

Reducing Satellite Communication Cost Using Terrestrial Peer-to-Peer for Lost Recovery

F. de Belleville*

ENSICA/TéSA, 1, place Emile Blouin, 31000 Toulouse, FRANCE

ENSEEIHT /IRIT, 2, rue Charles Camichel, BP 7122 - F 31071 Toulouse, FRANCE

L. Dairaine.†

NICTA, National ICT Australia, Australian Technology Park, Eveleigh NSW 1430, AUSTRALIA

ENSICA/TéSA, 1, place Emile Blouin, 31000 Toulouse, France

M. Gineste.‡

ENSICA, 1, place Emile Blouin, 31000 Toulouse, France

C. Fraboul§

ENSEEIHT /IRIT, 2, rue Charles Camichel, BP 7122 - F 31071 Toulouse, FRANCE

A practical solution to implement IP multicast service may consist in using a geostationary satellite. The broadcast nature and the large coverage zone of such systems make it possible for a source to reach a potentially very large number of receivers with only one hop. In the context of reliable multicast communications, a hybrid satellite/terrestrial approach based on communication costs is described. Due to particular data dissemination resulting from the satellite communication phase, ad-hoc lost recovery peer-to-peer mechanisms are specially designed and evaluated through simulations.

I. Introduction

The support for point to multipoint communications is a useful and performing service that is not sufficiently available in nowadays Internet networks. Numerous applications like multimedia streaming or software deployment and update would indeed take a great advantage of such a service. IP Multicast [1] offers an appropriate multipoint service in the context of the ubiquitous Internet Protocol (IP). Nevertheless, its availability for large communication groups over the whole internet is still very limited for the time being. This lack of deployment is mainly due to technical concerns and economical issues [2].

A practical solution to implement IP multicast service may consist in using a geostationary satellite. The broadcast nature and the large coverage zone of such systems make possible for a source to reach a huge number of receivers with only one hop. Satellite broadcasting may seem expensive at first sight, but per-receiver cost becomes less than using terrestrial network when the number of receivers increases. For this reason this study is focused on large scale reliable multipoint communications. Moreover we consider applications with no time constraints, because the long transmission delay of satellite links is not really compatible with such applications. Software updates or cache feeding in content delivery networks are examples of possible target applications.

The main objective of the proposed approach, named Hybrid Satellite Terrestrial Reliable Multicast (HSTRM), is to provide a low cost fully reliable multipoint service. Thus the satellite link may be used only to transmit packets useful for a large group of receivers. A way to partially achieve this goal is to use Hybrid ARQ type II [3] for retransmissions: with this technique, any packet transmitted can be exploited by any receiver waiting for information. The remaining issue is to ensure that the satellite link is not used for a small group containing few receivers only. During a satellite transmission the receivers which experienced losses have indeed to wait for more information to recover those losses. In consequence more time is required to transmit data to them. Thus the number of receivers which continue to really use the satellite broadcast will decrease all along the transmission. To solve this

* A.T.E.R, (Lecturer/Research assistant) Ingénierie des Réseaux et Télécommunications.

† Assistant Professor, Networking and Pervasive Computing group, NICTA

‡ A.T.E.R (Lecturer/Research assistant), Ph.D. Student ENSICA.

§ Professor, Head of Telecommunication and networking Department, ENSEEIHT

problem, our approach consists in estimating regularly the broadcast audience during the transmission [4]: data are then broadcasted via satellite only when the number of receivers is sufficiently high (i.e. when per-receiver cost is less than with terrestrial networks) and via terrestrial transmissions otherwise.

The present paper focuses on issues related to the terrestrial phase of the communication. In our context, peer-to-peer appeared to be the most suitable technique for seeking and exchanging data over the terrestrial network. Nevertheless, because of the atypical context (hybrid communication, satellite broadcasting, use of Forward Error Correction) and the specific objective of the approach (minimize the overall communication cost), a set of *ad-hoc* mechanisms has to be proposed. This led us to the proposition of a specific mechanism for terrestrial error recovery which takes advantage of both the satellite and the terrestrial networks. Some of its parameters are set according to theoretical computation results. The remaining ones are set thanks to simulation results. Performances of the proposed mechanism are eventually studied through simulations.

The paper is structured as follow. In the second section, the context of the study is described. The specific communication environment and the multicast transport service are then presented. Then, the principles of the Hybrid Satellite Terrestrial Reliable Multicast (HSTRM) protocol are given. A statistical study comparing costs for terrestrial, satellite and hybrid terrestrial/satellite multipoint is finally proposed. In the third section, the paper focuses on the terrestrial phase of the HSTRM protocol with the complete definition of an efficient low-cost peer-to-peer mechanism. In the last section, parameter settings and performance study of the mechanism are provided.

II. A Reliable Transport Protocol Designed for Hybrid Satellite-Terrestrial Network

In this section, the context of the study is specified. Then the target transport service is described, as well as a proposition of a hybrid satellite/terrestrial approach to achieve a reliable multicast transport service. The last part presents a statistical study of this approach.

A. Communication System

The considered system is a hybrid satellite/terrestrial network, i.e. end-users are connected to both satellite and terrestrial networks. Assumptions have been made in the context of the French DIPCAST project [5]: end-users are connected to terrestrial network via a high speed access network (e.g. xDSL or LAN), and to satellite system either via a high speed access network, or directly with a Very Small Aperture Terminal (VSAT).

The satellite system uses a geostationary satellite, and proposes a best effort multipoint communication service based on IP Multicast. This supposes that a protocol which manages joining and leaving procedures is integrated to the satellite system, as well as tree establishment algorithm. This problem of integration has addressed out during the DIPCAST project but is out of the scope of this paper.

In order to be representative of today Internet, the terrestrial network is not supposed to support multipoint transmissions. Thus any terrestrial communication in this hybrid system is a point-to-point transmission. Finally the following assumptions are made on the application:

- The application transmits data from one source to a large group of receivers (one-to-many communications)
- This application does not have strong delay constraints.
- A service transmitting session characteristics (file properties, start time, associated group address, etc.) is available. Those transmissions can be done via out of band means (e.g. e-mail), or via session management tools.
- No receiver can join the session after the beginning (late joining is not supported).

B. Multicast Transport Service

Since satellites are really advantageous for large scale data transmissions, applications which transfer files to a large number of receivers (several hundreds or more) are particularly considered in the present paper. Furthermore we consider applications requiring a full reliability, i.e. which must be assured that the whole group has received transmitted information. An example of such applications is the transport of multimedia files (e.g., video, music, games, etc.) towards a large set of users. Another considered issue concerns the overall communication cost. The utilisation of satellite links is indeed quite expensive. Nevertheless when the number of receivers increases, the per-user cost decreases. Thus any application being charged according to its bandwidth utilisation may prefer protocols which carefully watch communication cost.

Considering file transport using the best effort protocol IP Multicast over satellite, a fully reliable transport protocol must be used. In the following paragraphs some features of this protocol required in the above-mentioned context are exposed.

The transport protocol must guarantee that all receivers in the group receive transmitted information. Several multicast transport protocols propose a statistically reliable service [8]. Although it allows designing transport protocols with no return channel, full reliability is not ensured because there is no adaptation to losses that really happened. This technique is then not convenient for the aimed purpose.

Since satellite bandwidth is expensive, the transport protocol should ensure that any useless satellite transmission is avoided. This confirms that statistical reliability is not recommended in the presented context because systematic coding of information implies a potential waste of bandwidth [6]. Finally, as considered applications are designed for transmissions towards very large groups, underlying protocols (and specifically transport protocol) must scale very well. In particular for the transport layer, mechanisms of feedback suppression like [7] must be studied and configured for a satellite link.

C. A Hybrid Satellite/Terrestrial Approach

Numerous multicast transport protocols have been designed in the last few years to achieve efficient and scalable multicast transmissions [8]. From all researches on reliable multicast transport, a technique referenced as Hybrid ARQ type II [3] has emerged. It is an efficient way to reduce the used bandwidth and to improve the scalability. It consists in using FEC combined with Automatic Repeat reQuest (ARQ) in the following way: after a transmission, the source asks for the maximum number M of missing packets. Then it generates and transmits M new encoded packets. As those M packets have not already been transmitted, they are useful for any receiver which experienced losses. Hybrid ARQ type II allows to radically reduce the amount of retransmitted information in multipoint transmissions, and is then particularly interesting for any large scale full reliable transport protocol [3].

Most of the protocols presented in [8] are usable with satellite links because they support asymmetric transmissions. Nevertheless no previous work considers the overall communication cost. With the used reliability mechanisms, either packets are systematically retransmitted to the whole group, or each missing packet is retrieved with a point-to-point connection. In our context when point-to-point retransmissions are used, if a packet is requested by numerous receivers, satellite broadcasting may be profitable. On the contrary, if every missing packet is transmitted to the group via satellite, when only few receivers are concerned, few terrestrial point-to-point connections can be cheaper. According to this simple statement, the trade-off between satellite and terrestrial bandwidth use may be studied, and it is our belief that optimisation of this trade-off is an interesting way to reduce the overall communication cost.

An interesting approach may be to define a threshold R_{\min} representing the minimum number of receivers for satellite broadcasting to be advantageous (taking economical costs into account) [24]. A session would then behave as follows: at a predefined time, the satellite starts broadcasting over the IP multicast group. During this transmission, the source periodically estimates the session size. A receiver is considered to belong to the session as long as it has not received the whole transmitted information. Several papers have addressed the question of estimating multicast session size [9]. An appropriate mechanism has been proposed in the specific context of HSTRM [4] and is assumed to be available in this paper. Thus once the entire initial information has been transmitted, session size is likely to decrease (all receivers which experienced no losses quit the session). The source then goes on estimating session size while it transmits encoded redundancy packets to repair losses. When estimated session size goes below R_{\min} , the satellite transmission stops. Receivers which do not have enough information to decode received data (i.e. receivers experiencing high loss rates) then contact other receivers to recover missing data. This terrestrial recovery can be done using peer-to-peer services. When all receivers have fully received the information, the session stops.

D. Statistical Study

The present paragraph gives an illustration of the gain generated by the hybrid satellite/terrestrial approach. For that purpose a statistical study is presented, with the following assumptions:

Errors and Fades

For terrestrial transmissions, packet losses are only supposed to be produced by network congestions. These congestions cause at the transport layer a Packet Loss Rate (PLR) of 5% for point-to-point communications [10] and 10% for multipoint communications [11].

For satellite transmissions, three categories of receivers are considered. The first one corresponds to all receivers under a clear sky. Assumption is made that these ones do not experience any losses (possible issues due to scintillation are not considered) and that this category encompasses 90% of the receivers. The second category is supposed to include 9.9% of the receivers, and corresponds to end-users under a rainy sky. Fades due to light rain are supposed to cause a PLR of 20% at transport level. Finally, the last category encompasses receivers under a

stormy weather. This category is supposed to include 0.1% of the receivers which experience a PLR of 60%. According to [13] those values are realistic for a satellite communication using the Ka Band.

Note that all losses are supposed to be independent and uniformly distributed among packets. This assumption is not realistic, but since transport layer is supposed to implement ARQ type II technique, only the amount of lost packets is important for our computations. Assumption was also made that no packet are lost on the return channel.

Cost Function

In order to compare costs generated by hybrid satellite/terrestrial multipoint communications with pure terrestrial and satellite multipoint communications, it is necessary to first define a cost function. We choose to adopt a per-packet cost approach, and then define a cost function as follows:

$$F_x(R, K) = \alpha_x \times C_x(R, K), \quad X \in \{TU, TM, SM\}, \quad (1)$$

where K is the number of packets to transmit, R the number of receivers and α_x the per-packet transmission cost. $C_x(R, K)$ represents the average number of packets passing through the network in order to transmit K packets to R receivers (taking losses into account). The indices TU, TM and SM correspond to Terrestrial Unicast, Terrestrial Multicast and Satellite Multicast communications.

For terrestrial point-to-point communications R connections experiencing independent and uniformly distributed losses (represented by the PLR) are used to transmit K packets to R receivers. Then:

$$C_{TU}(R, K) = K \times R \times \sum_{i=1}^{\infty} i \times (1 - PLR) \times PLR^{i-1} = \frac{K \times R}{1 - PLR}. \quad (2)$$

For multipoint communications, the probability $P(N, PLR, R)$ that exactly N packets are needed so that all end-users receive K packets can be expressed as: the probability that each end-user receive K packets among N , minus the probability that R receivers receive K packets among $(N-1)$. Then $P(N, PLR, R)$ can be calculated as:

$$P(N, PLR, R) = \begin{cases} \left[\sum_{i=0}^{N-K} \binom{N}{i} PLR^i (1-PLR)^{N-i} \right]^R - \left[\sum_{i=0}^{(N-1)-K} \binom{N-1}{i} PLR^i (1-PLR)^{(N-1)-i} \right]^R, & \text{when } N > K \\ \left[(1-PLR)^K \right]^R, & \text{when } N = K \end{cases} \quad (3)$$

For terrestrial multicast communications, according to [12] transmitting data to R receivers is equivalent to $R^{0.8}$ point-to-point connections. The average number of packets passing through the network is then:

$$C_{TM}(R, K) = \left[\sum_{N=K}^{\infty} N \cdot P(N, PLR, R) \right] \times R^{0.8}. \quad (4)$$

For satellite multipoint communications, since we defined three categories of receivers and since no multicast tree is established, we have:

$$C_{SM}(N, K) = \max \left\{ \left[\sum_{N=K}^{\infty} N \cdot P(N, PLR_i, \beta_i R) \right], i \in [0, 2] \right\} \quad (5)$$

where $PLR_0=0$, $PLR_1=20\%$, $PLR_2=60\%$, $\beta_0=90\%$, $\beta_1=9.9\%$ and $\beta_2=0.1\%$.

For hybrid satellite/terrestrial transmissions, the cost is defined as the sum of the costs generated by terrestrial network utilization, and satellite system utilization.

Results

Using the cost function defined above, R_{\min} has been computed for group sizes ranging from 100 to 600,000. Then the hybrid satellite/terrestrial approach has been compared to terrestrial point-to-point and multipoint communications as well as to pure multipoint satellite communications. Figure 1 shows the results for $\alpha_U = \alpha_M = 1$, $\alpha_S = 100$ and $K = 100$. The different levels perceptible on the curve representing the hybrid communication cost are due to the model definition: when group size increases, the number of receivers under a rainy sky or a stormy sky increases as well. Thus it becomes necessary to repair more and more losses using satellite link. In this example, the hybrid satellite/terrestrial approach induces a gain ranging from 10% to 50% compared with the most advantageous of the three classical approaches.

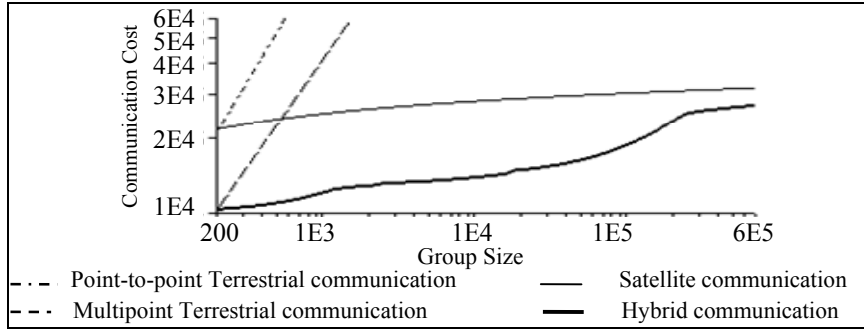


Figure 1: Cost generated by terrestrial, satellite and hybrid communications.

$$\alpha_{TU} = \alpha_{TM} = 1, \alpha_{SM} = 100 \text{ and } K = 100$$

The hybrid satellite/terrestrial approach relies on hybrid ARQ type II. As mentioned previously this technique implies to encode transmitted information. The remaining of the paper is focused on the design of an efficient low cost mechanism for the terrestrial phase of the hybrid approach.

III. An Efficient Low-Cost Peer-to-Peer Mechanism for Terrestrial Communications

The terrestrial phase of transmission starts as soon as the number of receivers became lower than a cost-related threshold R_{\min} formally defined into [24]. When the number of receivers is lower than this threshold, which will be noted MinRcv , satellite transmission is supposed to be not advantageous any more. During this phase, receivers seek missing data in order to supplement those which they received during the satellite phase of transmission. One major need for this approach is a mechanism for retrieving packet losses using the terrestrial network. In our context, peer-to-peer appeared to be the most suitable technique for exchanging data over the terrestrial network. Nevertheless, the analysis of existing peer-to-peer mechanisms reveals that none of these proposals take the cost criteria into account. Moreover, most the classical mechanisms are designed either for conditions where the localisation of data stored into only few peers or for the dissemination of a large amount of data.

A. Objectives and hypothesis

When the satellite transmission phase stops, the data repartition over the network is slightly different from the classical hypothesis taken in the context of peer to peer networking. HSTRM proposes a very large scale group communication service. As shown previously, if the satellite transmission stops, this means the number of receivers having not received the information has become relatively small in regards to the total number of receivers. Then, only a small fraction of the total number of receivers does not have all the information. Nevertheless, the protocol does not know how many data the lossy receivers got during the satellite transmission, with possible receivers having received no information at all. Moreover, considering the proximity in the network is not related to the geographical proximity, the system does not have information about the repartition of the information into the network. Recall that all receivers are connected to the terrestrial network by a high speed access network such as xDSL line.

HSTRM is designed to be used in the context of large group; then the terrestrial mechanism must be also suitable for large scale network. This means that the P2P mechanisms should not have an important impact on the underlying network or the receivers. This is particularly true for the P2P network establishment because all group members are involved into this phase. To allow an easy deployment of the protocol, the terrestrial mechanism should not impose specific requirements on the underlying network, such as IP multicast over wired network.

Finally, a very important constraint is related to the cost of use of the terrestrial communication. For the global proposal to be of interest, the proposed terrestrial recovery mechanism should not be more expensive than a simple satellite broadcast or terrestrial multicast.

B. Related work

In such a way to ensure HSTRM aims, different types of approaches have been analysed, and particularly the techniques of cooperative downloads, application-level multicast and peer-to-peer downloads.

The first type of mechanism is known under the appellation of cooperative downloads. In this case, the main objective is to relax the load imposed on a centralized server. This situation classically occurs when a server has a very popular content such as a software update. This problem has been addressed e.g. by Slurpie [14] and Bit

Torrent [15]. The aims of these solutions are to replicate as quickest as possible a large file into a set of distributed hosts. To achieve this, new receivers that want to get the file are redirect towards other users which already have partially or totally the file. The load is then distributed on the various participants to this download. Nevertheless, the main objective of these contributions differs from HSTRM. During the terrestrial phase, the point is not to duplicate content on the whole group where nobody has the information. At this phase of communication, only few receivers are directly concerned and usually look after a small amount of data (i.e., the blocks of data that have not been received during the broadcasting phase).

Another classical solution to allow a group of receivers to communicate with other receivers consists in using an application-level multicast service that consists in managing the data replication to various receivers in end-system high layer. This approach has been addressed in many studies [22,23]. Nevertheless, as the previous type of problem, the objectives differs from ours in transmitting a file to a whole set of receivers and the context is a transmission from a source to the group.

Another way to disseminate data among a set of receivers make use of peer to peer techniques. CAN (Content Addressable Network [16]) and CHORD [17] are two classical techniques proposed to localise efficiently information in a P2P network. They make use of hash functions: when a node looks for specific information, it calls the hash function that computes an identifier. This identifier can be considered as the information's address into the P2P network. This identifier is then searched into the network. These approaches allow localizing information among a set of receivers which are offering large information diversity. Recall to our context, the receiver looks after data that are already present in a large set of receivers. The localization then becomes far less complex than in a classical P2P context. Almost all receivers have a neighbour that already owns the requested information.

Local recovery mechanisms allow requesting retransmission to hosts located "near" the receiver. The problematic addressed into these works are then very related to our context. A proposal consists in dividing the receivers into several sub-groups, each sub-group having a host in charge [18]. When losses occur in a receiver, it seeks missing information within the sub-group. If nobody has this information, the host in charge contacts the source for that it retransmits the data by satellite. This approach targets a service very near HSTRM needs. Nevertheless, the choice to use a FEC coding at transport level can modify the properties of considered system. Moreover the study does not tackle at all the problems related to the P2P network setup, the choice of the host in charge, and the search for information. Finally, loss can lead to satellite retransmissions, while in our case all data must be transferred by means of the ground network.

The objective for a large number of works is to minimize the duration of the information downloading. To satisfy this need, several mechanisms are used to make the P2P network to converge towards a more powerful network. However, in our particular case study, among the whole number of peers, a large number of receivers will not make any requests (i.e. the number of receivers that entirely got the data is large from the main principle of the proposed approach). It is thus useless for the P2P network to arrange the peers in such a way to bring them closer between each other: they just must remain reachable.

Among all referred work, none takes into account the constraint of cost related to the use of the network. However, taking into account such constraint makes impossible for example the use of mechanisms allowing making the network of peers to converge (because these mechanisms imply a considerable additional use of the network). That also explains the little number of studies studying the setting of the P2P network in the error recovery context: the number of generated messages is not a criterion taken into account.

There is one considerable advantage related to the use of the hybrid approach: after the satellite diffusion phase, the FEC encoded information is naturally present in most of the peers. Consequently, there is no need of any replication and dissemination policies (see for example [19,20,21]). On the contrary, this natural dissemination of encoded data should be used to enhance the recovery phase. Nevertheless, the hybrid approach is associated to a certain number of problems specific which are not considered in the analyzed approaches. For example, none of the studied systems consider a large proportion of users already having the data (the opposite case is generally considered). This assumption allows simplifying the searching mechanisms. Finally the assumption of transmission in a (very) large groups, associated to the fact that this network is set up at particular instant (when satellite transmission stops), induces scalability problems. The problems presented above leads to propose a set of ad-hoc mechanisms tailored to adapt the studied context. These mechanisms are detailed in the following section.

C. Mechanisms for the terrestrial recovery phase

Principal objectives of the terrestrial mechanism for recovery errors come directly from the preceding remarks. They must be designed:

- to allow information finding among the whole receiver group;
- to avoid an expensive use of the terrestrial network;

- to be used in a large scale context, without implying unacceptable waiting times from the receivers point of view.

To achieve these aims, a P2P network is established so that the receivers can communicate between them. This network consists in a regular logical v -ary tree (see Figure 2).

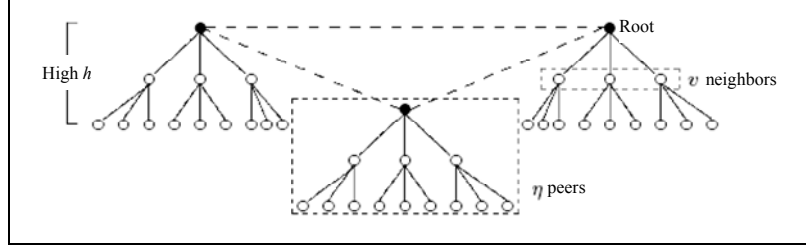


Figure 2. Logical representation of the P2P network.

Several advantages are related to such receiver's organization:

- the network creation is simple;
- this limits the number of known peers from a given receiver, which avoids the risks of message implosion and distributes the storage need to memorize the state over the tree;
- the tree traversal is relatively fast.

In the following mechanisms presentation, the number of logical trees is noted τ , the number of peers in each tree is η_i , and the height of trees is h_i (with $1 \leq i \leq \tau$). To constitute this network, it is necessary to have roots (which are the base of the network). Then, the source collects the addresses of the set of receivers which answers to the **Inquiry** messages during the initial phase of HSTRM [25] (those answers are also used for the estimation of the size of the group [4]). It then transmits these addresses to the entire group by means of messages **P2PRootList**. Those messages are sent every **P2PRLPeriod** units of time, allowing lossy receivers to get them. The source sends with this list a duration **TMaxP2P**, corresponding to the time the receivers have to be connected.

When the reception of the first message **P2PRootList** occurs, the receivers compute randomly a latency ranging between 0 and **TMaxP2P**. When the associated timer expires, the receivers choose randomly a root among the received list, and sends a message **ConnectP2P** to the selected root, specifying their IP address. If the selected root does not answer, another root is chosen.

The first reachable root will answer by a message **PeerConnected**, if it has less than v neighbours. In this case the considered receiver becomes a direct root neighbour. In the other case, the root returns a message **Redirected** which specifies a dependent peer, to which the newcomer receiver must contact. The receiver sends to the specified peer a message **ConnectP2P**. The peer answers by a message **PeerConnected**. The peer that the newcomer tries to reach can however be not reachable. In this case, the new peer will then contact its root which will retransmit the address of another peer.

These operations are illustrated in

with an example where $v=3$ ($@_i$ represents the address of receiver i). Moreover, in such a way to avoid a partition of the network because of a node breakdown, the peers keep in memory a certain number of potential root addresses. Thus, if a receiver becomes isolated (i.e., its neighbours are not reachable), it can always contact the other roots and to have access to the remaining of the P2P network. Any peer thus has access to the whole network.

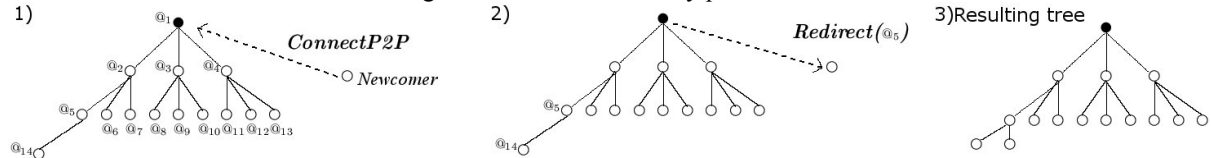


Figure 3. Connection of a newcomer into the P2P network.

IV. Parameter settings and Performance study

The following sections presents the choices on v , τ and **TMaxP2P** which was determined either by the way of a theoretical study or by simulations. The cost of the P2P network setup, the memory required to represent the network state, and the temporal spacing of connections are studied through theoretical calculations. Then, the cost

and the efficiency of information retrieval are studied by the way of simulations. Let us note that the results obtained by means of theoretical calculations were also confirmed by simulation.

A. Cost to setup the P2P network

It is possible to compute the number of messages exchanged during the establishment of the P2P network, when the process described in the paragraph III.C is used:

- For the τ roots, no message is used;
- For the $\tau \times \nu$ peers directly connected too the τ roots, 2 messages are exchanged ;
- For the $R - \tau \times (\nu + 1)$ other peers, 4 messages are necessarily used.

Denote as $F_{P2P}(R, \nu, \tau)$ the cost to setup the P2P network. According to the previous remarks, the expression of $F_{P2P}(R, \nu, \tau)$ is (where α_{TU} is the terrestrial per-packet transmission cost):

$$F_{P2P}(R, \nu, \tau) = \alpha_{TU} \times (2 \times \tau \times \nu + 4 \times (R - \tau \times (\nu + 1))) \quad (6)$$

This value constraints the minimum size of the data that can be transmitted so that the hybrid communication remains advantageous. According to the paragraph II.D, the cost related to the transmission of K packets is $F_X(R, K)$. And when S bites are to be sent, $\lceil S / (K \times TDU) \rceil$ blocs of K packets are needed so that the whole file is transmitted. In consequence, the cost C_{file} related to the transmission of the file is:

$$C_{file} = \lceil S / (K \times TDU) \rceil \times F_X(R, K) \quad (7)$$

It is then possible to compute the minimum size S so that the cost to setup the P2P network does not exceed a certain percentage of the file transmission cost. For example with $R=600,000$, $\nu=10$ and $\tau=166$ (those values are explained below), the condition $F_{P2P}(R, \nu, \tau) \leq 0.1 C_{file}$ implies that more than 60 MB have to be sent.

B. Memory requirements

According to the process described in the paragraph III.C, all the root either accept the newcomers, or redirect them toward a peer located at the bottom of the logical tree. In consequence the roots have to keep the addresses of the peers located at the bottom of the tree (see Figure 4). This paragraph focuses on the evaluation of the memory required for that purpose.

With the notations introduced into the Figure 2, the following inequalities are respected: $\sum_{i=0}^{h-1} \nu^i < \eta \leq \sum_{i=0}^h \nu^i$. In order to approximate the amount of memory needed, we only consider the case when the equality is verified (this is a conservative approach):

$$\eta = \sum_{i=0}^h \nu^i = \frac{\nu^{h+1} - 1}{\nu - 1} \quad (8)$$

from (8), we deduce: $h = \ln(\eta(\nu - 1) + 1) / \ln(\nu) - 1$. In our context, R receivers are distributed among τ logical trees. Thus each tree contains on average R / τ peers. In consequence each root have to keep L addresses in memory,

where $L \propto \nu \left(\frac{R}{\tau} (\nu - 1) + 1 \right) / \ln(\nu) - 1$. The objective is then to find convenient values for τ and ν . When ν increases, the number of neighbours increases for each peer, and, as a consequence, the robustness of the overall network improves. Nevertheless, the number of addresses stored increases for each root. The parameter τ has equally a significant impact: when it increases L decreases. In exchange the network becomes more divided. The Figure 5 shows the amount of memory used for the addresses storage for different values of τ and ν .

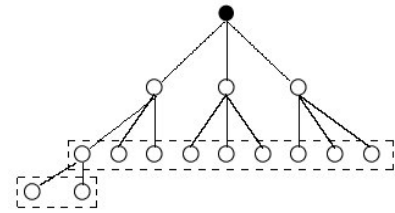


Figure 4: addresses kept in memory ($\eta = 15, \nu = 3$)

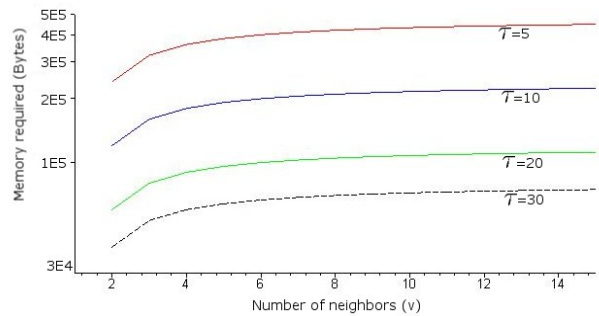


Figure 5: Memory requirements (R = 600 000).

As it can be shown in the Figure 5, the parameter τ allows reducing drastically the memory requirements. Once this parameter is fixed, ν has only small effects on memory space. As a result, the choice of $\tau = 30$ when $R=600000$ requests only less than 80Ko per root.

In addition, each node of the tree has to keep also IP addresses of a number of peers to ensure a robust connectivity. As a result, the nodes needs to keep the addresses of known roots, the addresses of all intermediary nodes between them and the root and the addresses of their child. The number of addresses stored is $\tau + h + \nu$. For realist values of τ , either this size is negligible in regards to L , or R is sufficiently low for the number of stored addresses to be not a real problem. Only constraints on L have to be kept in mind.

C. Connections spacing

The choice of τ is not only dependant of memory considerations. In fact, the value of this parameter has an important impact over the number of requests sent to the roots. The principle of the connection spacing mechanism is fairly simple: it consists to distribute over time the connection requests from the receivers, in such a way to avoid root congestions. The very classical Nack Suppression mechanism is used as follow: when a **P2PRootList** is received, each receiver that are not part of the list wait for a time between 0 and $TMAXP2P$ before replying. The present section studies the issues related to the choice of τ and $TMAXP2P$. The average number of connections requests A_C that each root receive is connected to these two parameters: $A_C = R / (\tau \times TMAXP2P)$. Several constraints must then be taken into account:

- A_C must not be higher than a predefined threshold A_{CMax} .
- The time $TMAXP2P$ which represents the P2P network setup time must not be higher than a maximum, denoted as $TMAXP2P_{Max}$, so that the time to establish the P2P network remains acceptable.
- τ should not be more than the average number of answers collected for a single **Inquiry** message so that a simple exchange of message is sufficient to get all the addresses needed. Denote as τ_{moy} this average number.
- τ should be chosen in such a way that each logical tree contain at least η_{min} peers (i.e. the network is not too divided in regards to the total number of peers).

Considering the previous constraints, the choice of the parameter τ and $TMAXP2P$ can be done with the following algorithm:

```

If ( $R < \tau_{moy} \times A_{CMax} \times TMAXP2P_{Max}$ ) Then {  $TMAXP2P_{Max} = R / (\tau_{moy} \times A_{CMax})$  }
    If ( $R < \eta_{min} \times \tau_{moy}$ ) Then {  $\tau = R / \eta_{min}$  }
    Else {  $\tau = \tau_{moy}$  }
Else {
     $\tau = R / (A_{CMax} \times TMAXP2P_{Max})$ 
     $TMAXP2P = TMAXP2P_{Max}$ 
}

```

Figure 6: Algorithm for the choice of τ and $TMAXP2P$

After the definition of this algorithm, our objective was to study the performances of the network established. For that purpose, several simulations were conducted.

D. Simulation results

A simulation study has been achieved to verify the effectiveness of the techniques described in the previous paragraphs. Simulations have been carried out also to determine a suitable value for the number of neighbors. The simulation software has been developed with Java language. All the simulations have been setup with the following parameters:

- $\tau_{moy} = 30$: This value is coherent with the group size estimation study provided in [4]. Then, when the number of root is lesser than τ_{moy} , only one Inquiry message is necessary to obtain all the necessary addresses.
- $A_{cMAX} = 20$ message per seconds corresponding approx. to a low throughput of 3kb/s

- $TMAXP2P_{MAX}=180s$ in such a way the value of $\tau = \tau_{moy}$ per seconds corresponding a throughput of 3kb/s
- $600 \leq N \leq 600000$ receivers.

A large number of simulations have been achieved in this context, in a first time when all the receivers of the initial group are reachable during the terrestrial phase, and in a second time when a subset of the original receivers are unreachable. In this context, the studies cover various performance aspects of the terrestrial phase, particularly the memory requirements, the load over the network, the cost linked to the information localization and retrieval. The second set of simulations was also addressing robustness issues in evaluating the impact on the overall mechanism of a set of unreachable receivers. In this paper, we only consider a subset of the studies achieved when all receivers are reachable. Only the results concerning cost and efficiency of information search is provided. The complete results of study aiming at determining the impact of a set of non reachable receivers on the provided service is available in [25].

Information searching cost

The objective of this set of simulations is to ensure that the cost related to the search of data is not too high; so that the proposed hybrid approach remains advantageous. In the achieved simulations, **MinRCV** (the value of **MinRCV** is chosen according to [24]) are randomly chosen among all the receivers: they represent the receivers which do not receive the entire information. In consequence, they contact other peer in the overlay network to retrieve lacking information. The number of search messages is then computed. It appears in the experiences that when the group size increases, the number of messages decreases. This behaviour is due to the fact that the ratio of lossy receivers decreases when the group size increases (**MinRCV** is independent of the group size). In consequence the probability that a peer is connected to another peer who received the whole information increases. Figure 7 shows the number of generated messages for different values of \mathcal{U} with 600 receivers and 300 lossy receivers (each curve represents the average curve obtained from 300 Monte Carlo runs). According to the Figure 7 several cases may be distinguished:

- The receivers who are a leaf of the logical tree and whose father received all the data generate only one search message.
- The receivers located at the middle of the tree and whose father and son received all the data generate $(\mathcal{U} + 1)$ search messages.
- The remaining lossy receivers generate more messages because the request is propagated in the tree.

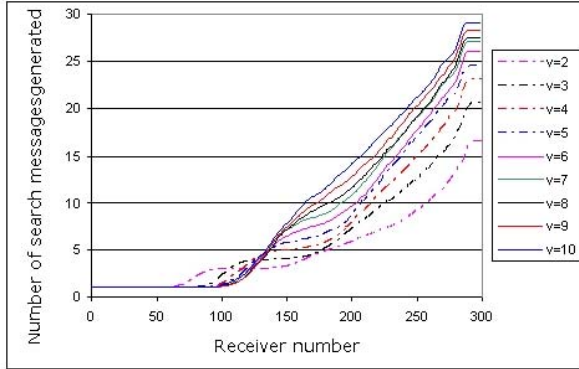


Figure 7: number of search messages generated for each lossy receivers. $R=600$, $MinRCV=300$

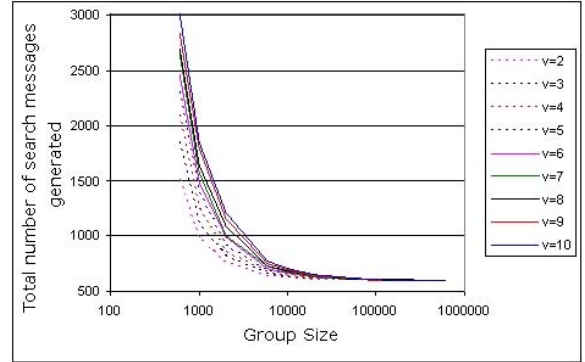


Figure 8: total number of search messages generated

Moreover, according to this figure, when \mathcal{U} increases, more peers generate only one message, but the total number of messages generated (represented by the integral of the curve) increases. The remaining issue is then to find a trade-off between the total cost of the search phase and the efficiency of the proposed mechanism. Figure 8 presents the evolution of the total number of search messages generated as a function of the group size for different values of \mathcal{U} . As one can see on this figure the influence of the parameter \mathcal{U} decreases when the group size increases. Moreover, when the group is more than 6,000 receivers, the cost is almost independent of \mathcal{U} . As a consequence, the value of \mathcal{U} should be fixed as 10 because it increases the robustness of the network and the efficiency of the mechanism.

Searching performance

We evaluate the performance as *i*) the number of peers who successfully retrieve the information and *ii*) the minimum number of hops realised in the overlay network before the information is found. For the first criterion, during all the simulations realized, no receiver was unable to retrieve the information. In consequence the mechanism can be considered as very efficient.

The second criterion characterize the minimum search time before the information is successfully retrieved. For the reasons evoked previously, results are less satisfying with small groups. Figure 9: represents the values obtained with a small group ($R=600$) and for different values of v . According to the results when v increases *i*) the number of peers who do not find the data after two hops decreases, and *ii*) the maximum number of hops required decreases. This is merely due to the fact that the number of peers contacted after two hops increases when v increases.

Finally to characterize the overall efficiency of the mechanism, we considered the maximum of the minimum number of hops required to retrieve the information (i.e. the maximum of the curves represented in the Figure 9:). The Figure 10 shows the evolution of the maxima obtained as a function of the group size. As expected for small groups, the maximum decreases as v increases. On the other hand, the trend is inverted for large groups. Nevertheless, the influence of v is really important for small group, and quite limited for large groups. As a consequence the value of v should be chosen as high as possible. According to the results presented in this paper, $v=10$ is a satisfying value because the mechanism is robust and efficient at low cost.

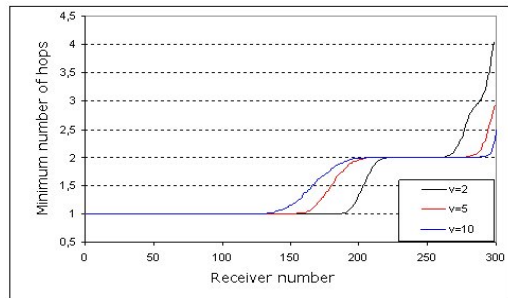


Figure 9: Minimum number of hops for each receivers. $R=600$. $\text{MinRCV}=300$.

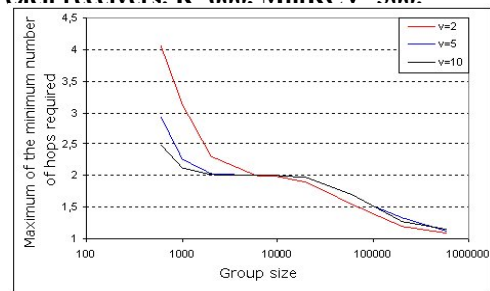


Figure 10: Minimum number of hops for each receivers. $R=600$, $\text{MinRCV}=300$.

V. Conclusion

According to the proposed hybrid approach, only a small part of the initial receivers group has to retrieve data using the terrestrial network. Moreover, most of the receivers' just need a few part of the FEC encoded information. In this context we propose a mechanism which takes advantage of the satellite transmission to establish a logical network composed of T v -ary trees. In particular, this mechanism uses timers to avoid the saturation of the terrestrial network when large groups are involved in the establishment of the network. Then, peers can search for data contacting their neighbors in the logical structure. The mechanisms are completely defined, configured using theoretical results, and evaluated through simulations. According to the simulation results, the mechanisms allow to efficiently find data in the network at very low cost.

References

- ¹ S. Deering: Host Extensions for IP Multicasting. Request for Comments RFC 1112, 1989
- ² C. Diot, B. Neil Levine and Bryan Lyles and Hassan Kassem and Doug Balensiefen: Deployment issues for the IP multicast service and architecture. IEEE Network, vol.14, N. 1, p.78-88, 2000
- ³ J. Nonnenmacher, E. Biersack, D. Towsley: Parity-based loss recovery for reliable multicast transmission. IEEE ACM Transactions on Networking, vol.6, p.349-361, 1998
- ⁴ F. de Belleville, L. Dairaine, C. Fraboul, and J.Y. Tourneret. Group size estimation for hybrid satellite/terrestrial reliable multicast. In IFIP World Computer Congress - Broadband Satellite Communication Systems Workshop, 2004
- ⁵ DVB comme support d'IP multiCAST par satellite. RNRT project No. 67. <http://www.telecom.gouv.fr/rnrt>
- ⁶ M. Jung, J. Nonnenmacher, E. Biersack: Uni-directional versus Bi-directional Communication. Kommunikation in Verteilten Systemen, p. 264-275, 1999
- ⁷ J. Nonnenmacher, E. Biersack: Optimal Multicast Feedback. Proceedings of IEEE Infocom, p. 964-971, 1998
- ⁸ M.W. Koyabe, G. Fairhurst: Reliable Multiast by Satellite: A Comparison Survey and Taxonomy. International Journal of Satellite Communications, Vol: 24(1), p. 21-26, 2001
- ⁹ S. Alouf, E. Altman, P. Nain: Optimal on-line estimation of the size of a dynamic multicast group. Proceeding of IEEE INFOCOM'02, New York, USA, 2002
- ¹⁰ V. Paxson: End-to-end Internet packet dynamics. In Proc. ACM SIGCOMM, pp. 139--152, 1997

- ¹¹ M. Yajnik, J. Kurose, D. Towsley: Packet loss correlation in Mbone multicast network. In Proc. IEEE Global Internet Conference, Part of GLOBECOM'96, 1996
- ¹² C. John, Chuang, A. Marvin, Sirbu: Pricing Multicast Communication: A Cost-Based Approach. Telecommunication Systems, Vol. 17, p. 281-297, 2001.
- ¹³ J. Pinder, L. Ippolito, S. Horan: Four years of Experimental Results from the New Mexico ACTS Propagation Terminal at 20.185 and 27.505 GHz. IEEE on Selected Areas in Communication, Vol.17, N° 2, p.153-162, 1999
- ¹⁴ R. Sherwood, R. Braud, B. Bhattacharjee. Slurpie: A cooperative bulk data transfer protocol. In IEEE INFOCOM'04, 2004.
- ¹⁵ I. Mikel, G. Urvoy-Keller, E. Biersack, P. Felber.,A. Al Hamra, L. Garces-Erice. Dissecting BitTorrent : five months in a torrent's lifetime. In PAM'04, 5th annual Passive & Active Measurement Workshop, 2004.
- ¹⁶ S. Ratnasamy, P. Francis, M. Handley, R. Karp, S. Shenker. A scalable content-addressable network. In Proceedings of ACM SIGCOMM, San Diego, USA, 2001
- ¹⁷ I. Stoica, R. Morris, D. Karger, M. Kaashoek, K. Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications, pages 149–160. ACM Press, 2001.
- ¹⁸ G. Cao, Y. Wu. Reliable multicast via satellites. In International Conference on Information Technology : Coding and Computing (ITCC '01), 2001.
- ¹⁹ S. Rhea, C. Wells, P. Eaton, D. Geels, B. Zhao, H. Weatherspoon, J. Kubiatowicz. Maintenance-free global data storage. IEEE Internet Computing, 2001
- ²⁰ L. Lancérica, L. Dairaine, J. Lacan. Evaluation of content-access QoS for various dissemination strategies in peer to peer networks. In proceedings 11th IEEE International Conference on Networks ICON, 2003
- ²¹ F. Cuenca-Acuna, R. Martin, and T. Nguyen. PlanetP : Using gossiping and random replication to support reliable peer-to-peer content search and retrieval. Technical Report DCS-TR-494, Department of Computer Science, Rutgers University, 2002
- ²² A. Costello, S. McCanne. Search party: Using randomcast for reliable multicast with local recovery. In IEEE INFOCOM 99, pages 1256–1264, 1999
- ²³ J. Nonnenmacher, M. Lacher, M. Jung, E. W. Biersack, and Georg Carle. How bad is reliable multicast without local recovery, In IEEE INFOCOM'98, 1998
- ²⁴ F. de Belleville, L. Dairaine, C. Fraboul, J. Lacan, Une Approche Hybride Satellite/Terrestre pour le transport fiable multipoint à grande échelle, in proceedings of CFIP 2003 (Colloque Francophone sur l'Ingénierie des Protocoles), Paris, 2003
- ²⁵ F. de Belleville, Transport multipoint fiable à très grande échelle : intégration de critères de coût en environnement Internet hybride satellite / terrestre. PhD thesis. Institut National Polytechnique de Toulouse, December 2004