L.Fink, "A LAN among MAN's", Computer communication review vol 20 n°3, July 1990.
- [Sevcik 87] K.Sevcik, M.Johnson, "Cycle Time Properties of the FDDI Token Ring Protocol", IEEE transactions on software engineering vol SE13 n°3, March 1987.
- [Teener 90] M.Teener, R.Gvozdanovic, "FDDI-II Operation and Architectures", 14th conference on Local Computer Networks, 1990.
- [Ulm 82] J.M.Ulm, "A Timed Token Ring Local Area Network and its Performance Characteristics", 7th conference on local computer networks, p50-56, Minneapolis-USA, February 1982.

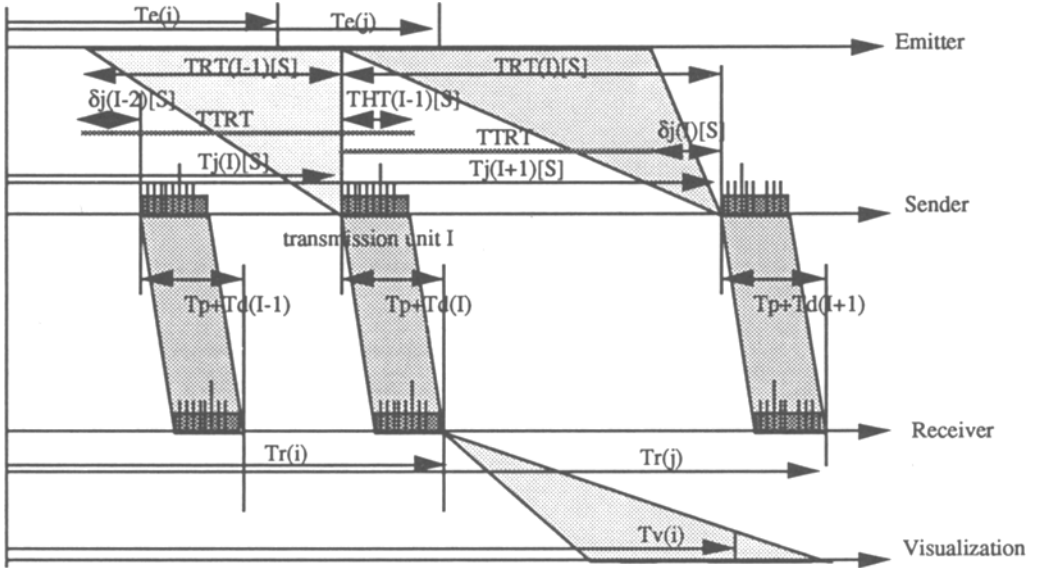


Figure 1 - Constituting of the transmission units