

GROUP SIZE ESTIMATION FOR HYBRID SATELLITE/TERRESTRIAL RELIABLE MULTICAST

F. de Belleville^{1,2}, L. Dairaine^{1,2,4}, C. Fraboul^{1,3}, J.Y. Tourneret^{1,3}

¹*TéSA : Laboratoire coopératif en Télécommunication Spatiales et Aéronautiques
2, rue Charles Camichel BP 7122 - F-31071 Toulouse*

tourneret@tesa.prd.fr

²*ENSICA : Ecole Nationale Supérieure d'Ingénieurs en Constructions Aéronautiques
1, Place Emile Blouin, 31056 Toulouse, Cedex 5, France*

{Florestan.de-Belleville, Laurent.Dairaine}@ensica.fr

³*ENSEEIH /IRIT : 2, rue Charles Camichel, BP 7122 - F 31071 Toulouse*

Christian.Fraboul@enseeih.fr

⁴*LAAS-CNRS : 7 avenue du Colonel Roche, 31077 Toulouse Cedex 4*

Abstract This paper addresses the problem of group size estimation for hybrid satellite/terrestrial multipoint communications. Estimators based on the maximum likelihood principle are investigated. These estimators assume that a *Nack suppression* mechanism is implemented at transport layer. The performance of these estimators is studied theoretically and via simulations. The integration of an appropriate group size estimator in a transport mechanism is finally considered.

Keywords: Multicast , Reliable Transport, Satellite, Large Group, Estimation

1. INTRODUCTION

The support for multipoint communications is an interesting extension of the services proposed in nowadays Internet. However, availability of IP Multicast over terrestrial networks is still limited because of technical and economical issues. A more practical solution to deploy the IP Multicast service may consist of using a geostationary satellite. Such systems have indeed a large coverage zone. Moreover, packets are naturally broadcast over this zone. Thus, using a satellite system allows a source to reach a huge number of receivers with only one hop. Satellite communications seem too expensive at first sight. Nevertheless, in the context of communications toward large groups, per-receiver cost

decreases when the group size increases. For very large groups, it becomes even less than when a terrestrial network is used [?]. For this reason, this study focuses on large scale reliable multipoint communications. Moreover, satellite links exhibit long transmission delays. Therefore we consider applications without time constraints, such as software updates.

Satellite communications are prone to transmission errors (due to atmospheric perturbations). These errors induce packet losses at some receivers. During a satellite transmission, these receivers have to wait for more information to recover these losses. The number of such receivers will decrease all along the transmission. Since the size of Multicast groups decreases, the cost per receiver (for the satellite communication) grows. Consequently, the proposed approach consists of transmitting data via satellite only when the number of receivers is sufficiently high and via terrestrial transmissions otherwise [?].

One major need for this approach is a mechanism ensuring group size estimation. Several group size estimators have already been studied in the past. However, their efficiency for hybrid satellite / terrestrial multipoint communications has not been studied, to our knowledge. This paper focuses on issues related to group size estimation in this specific context. It is structured as follows: the problem is formulated in section 2, different group size estimators are investigated in section 3, section 4 studies an association between *Nack suppression* and group size estimation, conclusions are reported in section 5.

2. PROBLEM FORMULATION

2.1 Communication System

This paper considers a hybrid satellite/terrestrial network. End-users are assumed to be connected to the terrestrial network via a High Speed Access Network (HSAN) like xDSL or LAN. They are also connected to the satellite system via a HSAN, or directly with a Very Small Aperture Terminal (VSAT). The satellite system uses a geostationary satellite. It provides a best effort IP Multicast service. This system, considered in many research projects such as DIPCAST [?] or GEOCAST [?], is technically and economically realistic.

Moreover, the following assumptions about the application layer are made:

- The Multicast session is announced beforehand. This announce is realized using either out-of-band means (e.g. e-mail), or session management tools such as the Session Description Protocol (SDP) [?].
- No receiver can join the session after its beginning.

2.2 A Hybrid Satellite/Terrestrial Approach

The hybrid satellite / terrestrial approach consists of optimizing the overall communication cost for full reliable large scale multipoint communications

(i.e. including several thousands of receivers, or even more). In order to achieve this goal, an efficient technique consists of estimating the minimum number of receivers ensuring a profitable satellite broadcasting (with respect to the terrestrial network) [?]. This minimum number, denoted as R_{min} , can be determined by using appropriate cost functions (reflecting economical costs, network load, etc.).

The behavior of a typical session can be summarized as follows:

- the satellite transmission starts at a predefined time,
- the source periodically estimates the session size during this transmission. The session size is defined as the number of receivers which have not received the complete information. Thus, once all initial information has been transmitted, the session size is likely to decrease (all receivers which have not experienced losses leave the session),
- the source transmits FEC (Forward Error Correction) encoded packets to repair losses. This technique, referred to as *ARQ type II*, improves scalability (see [?] for more details),
- the satellite transmission stops when the estimated size goes below the threshold R_{min} ,
- the remaining receivers (which do not have all information) contact the others (by using e.g. a peer-to-peer block downloading service) to retrieve missing data,
- the session stops when all receivers have received the full information.

This approach assumes that an effective mechanism is available to estimate group size. Moreover, transport protocol mechanisms have to scale very well, as the considered applications are designed for transmissions toward very large groups.

3. GROUP SIZE ESTIMATION

3.1 System Characteristics

As numerous receivers have to inform the source about data they did not receive, a *Nack suppression* mechanism (see e.g. [?]) has to be implemented. This mechanism limits the number of responses generated from the group (to avoid *Nack implosion* and improve scalability). This paper considers a timer-based mechanism, where receivers pick independent and random timeout periods. Timeout periods are generated according to an exponential distribution whose probability density function (pdf) is

$$f(x) = \begin{cases} \frac{1}{e^{\lambda-1}} \frac{\lambda}{T} e^{-\frac{\lambda}{T}x} & , \quad 0 \leq x \leq T \\ 0 & , \quad \text{otherwise.} \end{cases}$$

The determination of parameters λ and T is discussed in section 4. The receiver which picks to the lowest timeout sends a message to the source. The source then sends a *stop* message to the group. All the receivers reset their timers, upon reception of this message. Consequently, the source only receives the responses between the first message and the reception of the *stop* message. These responses correspond to timers picked in $[x_1^*, x_1^* + c]$, where x_1^* is the lowest timeout and c is the round trip time. Friedman and Towsley pointed out that heterogeneity between the transmission delays of the receivers has an impact on the feedback suppression mechanism [?]. We assume that the correction they proposed in [?] is implemented so that this heterogeneity does not affect the estimation.

Satellite communications are subjected to burst losses. As a consequence, some receivers (experiencing losses) cannot receive messages from the source for long periods (on average several seconds). As a result, the Nack suppression mechanism has to be based on an explicit approach. In this case, the source periodically sends a *pool* message to the group of receivers. Note that this message will reach only the visible receivers (i.e. which do not experience losses). Each receiver generates a timeout after receiving the *pool* message. The explicit approach contrasts with an implicit one for which all receivers periodically pick timeouts. In the implicit case, one cannot prevent non line-of-sight receivers to send response messages. This may result in a message implosion.

Receivers which belong to the multicast session are called *active receiver* when they are able to receive source messages, and *inactive receivers* when they cannot receive these messages.

Two steps have finally to be distinguished during the transmission:

- All original information is broadcast in the course of the first step. The session size can not vary during this step, since late joining is not supported (though some receivers may be unreachable).
- Redundancy packets are sent to repair losses, in the course of the second step. The session size will decrease during this step, as more and more receivers gradually obtain all information.

3.2 Objectives of the Estimation Mechanism

The main objective of group size estimation is to track session size in order to stop satellite transmission when the number of receivers which belong to the session becomes too small. To achieve this goal, it is important to estimate the number of active receivers at each instant. The main contributions of this paper are summarized below:

- the Broadcast audience estimation procedure studied in [?] is adapted to satellite/terrestrial multicast. In particular, it is important to quantify the estimation performance obtained with real observations,
- The theoretical study conducted in [?] is further investigated. In particular, a new unbiased estimator is derived.
- A mechanism allowing session size evaluation is proposed.

3.3 Estimation Methods

An experience is a set of random timeout periods caused by a *pool* message. For the i th experience, denote as $(x_{1,i}, x_{2,i}, \dots, x_{N,i})$ the values generated by the N receivers. Denote as $(x_{1,i}^*, x_{2,i}^*, \dots, x_{N,i}^*)$ the associated order statistics such that:

$$x_{1,i}^* = \min_{j=1,\dots,N} x_{j,i} \leq x_{2,i}^* \leq \dots \leq x_{N,i}^* = \max_{j=1,\dots,N} x_{j,i}.$$

This paper addresses the problem of estimating the number of active receivers N from k experiences in two cases:

- **Case 1:** only the first arrivals $\{x_{1,i}^*, i = 1, \dots, k\}$ are considered by the receiver,
- **Case 2:** all responses belonging to $[x_{1,i}^*, x_{1,i}^* + c]$ are considered by the receiver.

3.3.1 Maximum Likelihood Estimator (MLE) from the First Arrivals.

Consider an absolutely continuous random variable X with probability density function $f(x)$ and cumulative distribution function $F(x)$. It is well known that the probability density function of the sequence of minima $(x_{1,i}^*)_{1 \leq i \leq k}$ can be expressed as (see [?]for more details):

$$L(\mathbf{x}_1^* | N) = \prod_{i=1}^k N f(x_{1,i}^*) [1 - F(x_{1,i}^*)]^{N-1},$$

where $\mathbf{x}_1^* = (x_{1,1}^*, \dots, x_{1,k}^*)$. The MLE of N (based on the observed vector \mathbf{x}_1^*) can be obtained by maximizing the likelihood $L(\mathbf{x}_1^* | N)$ with respect to $N \in \mathbb{N}$. Straightforward computations show that the maximum of $L(\mathbf{x}_1^* | x)$ is obtained for

$$x = \hat{N}_k = \frac{k}{-\sum_{i=1}^k \ln(1 - F(x_{1,i}^*))}. \quad (1)$$

However, \hat{N}_k is not the MLE of N since its value does not generally belong to \mathbb{N} . The MLE of N is defined by:

$$\hat{N}_{ML} = \arg \max \left\{ L \left(X_1^* | \text{ip}(\hat{N}_k) \right), L \left(X_1^* | \text{ip}(\hat{N}_k) + 1 \right) \right\}, \quad (2)$$

where $\text{ip}(x)$ is the integer part of x . Note that the properties of the estimator \hat{N}_{ML} (in terms of bias and variance) are difficult to study because of the $\arg \max$ operation.

The authors of [?] propose to estimate N by using the estimator \hat{N}_k , for broadcast audience estimation. However, it is easy to show that this estimator is biased. Indeed, $Z_i = -\ln(1 - F(X_{1,i}^*))$ is distributed according to an exponential distribution. By assuming that the k experiences are independent, $(2N) \sum_{i=1}^k Z_i$ is distributed according to a chi-square distribution with $2k$ degrees of freedom. Consequently, the mean of \hat{N}_k expresses as $E(\hat{N}_k) = \frac{k}{k-1}N$ (for $k > 1$). Based on this comment, we propose the following unbiased estimator:

$$\hat{N}_k^* = \frac{k-1}{-\sum_{i=1}^k \ln(1 - F(X_{1,i}^*))}, \quad (3)$$

whose variance can be computed (for $k > 2$): $\text{var}(\hat{N}_k^*) = \frac{N^2}{k-2}$.

To summarize, this section has presented two group size estimators \hat{N}_{ML} and \hat{N}_k^* . The first estimator \hat{N}_{ML} is based on the maximum likelihood principle. Unfortunately, its properties are difficult to study. The second estimator \hat{N}_k^* can be seen as an approximation of the MLE. The performance of this estimator (in terms of bias and variance) has been studied analytically.

3.3.2 MLE from Observations in $[X_{1,i}^*, X_{1,i}^* + c]$.

Denote as R_i the number of observations $x_{j,i}^*$ belonging to $[x_{1,i}^*, x_{1,i}^* + c]$. As explained in [?], the vector $(X_{1,i}^*, R_i)_{1 \leq i \leq k}$ is a sufficient statistic for the estimation of N . This paper derives the ML estimator of N based on the observation of $(X_{1,i}^*, R_i)$, $i = 1, \dots, k$. The joint distribution of $(X_{1,i}^*, R_i)$ is the set of densities

$$\begin{aligned} f_{r_i}(x_{1,i}^*) &= N \binom{r_i-1}{N-1} f(x_{1,i}^*) [F(x_{1,i}^* + c) - F(x_{1,i}^*)]^{r_i-1} \\ &\quad \times [1 - F(x_{1,i}^* + c)]^{N-r_i} \\ &\propto \frac{\Gamma(N+1)}{\Gamma(N-r_i+1)} [1 - F(x_{1,i}^* + c)]^{N-r_i}, \end{aligned}$$

where $\Gamma(\cdot)$ is the Gamma function and $r_i \in \{1, \dots, N\}$ is the number of observations in $[x_{1,i}^*, x_{1,i}^* + c]$ ($f_p(x_{1,i}^*) dx_{1,i}^*$ is the probability of having $N_{r_i} = p$ and $X_{1,i}^* \in [x_{1,i}^*, x_{1,i}^* + dx_{1,i}^*]$). Assuming the different experiences are independent, the MLE of N from the observation of $(X_{1,i}^*, R_i)$, $i = 1, \dots, k$ can be obtained by maximizing the following criterion with respect to N :

$$\sum_{i=1}^k \ln \left[\frac{\Gamma(N+1)}{\Gamma(N-r_i+1)} \right] + (N-r_i) \sum_{i=1}^k \ln[1 - F(x_{1,i}^* + c)]. \quad (4)$$

This maximization can be computationally expensive in practical applications. In order to avoid the maximization procedure, a Poisson approximation was proposed in [?]. This approximation leads to the following estimator

$$\tilde{N}_k = \frac{k + \sum_{i=1}^k (R_i - 1)}{-\sum_{i=1}^k \ln(1 - F(X_{1,i}^*)) + \sum_{i=1}^k p_i}, \quad (5)$$

with

$$p_i = \begin{cases} \frac{F(X_{1,i}^*+c) - F(X_{1,i}^*)}{1 - F(X_{1,i}^*)} & \text{if } X_{1,i}^* + c < T \\ 1 & \text{if } X_{1,i}^* + c \geq T \end{cases}$$

and $[0, T]$ is the support of the probability density function $f(x_{1,i}^*)$. Note that (5) reduces to (1) when $p_i = 0$ and $R_i = 1$. This particular case corresponds to the situation where only the first arrivals are considered.

3.3.3 Simulation Results.

Many simulations have been conducted to validate the previous theoretical results. Figure 1-a shows the mean of the different group size estimates, obtained from 20 Monte Carlo runs (the parameters are $N = 50.000$ and $k = 5$). The estimator based on all responses is clearly more efficient than those based only on the first arrivals. This result is confirmed in figure 1-b which shows the variances of the different estimates as a function of k (computed from 2,000 Monte Carlo runs). Note that the maximum values of k is $k_{max} = 10$. This value corresponds to the maximum number of *pool* messages sent before each group size estimation. High values of k would imply long delays before each estimation, or a high sending rate of *pool* messages. Thus, high values of k are prohibited since they cannot be tolerated by the upper protocol.

a. Mean of group size estimates (N = 50,000, k = 5) b. Variance of group size estimates versus k (N = 50,000)

Figure 1. Mean and variance of Group Size Estimates.

4. NACK SUPPRESSION AND GROUP SIZE ESTIMATION MECHANISM

The problem of group size estimation is associated with the problem of *Nack implosion* as explained previously. In particular, the parameters λ and T of the pdf $f(x)$ have an impact on the number of generated messages. Consequently, two questions arise: 1) Is there a risk of *Nack Implosion*?, and 2) How is the estimation accuracy related to the parameters λ and T ? Many simulations have been conducted in order to analyze the impact of λ and T on group size estimators. This paper focuses on the estimator \tilde{N}_k , since it outperforms the estimators based on the first arrivals.

This paper considers two kinds of simulations: Parameter λ and T are **static** *in the first set of experiments* (i.e. the parameters do not change with time). In this case, the Nack Suppression mechanism is configured so that only few messages are generated when $N = N_{max}$ (N_{max} is the expected maximum group size, $N_{max} = 10^6$ in our simulations). The parameters of $f(x)$ can then be computed as follows (see [?] for more details):

$$\begin{cases} \lambda = 1.1 \ln(N) + 0.8, \\ T = \lambda c \left[\ln \left(R + \frac{N}{e^\lambda - 1} \right) - \ln \left(1 + \frac{N}{e^\lambda - 1} \right) \right]^{-1}, \end{cases} \quad (6)$$

where $N = N_{max}$ and R is the expected number of responses ($R = 30$ in the simulations). *In the second set of simulations*, parameters λ and T are **updated**. After each estimation, the parameters of $f(x)$ are computed from the estimated group size (i.e. by using equation (6) with $N = \tilde{N}_k$).

Figure 2-a shows the variance of the estimated group size (computed from 2,000 Monte Carlo runs) as a function of k . Updating the parameters of $f(x)$ has clearly a beneficial impact on the estimation of N , since the variance of the estimates decreases.

- a. Variance of \tilde{N}_k versus k ($R = 30$, $N_{max} = 10^6$, $N = 50,000$). b. Maximum Number of Responses versus N_{min} .

Figure 2. Impact of the adaptation and N_{min}

Note that the source has to send a *pool* message for each estimation (according to section 3.1). Consequently, it is not problematic to send the estimated group size, as it can be included in the *pool* message. It is also important to note that adjusting parameters λ and T can induce a risk of *Nack Implosion*. Indeed, packet losses are due to atmospheric perturbations (as explained in section 1). These perturbations affect geographical zones. Thus, numerous receivers can simultaneously suffer from a loss of connection (for example when a perturbation is above a large city). In this situation, the estimation mecha-

nism will underestimate the group size. This may lead to the generation of a huge number of packets when the non line-of-sight receivers are visible again.

To avoid this *Nack Implosion*, it may be preferable to set a minimum value N_{min} for the estimation of (λ, T) such that:

- when $\tilde{N}_k > N_{min}$, λ and T are updated according to (6) with $N = \tilde{N}_k$,
- when $\tilde{N}_k \leq N_{min}$, λ and T are updated according to (6) with $N = N_{min}$.

The determination of N_{min} is an important problem. A small value of N_{min} induces a large risk of *Nack Implosion*. Conversely, the benefit of the adaptation is reduced when the value of N_{min} is too high. Many simulations have been conducted in order to determine an appropriate value of N_{min} . More precisely, the mechanism is configured with different thresholds. For each simulation, the group size varies instantaneously from $N = 200$ (i.e. a value below the threshold) to $N = 10^6$. The maximum number of generated messages for each value of the threshold N_{min} is depicted in figure 2-b (each maximum has been computed from 20 Monte Carlo runs). These results allow to control the number of generated responses. For instance, $N_{min} = 6000$ ensures that the number of responses is less than 1000. Note that the thresholding operation is necessary to avoid *Nack Implosion*. However, this operation is not often used since groups are expected to be really large.

The group size estimation procedure has been evaluated in several real cases of satellite transmissions. Figure 3-a shows group size estimates obtained during a particular experiment where (λ, T) has been **updated** every $k = 5$ time units. The quality of the group size estimates is very similar in presence or

a. Group size estimates b. Number of generated messages

Figure 3. Group Size Estimates and Number of messages generated during a satellite transmission ($k = 5$, $N_{min} = 6000$).

absence of threshold ($N_{min} = 6000$ when thresholding). Figure 3-b shows the number of responses obtained during the previous transmission. In absence of thresholding, several hundred messages are generated when the number of visible receivers suddenly increases. This situation does not occur when the threshold has been set to $N_{min} = 6,000$. The use of a threshold N_{min} appears to be a good solution to prevent the *Nack Implosion* problem, since it has no impact on the quality of estimation.

5. CONCLUSION

This paper addressed the issue of group size estimation for hybrid satellite/terrestrial multipoint communications. Several estimators based on the maximum likelihood principle were investigated. These estimators assumed

that a *Nack suppression* mechanism was implemented. The estimators constructed from several responses outperform those based on the first arrivals. Its integration in a mechanism of group size estimation was finally studied.

The main objective of the proposed estimation scheme was to detect situations when group size is lower than a predefined threshold ($N \leq R_{min} \approx 300$ receivers). A performance analysis showed that the estimator \tilde{N}_k is a good candidate to achieve this goal (the mean square error of \tilde{N}_k was about 10%).