Abstract of paper entitled

Data Broadcasting: "DIDON" and "DIODE" Protocols

by Yves Noirel, CCETT/Rennes, France

Summary:

A definition of the broadcasting mode is given by using the concepts of *audience* and *programming*. There follows a description of the DIDON protocol and the DIODE protocol. The DIODE protocol is a specialized application of the DIDON protocol, when data is broadcast upon request.

The DIODE concept is explained in relation to the various data retrieval systems, and emphasis is placed on its role in learning about audience structure through the use of new mechanisms in collecting user requests.

The DIODE concept is particularly well suited to teletext systems transmitted via cable networks. It allows efficient use of a return channel to remotely control the downloading of teletext pages into multipage decoders built into television sets.

Paper presented by: Yves Noirel

Born in 1946, graduate of the Ecole Nationale Superieure des Telecommunications/TDF (Telediffusion de France), Senior Engineer/Director of the Data Broadcasting Protocol Laboratory at the CCETT/Rennes in France.

I — INTRODUCTION

In just a few years' time, data broadcasting has gone from the status of a concept greeted somewhat dubiously to that of a technology internationally recognized as very promising. This evolution was brought about by means of a double distinction: first of all, a distinction from the traditional image that radio and television had given to broadcasting; secondly, a distinction from the restrictive image that teletext was beginning to give to data broadcasting.

Avoiding this double pitfall meant providing a definition of broadcasting which would be independent from the type of programs broadcast. This approach was thought to be particularly legitimate in that it constituted a step toward the separation, felt to be increasingly necessary, between information carrier and information provider, between container and thing contained. It thus became possible to build a definition which was no longer focused on one particular presentation of broadcast information, the audiovisual presentation. The specificity of broadcasting is to be found in that particular type of presentation.

That which is specific to broadcasting, and which distinguishes it from other systems of correspondence between two users, is the notion of an audience. The enhancement of the services offered by broadcast systems must therefore take place by broadening the means of configuring this audience. In today's radio and television, programs are established and then transmitted by the press. The schedules indicate, more or less accurately, the time of broadcast, that is, the time when the various products will be available to the users. The influence the user can exert on this programming is one that has a very long response time, by means of surveys that provide a basis for evaluating whether the supply is adequate to the demand. Broadcasting digitally coded information provides a means of broadening the notion of a program, due to the ease with which this type of information can be stored at both transmission and reception end.

The design of a data broadcasting system must therefore be based on two kinds of research, which are complementary: research concerning the set of rules governing the exchange of information between transmitting and receiving equipment, and research concerning the strategies to be implemented in receiving equipment in order to fulfill the requirements expressed by the transmitted equipment. The two types of research result respectively in the definition of broadcasting protocols and transmission guidelines.

II — BROADCASTING PROTOCOLS

The design of broadcasting protocols was guided by the architectural principles applicable to teleinformatic systems, and known by the name of "Open Systems Interconnection". These principles, initially developed within the I.S.O., are now universally used as working tools to design new telecommunications systems. The central notion is the layer, with each layer grouping functions which are similar, either in their nature, or in the technology employed. The functions in each layer are such that the layer can be totally redesigned to take technological advances into account, or to satisfy the needs of a new application, without altering the interfaces with adjacent layers. Every system is described by seven layers, the first four of which concern transmission, and are therefore involved in the definition of broadcasting protocols. Without enumerating the many advantages of this approach, we would nevertheless mention the most perceptible ones:

- the possibility of designing protocols that can be used for various media, with the adaptation of the protocols to the different media taking place at lower layer(s).
- the possibility of multiplexing several digital channels on a single medium, and, within each channel, of carrying different kinds of information to be used together to provide a service.
- the possibility of offering a separate transport service on each digital channel, each transport service being adaptable to the needs of the user (residual error rate, encryption,

amount of information, average and instantaneous data rates).

The DIDON data broadcasting protocol corresponds to the first four layers described in the NABTS (North American Broadcast Teletext Specification). The application of this protocol for broadcasting data on request is called DIODE.

The exact description of the DIODE protocol is beyond the scope of the present paper. We would merely indicate that this protocol defines the dialogue between the transmitting equipment and the receiving equipment in a system of data broadcasting upon request. It is composed of a list of messages which can be exhanged, a description of these messages and the causal relations defining the way they are linked.

The central idea around which the DIODE protocol is built is that the most important characteristic of broadcasting is not that the signals transmitted are one-way, but rather that the information broadcast can be received by a theoretically unlimited number of terminals simultaneously. On the basis of this idea, the transmitting equipment may be said to be like the conductor of an orchestra with regard to the receiving equipment. This idea is also the basis on which the following paragraph develops the ideas enabling us to situate DIODE within the family of information retrieval systems.

III — TRANSMISSION GUIDELINES

3.1 HOW DIODE IS SITUATED AMONG THE DISTRIBUTION SYSTEMS

3.1.1 Resource Sharing in an Information Retrieval System

3.1.1.1 Information Retrieval Systems

An information retrieval system enables the users to retrieve part of the information put into the system by suppliers (who may be the users themselves). The distribution functions of the system consist in offering the users the possibility of requesting information and in supplying each user with the information requested.

In figure 1, k is the amount of information available in the system. The total volume of information available is:

$$v = \sum_{i=1}^{k} Si,$$ Si being the size of information i

----------1----------
----------2----------
----information---
----------k----------

DISTRIBUTION SYSTEM

\+ \+ \+ \+ \+ \+ \+ \+

USERS

System for Distributing k Items of Information

Figure 1

3.1.1.2 Resources of an Information Retrieval System

The user population produces demands for information retrieval which require, if they are to be satisfied, the consumption of common resources of the system:

- terminal → system communications
- system → terminal communications
- system processing power
- system Input-Output capacity
- system storage capacity
- system links with remote suppliers who provide updated information and serve as memory extensions for storage

A request for retrieval is therefore translated as a request for consumption (figure 2). The system's common resources are divided into three groups, corresponding to the different phases of servicing a request for retrieval (this classification is useful for the evaluation and comparison of performances of different information retrieval systems):

- welcome resources (terminal → system communications, processing required to receive and analyse the request, memory for the context of the retrieval session, etc.)
- resources for acquisition and formatting of information (Inputs-Outputs, links with remote information providers, processing, buffer memory, etc.)
- resources for system → terminal communications

users
interactions
Resources:
\+ (terminal → system communications
\+ (system → terminal communications
\+ (processing
\+ (inputs - outputs
\+ (storage
\+ (links with distant information providers

Resources of an Information Retrieval System
Figure 2

From this point of view, the design of an information retrieval system falls within the general framework of resource sharing in a system where the consumers arrive at random.

We are taking a "telematic" system with a large number of users as our frame of reference. The process of the arrival of users wishing to retrieve information is presumed to be independent from the number of users retrieving information at a given time.

As system resources are finite, a request for information may, if the system is overloaded, not find sufficient resources to satisfy it immediately. This request will therefore have to wait for service until such time as the resources required become available.

Generally speaking, each resource or set of resources is associated with a resource availability queue.

For an indivisible resource, i.e. one monopolized, at a given time, by a consumer, user consumption is expressed in resource use time, called service time.

As the queue associated with the resource is presumed to have infinite capacity, the sharing of a limited resource means a certain amount of time spent by the consumer of this resource. The time spent is the equivalent of the service time (which is greater or lesser, for a given consumption, depending on the importance of the resource) plus the time spent waiting for the resource to be available (this waiting time depends on the distribution in time of the arrival of consumers).

We shall use the following elementary result, from queue theory, concerning the delay due to sharing a resource (with an infinite queue capacity) in the following conditions:

- the rule is first come - first served
- a consumer's service time is constant, with a value of s
- the law of consumer arrival is Poisson's law with a parameter λ (the time between two arrivals follows an exponential law with an average of $1/\lambda$)

Thus the average time (average service time + average queue time) spent by a user is given by the following formula:

$$(1) \quad W = s + \lambda s^2 / (2(1 - \lambda s))$$

3.1.1.3 System → Terminals Communication Resource

If the system → terminals communication resource is the rarest resource in the information retrieval system, it means that system performance will vary with the load. In this paragraph, we suppose that the system → terminals communication resource is the only limited resource in the system (we are not taking other resources into account) and that, at a given time, it may be totally or partially used (dynamically shared data rate) to service a request for information retrieval.

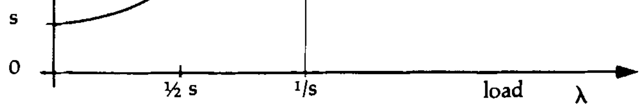
a) The system serves each user independently.

If the system serves each request independently, with the first come-first served rule, the average time (queue + service), called Wind, spent by a user is given by the formula (1) in § 3.1.1.2:

$$\text{Wind} = s + \lambda s^2 / (2(1 - \lambda s)).$$

Figure 3 gives the model for the system with a queue and the Wind curve representing the variations in average time spent according to the average number of arrivals per time unit.

Wind does not depend on k, the amount of information. When the load (described by λ) nears the maximum rate (1/s) admissible in the system, queue time tends toward infinity.



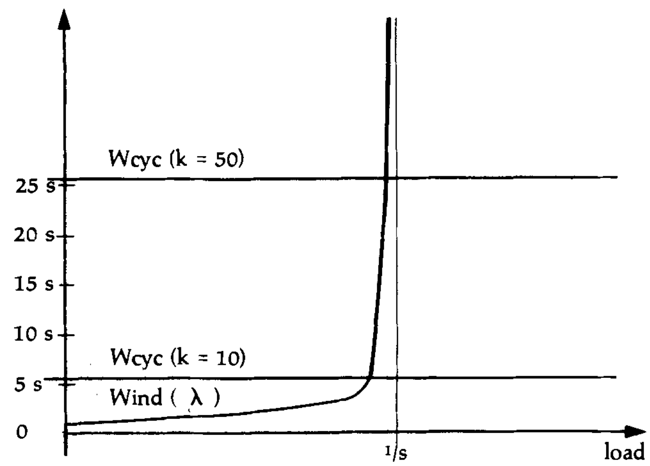
Delay due to resource sharing when the system services requests independently (constant service time)

Figure 3

b) The system broadcasts all information in a repetitive cycle.

If the system broadcasts all the available information in a repetitive cycle, using the totality of the resource, information is broadcast periodically with a period of ks. Requests to retrieve information i are all satisfied simultaneously by the broadcast of information i. The average time spent per request does not depend on λ and is:

$$(2) \quad \text{Wcyc} = s + ks/2$$



Information delivery time with independent service and cyclical broadcast for k = 10 and k = 50

Figure 4

c) The system broadcasts part of the information in a repetitive cycle, and services requests independently for the rest of the information.

A fraction p ($0 \leqslant p \leqslant 1$) of the broadcasting resource is allotted to serving requests for retrieval of information not broadcast cyclically. The choice of p and of the information to be broadcast cyclically uses the average rate λ of request arrival, the probability distribution (P1, P2, . . .,Pk), and the criterion of service optimization. We shall give only a numerical example here:

— the system distibutes 200 items of information (k = 200)
— available data rate is 4 Mbits/s
— each item of information takes an amount of time s = 1/5 second to be broadcast at full rate (which corresponds to a volume of 100 Kbytes)
— requests for retrieval of information i ($1 \leqslant i \leqslant 200$) arrive according to Poisson's law with an average rate of λ i (in our model λ i = Pi) with:

$$\lambda i = \begin{cases} 10 \text{ arrivals per second for } 1 \leqslant i \leqslant 10 \\ 1 \text{ arrival per second for } 11 \leqslant i \leqslant 50 \\ 0.01 \text{ arrival per second for } 51 \leqslant i \leqslant 200. \end{cases}$$

It is not possible to service all the requests independently because the average rate of arrival is greater than system capacity, which is 5 services per second.

If the 200 items of information are broadcast cyclically, the period has a value of 40, hence an average queue time of

20 seconds and average time spent (queue + service) of 20.2 seconds.

If we reserve a fraction $p = 1/3$ of the rate to service requests independently, if information items 51 to 200 are served this way, and if the remaining fraction (2/3) of the resource is allotted to cyclic broadcast of information items 1 to 50, then:

— average time spent for retrieval of one item 51 to 200 of information:

$$s/p + (1/2) \left(\sum_{j=51}^{200} \lambda_j \right) (s/p)^2 / \left(1 - \left(\sum_{j=51}^{200} \lambda_j\right) (s/p)\right) = 3.3 \text{ sec.}$$

— the average time spent to retrieve an item 1 to 50 of information is:

$$s/(1 - p) + (1/2)\, 50\, s \,/\, (1 - p) = 7.8 \text{ sec.}$$

With this same division of information between the two types of service, it is possible, by adjusting (diminishing) parameter p, to obtain identical average times (queue + service) whatever item of information is retrieved. This leads to a use rate ($s \sum_{j=51}^{200} \lambda_j / p$) of the resource allotted to the service "by request" which is very close to 1 (this use rate is already 90% in the previous example), which makes the system all the more sensitive to errors in estimation and to load variations (see Wind curve (λ) in figure 3).

By taking the configuration of the audience (characterized here by the probability distribution $(P_i)_{i = 1, 2, \ldots, k}$ and the average rate λ of requests) into account to divide the information to be distributed between the "cyclic distribution" and "individual service per request" modes, it was possible to improve the average information access time, as compared to that necessary for a system operating in only one or the other of these two modes.

3.1.2 Current Systems of Information for Large Numbers of Users: Cyclically Broadcast Teletext, Interactive Videotex

3.1.2.1 Cyclically Broadcast Teletext

Magazines composed of several pages of text are broadcast in a repetitive cycle. Users consult these magazines on their TV sets by means of a keypad. The pages requested are stored and displayed when they are broadcast. Waiting time does not depend on the number of simultaneous users (an additional user does not entail any additional use of the system's common resources), but it is connected with the number of pages available (see §3.1.1). In order to make an additional page available, it must be broadcast cyclically, thus requiring that a fraction of the broadcast rate be allotted to it, which increases the waiting time for the pages which are already programmed.

As the system does not scrutinize audience configuration, resource sharing between items of information to be broadcast (programming) is not dynamic: information no one wants may be broadcasted.

3.1.2.2 Interactive Videotex

Videotex uses the switched telephone network to serve users interactively.

It provides access to large quantities of information - several hundred thousand pages - which can be distributed among independent data bases with which the user dialogues.

Its performances as regards distribution, which are characterized by the response time and the probability that the user may find all access lines busy, depend on the number of simultaneous users (see §3.1.1 for systems servicing user requests independently).

3.1.3 DIODE

3.1.3.1 Structure

DIODE combines a broadcast channel (system → terminal communications) with the return channels of a cable network or the switched telephone network links (terminal → system communications).

Figure 5 gives a block diagram of the system.

Users send requests to a machine called "Delivery Front End". The Front End is located at the interconnection between three networks:

- The telephone network, or the return channels of a cable network, through which the requests are sent.
- The data transmission network used for exchanges with remote information providers. These exchanges enable the Front End to have updated infromation, and to retrieve information requested that it does not have in its memory.
- The DIDON data broadcasting network through which it sends the information to the users.

3.1.3.2 Delivery Front End Functions

The functions of the Front End include:

- the classical functions of a communication node in a teleinformatic system, i.e. transmission procedure management, message queueing, data flow routing
- information retrieval session management: analysis of user request, exchanges with information providers, management of DIODE protocols at transmission, assignment of DIDON digital channels, respecting transmission parameters, etc.
- service performance optimization: regulation of request arrivals (audience structuring), preparation of a program of information to be broadcast on the basis of requests received, optimum location of files in accordance with configuration, etc.

3.1.3.3 Adaptation to Variations in Audience Configuration

The Delivery Front End's role is to enable DIODE to utilize the notion of a broadcast program while taking audience configuration into account.

It is costly, in terms of broadcasting resources and performances, to add infrequently requested information to a system that broadcasts its contents in a repetitive cycle. It is equally costly to distribute frequently requested information in a system that serves its users independently.

To give a numerical example, we may take the following hypotheses:

Considering:

- two information retrieval systems, A and B, with similar size system → terminal communication resources, that resource being the only limited resource in each of the systems; system A serves users independently, and system B broadcasts the information cyclically
- a group of K = 110 items of information, with retrieval of each item taking the same service time s of the system → user communication resource, and user

requests for each item arriving according to Poisson's law with average rates $\lambda 1$, $\lambda 2, \ldots,$ $\lambda 110$, with:

$$\lambda 1\, s = \lambda 2\, s = \ldots = \lambda 100\, s = 0.005$$

$$\lambda 101\, s = \lambda 102\, s = \ldots = \lambda 110\, s = 0.5$$

We want to divide the information between A and B.

By putting items: 1,2, . . . 100 in A
items: 101, 102, . . . 110 in B

we obtain an average retrieval time of:

Wind (λ = 1/2s) = 1.5 s   for items 1 to 100

Wycy (k = 10) = 6 s   for items 101 to 110

If item 101 is put in A instead of in B, waiting time becomes infinite in system A, which is saturated with requests for item 101.

If items 91 to 100 are put in B instead of in A, with the other items remaining as they were, we obtain an average retrieval time:

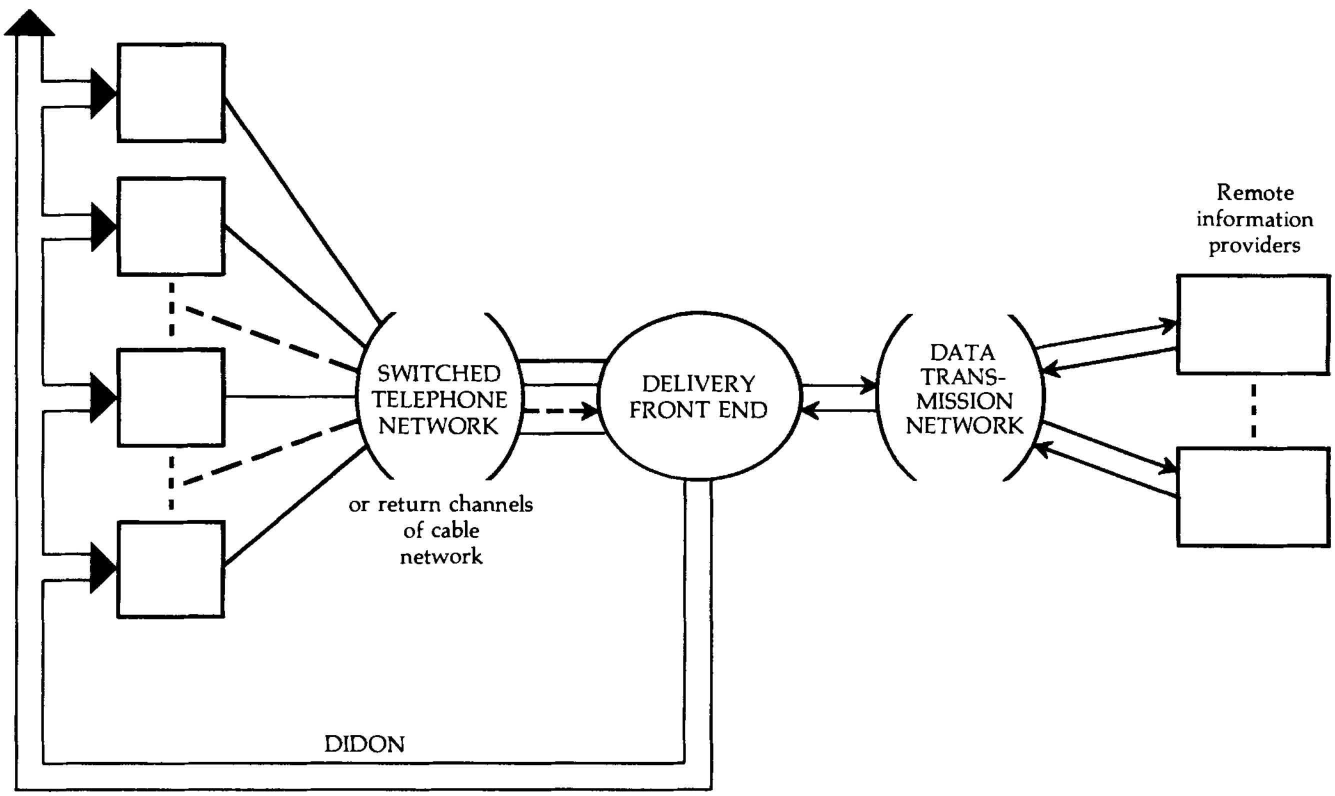
Wind (λ = 0.45/s) = 1.41 s   for items 1 to 90

Wcyc (k = 20) = 11s   for items 91 to 110

The average retrieval time has increased by about 80% for more than 90% of the requests for retrieval. The gain in time for the others is negligeable.

The above numerical example shows the extra cost in resources and in performances brought about by a bad choice of an information distribution system (cyclic broadcasting or independent service for each user).

One of the advantages of DIODE, which operates on the principles of "broadcast on request with grouping of request for the same information", lies in the flexibility with which it adapts to variations in audience configuration, and in its ability to distribute all information without "cost" distinctions according to the frequency of requests.



DIODE SYSTEM
Block Diagram
Figure 5