Modeling Dynamic Satellite Bandwidth Allocation for Situation-Awareness Applications

Laura Galluccio, Alessandro Leonardi, Giacomo Morabito, Sergio Palazzo, and Corrado Rametta

Dipartimento di Ingegneria Elettrica, Elettronica ed Informatica, University of Catania, Italy
{lgalluccio,aleonardi,gmorabi,spalazzo,crametta}@dieei.unict.it

Abstract. Dynamic allocation is a crucial issue when coping with satellite-provided services where many demanding users share a limited bandwidth. This is the case of situation-awareness applications, for example, where several users generate information (possibly multimedia) that must be delivered to a control center through a shared satellite link. In this context, bandwidth allocation should be adaptive so as to support differentiation in the treatment offered by the communication system to distinct data flows, depending on the specific criticality of the individual flows. In this paper we provide an analytical framework for modeling the performance of different approaches which can be adopted to face such a problem. As compared to existing literature in the field, the original contribution of this work is providing a general mathematical framework to support dynamic bandwidth allocation as a function of the criticality of the information being managed.

Keywords: Satellite, Dynamic Bandwidth Allocation, Criticality.

1 Introduction

Satellite services give coverage to remote and scarcely connected areas. This is a very important contribution to foster situation-aware and military operations, as well as to connect faraway villages in emerging countries. However, typically there are only a few access points to the satellite network, denoted in the following as *gateways*. If many users want to access the Internet, they should thus share the same set of gateways. So bandwidth sharing mechanisms are needed. The problem drastically worsens when users request for support of highly demanding real-time applications, such as video or audio, since the satellite bandwidth is costly and should be allocated according to the different traffic priority, i.e. the data criticality. Accordingly, highly dynamic and adaptive bandwidth allocation schemes should be designed. Traditional architectures for services differentiation like IntServ and DiffServ [8] are not sufficient since neither dynamic bandwidth reallocation is supported nor a feedback exists, so that source coding parameters can be tuned run-time. In this paper, differently from previous literature in the field, we mainly focus on mathematical modeling of a satellite service access

G. Giambene and C. Sacchi (Eds.): PSATS 2011, LNICST 71, pp. 19–32, 2011.

system where multiple sources are grouped in clusters. Sources are characterized through their emission process and the criticality of the information they carry.

Numerous models of traffic sources have emerged in the last years, both multimedia and not. Also, some of these models assume satellite communication channels [4,5,11] as likely in situation-awareness operations. However, none of these schemes considers dynamic bandwidth allocation when related to multiple sources transmitting on the same channel, and the priority that traffic could have due to the criticality of the area where it was generated. Similarly to our work, in [9] is proposed a scheduler aimed at sharing the limited uplink resources of a satellite system. In particular, the discussed solution refers to the case where many bursty users emit traffic with different QoS requirements, and different data rates are provided to each terminal to differentiate the QoS priority levels. The approach provides fairness and maximizes the exploitation of the capacity of the system, also in case of time-varying channels.

In this paper we provide a mathematical model to compare the performance of different schemes for situation-awareness operations. More specifically, we present a system where multiple clusters of non gateway devices, not connected to the Internet, are deployed and want to send information through a cluster gateway equipment. All the cluster gateways share a satellite link towards the headquarter of the operations. We assume that the traffic issued by each of these gateways could be characterized by a priority associated to the criticality of the area where it was generated. To this purpose, an intelligent mechanism for the management of the transmission at the gateway output should be devised. This obviously depends on the specific scheme being employed.

Specifically, we develop a framework of the overall system and we use it to test different schemes for situation-awareness operations. We provide this comparison by exploring parameters to support traffic differentiation at the queue in terms of allocated bandwidth, adaptive queue management and adaptive source coding as a function of the criticality level of both the cluster area and the overall system. However, supporting adaptive protocols as a function of the criticality level of the different areas in the system could be complex and costly. In fact, as a consequence of the adaptability of the system, the complexity of the procedures could also imply an increase in the delivery time. This could become a detrimental aspect in scenarios where responsiveness is of primary importance. So, the main contribution of this work is to provide network designers with a tool to design an efficient system by also taking into account cost constraints, reliability and effectiveness issues.

2 Overview

We consider that sources are distributed in a geographic area. Sources located nearby and served by the same gateway are grouped into a cluster C_i. Each subarea where a cluster is located and, thus, each cluster, are characterized by a certain criticality level as a consequence of a given event occurring in the related area, e.g. a earthquake, a mine explosion, etc. Also, some sources in each area could become particularly critical. The identification of this criticality can lead to different approaches:

- *Plain approach* (PA): in this case the behavior of the traffic sources, as well as the bandwidth share assigned to each gateway, do not depend on the criticality levels. At the gateways, all packets are treated in the same way, regardless of the criticality of the information they carry.
- *Packet differentiation at the GW approach* (PDA): differently from the PA case, packets treatment at the gateway depends on the criticality of the information they carry. We assume that a D-RED buffer management scheme is utilized [1], i.e., the gateway applies a RED buffer management scheme in which the characteristic parameters depend on the criticality of each source. In this way, packets characterized by low criticality level are discarded with higher probability than packets with high criticality level.
- *Aggregate criticality-aware approach* (ACA): in this case the bandwidth share that will be used by each gateway depends on the criticality of the area where the cluster is located and not on the individual criticality levels of the sources. Buffer management at the gateway will be the same as in the PDA case.

3 Model

In this section we will model the system status. Accordingly, we should consider both the cluster to which the source node under consideration belongs and the remaining clusters. Moreover, the criticality of the area where the entire system is located will impact on the criticality of the specific sub-area where the source node under consideration belongs. This will also influence its cluster criticality level. However, observe that the time scale of criticality variations is significantly larger than the one associated to packet transmission or buffer queueing. Accordingly, in the next sections, we will consider the number of clusters with high criticality level as assigned and we will perform an analysis of the system status focused only on a single cluster. Then, final results will be obtained by weighting the results provided in this case with the probability to have a given number of critical clusters $NC_{C_1} = l$ over the NC^{max} clusters in the network.

In the following sections we will model the various components of the system. More specifically, we consider the following states:

- $S^{(Q)}(t)$: state of the server process at the gateway node;
- $S^{(TS.E)}(t)$: state of the emission process at a given source (in the following identified as tagged source, TS);
- $S^{(TS.CL)}(t)$: state of the criticality level process at a given source;
- $S^{(C_i-TS.E)}(t)$: state of the emission process at sources in the cluster C_i, other than the given source;
- $S^{(C_i-TS.CL)}(t)$: state of the process representing the number of critical sources in cluster C_i other than the given source;
- $S^{(C_i.ACL)}(t)$: state of the process representing the criticality level of the area the cluster belongs to;

3.1 Process Representing the Criticality Level of the Area

As a consequence of the occurrence of an event, e.g. the explosion of a mine or the invasion of a region, a certain area where various sources are located can become critical. The criticality of an area where a cluster of nodes is located can be represented as a two-states Markov chain. More specifically we identify as $S^{(C_i.ACL)}(t)$ the state of this process. Its state space $\Im^{(C_i.ACL)}$ consists of only two possible states: not critical (0) and critical (1). Also, the transition state matrix can be identified as $Z^{(C_i.ACL)}$ and the stationary state probabilities can be obtained by solving the set of equations as in [10]

$$\begin{cases} \mathbf{\Pi}^{(C_i.ACL)}\mathbf{Z}^{(C_i.ACL)} = \mathbf{0} \\ \sum_{j\in\Im^{(C_i.ACL)}} \Pi_j^{(C_i.ACL)} = 1 \end{cases} \tag{1}$$

where $\mathbf{0}$ is an array of 2 elements all equal to zero. The characterization of the criticality level of the area where the cluster C_i is located is of paramount importance. In fact, in the following, we will see how the emission rates at the sources are conditioned and impacted by this criticality.

3.2 Process Characterizing the TS

We now focus on a generic source in a cluster C_i denoted as *tagged source* (TS). There is a large literature demonstrating that a traffic source can be modeled accurately in the time by means of *Markov Modulated Poisson Processes* (MMPP). According to such models, the process characterizing the behavior of a traffic source depends on the current state of an underlying Markov chain.

The state space of this process can be written by considering that $S^{(TS)}(t) = (S^{(TS.E)}(t), S^{(TS.CL)}(t))$ where $S^{(TS.E)}(t)$ and $S^{(TS.CL)}(t)$ have been introduced above and represent the state of the emission process at the tagged source and the state of the criticality level process at the TS, respectively. Accordingly, the state space $\Im^{(TS)}$ can be expressed as $\Im^{(TS)} = \Im^{(TS.E)} \times \Im^{(TS.CL)}$ where $\Im^{(TS.E)}$ and $\Im^{(TS.CL)}$ represent the state spaces of the two underlying processes. For worth of simplicity in the following we will assume that, again, there are only two states in the process characterizing the criticality of the TS, i.e. $\Im^{(TS.CL)} = \{0, 1\}$. The behavior of the TS depends on the criticality level of the information generated by the TS with respect to the number of sources in the same cluster C_i and the criticality level of the area where nodes of the same cluster C_i are located. Accordingly, if we denote as N_{C_i} the number of traffic sources in the cluster C_i, then the emission rate $\lambda^{(TS)}(t)$ of the TS source will be

$$\lambda^{(TS)}(t) = f\left(N_{C_i}, S^{(TS.CL)}(t), S^{(C_i.ACL)}(t), S^{(TS.E)}(t)\right) \tag{2}$$

where $f(\cdot)$ is a function which can be rewritten as the product of two terms:

- $f_{\text{State}}\left(S^{(TS.E)}(t)\right)$ which is a term that depends only on the current state of the tagged source emission process. This is the emission rate that would be used by TS in case there is no adaptation to the current system condition, i.e. no consideration of the criticality level.

– $f_{\text{TS_Share}}\left(N_{C_i}, S^{(TS.CL)}(t), S^{(C_i.ACL)}(t)\right)$which is a term that accounts for the share of bandwidth allowed to the TS source. This function takes into account both the criticality level of the TS when compared to the criticality level of the cluster area and the number of nodes in the cluster C_i. Note that, in case the emission process of TS does not change its behavior depending on the criticality level, then $f_{\text{TS\_Share}}\left(N_{C_i}, S^{(TS.CL)}(t), S^{(C_i.ACL)}(t)\right)$ is only a function of the number of nodes in the cluster C_i, i.e. $f_{\text{TS\_Share}}(N_{C_i})$. On the contrary if the behavior of the TS source depends on the criticality levels, then $f_{\text{TS\_Share}}\left(N_{C_i}, S^{(TS.CL)}(t), S^{(C_i.ACL)}(t)\right)$ is expected to increase as the criticality level of the TS increases with respect to the criticality level of the area where the cluster is located. Similarly, the value of $f_{\text{TS\_Share}}\left(N_{C_i}, S^{(TS.CL)}(t), S^{(C_i.ACL)}(t)\right)$ is expected to increase as the criticality level of the cluster area increases.

Observe that the emission rate values, in general depend on the criticality level of the TS although in some cases it could happen that the emission rate array is independent of it. Also consider that a change in the criticality level of a TS could not result in a variation in the emission rate depending on the state, but only lead to a change in the type of packets being sent (i.e. the packet priority), although the rate could remain the same.

To characterize the emission process at the TS source, we need a transition rate matrix $\mathbf{Z}^{(TS.E)}$ and an array of the stationary state probabilities $\mathbf{\Pi}^{(TS.E)}$ which should satisfy a relationship analogous to the one in eq. (1).

Concerning the TS criticality level, it could be represented using a two-states Markov chain. In order to characterize this Markov chain, we use a state transition rate matrix $\mathbf{Z}^{(TS.CL)}$ which generic element $z_{h,j}^{(TS.CL)}$ represents the rate of transition of the criticality state at each TS from state h to j. Let us note that the rate of transitions from a state to another will depend on the criticality of the area the cluster belongs to. More specifically,

$$z_{h,j}^{(TS.CL)} = \begin{cases} \alpha_h^{(TS.CL)} & \text{if } j = h+1 \\ \beta_h^{(TS.CL)} & \text{if } j = h-1 \\ -\sum_{\substack{j \in \Im^{(TS.CL)} \\ j \neq h}} z_{h,j}^{(TS.CL)} & \text{if } h = j \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The terms in the transition rate matrix, $\alpha_h^{(TS.CL)}$ and $\beta_h^{(TS.CL)}$ are related to the process describing the criticality of the area where the cluster is located.

$$\alpha_h^{(TS.CL)} = \begin{cases} \alpha_0^{(TS.CL)} \\ \alpha_1^{(TS.CL)} \end{cases} \qquad \beta_h^{(TS.CL)} = \begin{cases} \beta_0^{(TS.CL)} & \text{if } S^{(C_i.ACL)}(t) = 0 \\ \beta_1^{(TS.CL)} & \text{if } S^{(C_i.ACL)}(t) = 1 \end{cases} \tag{4}$$

To completely characterize this process, we also need the array of the steady state probabilities. The latter, together with the transition rate matrix will satisfy a relationship similar to the one shown in eq. (1). Once the underlying Markov chains have been introduced, the state of the TS source can be characterized through the transition rate matrix. If we identify as $s_1^{(TS.E)}$ and $s_2^{(TS.E)}$ two states of the TS emission process and $s_1^{(TS.CL)}$ and $s_2^{(TS.CL)}$ two states of the TS criticality level process, the terms of the transition rate matrix associated to the process describing the global behavior of the TS, can be written as:

$$z_{h,j}^{(TS)} = \begin{cases} z^{(TS.E)}_{(s_1^{(TS.E)}, s_2^{(TS.E)})} & h \neq j \ , j = s_1^{(TS.CL)} = s_2^{(TS.CL)} \\ z^{(TS.CL)}_{(s_1^{(TS.CL)}, s_2^{(TS.CL)})} & h \neq j \ , j = s_1^{(TS.E)} = s_2^{(TS.E)} \\ -\sum_{\substack{j \in \Im^{(TS)} \\ j \neq h}} z_{h,j}^{(TS)} & \text{with } h = j \\ 0 & \text{in other cases} \end{cases} \tag{5}$$

The steady state probabilities of the TS can be calculated as a solution of the system in eq. (1) with TS instead of $C_i.ACL$. Consider that with such an approach for modeling the TS source, we are able to limit the explosion in the number of states; in fact, with this MMPP modeling of the TS process, we reduce the number of states to $\Im^{(TS)} = \Im^{(TS.E)} \times \Im^{(TS.CL)}$. This represents a primary advantage if compared to traditional literature in the field where ON-OFF sources require only 2 states but do not allow characterization of the sources with high detail [10]. On the contrary, traditional modeling of video and audio sources implies a drastic increase in the number of states as discussed in [7,10].

3.3 Process Representing the Number of Critical Sources in Cluster C_i Other Than the TS

This process, denoted as $S^{(C_i-TS.CL)}(t)$, can be modeled through a birth-death chain with N_{C_i} states where N_{C_i} has been defined as the number of traffic sources in cluster C_i, including the tagged source. Accordingly, the state space of this process is $\Im^{(C_i-TS.CL)} = \{0, 1, \ldots N_{C_i} - 1\}$. This Markov chain could be described through a transition state matrix $\mathbf{Z}^{(C_i-TS.CL)}$ and an array of the stationary state probabilities $\mathbf{\Pi}^{(C_i-TS.CL)}$. The terms in matrix $\mathbf{Z}^{(C_i-TS.CL)}$ can be written as in eq. (3), with $C_i - TS.CL$ instead of $TS.CL$ The transition rate matrix terms, $\alpha_h^{(C_i-TS.CL)}$ and $\beta_h^{(C_i-TS.CL)}$, depend on the process of the criticality level of the area where cluster C_i is located.

$$\alpha_h^{(C_i-TS.CL)} = \begin{cases} \alpha_0 \cdot (N_{C_i} - 1 - h) \\ \alpha_1 \cdot (N_{C_i} - 1 - h) \end{cases} \quad \beta_h^{(C_i-TS.CL)} = \begin{cases} \beta_0 h & \text{if } S^{(C_i.ACL)}(t) = 0 \\ \beta_1 h & \text{if } S^{(C_i.ACL)}(t) = 1 \end{cases} \tag{6}$$

Finally, let us remember that the transition rate matrix and the array of the steady state probabilities should again satisfy a system similar to eq. (1).

3.4 Process of Emission at Sources in Cluster C_i Other Than the TS

Once the emission process at a source TS in the cluster C_i has been identified, we should model the emission process at the other $N_{C_i} - 1$ sources. In order to make the analysis independent of the number of sources multiplexed, we introduce an approximation which consists of focusing only on the TS source and modeling the aggregate of the other sources loading the buffer by means of a single arrival process. In order to minimize the effects of the approximation, we model the aggregate traffic with a MMPP process matching its pdf and autocorrelation function. In fact, it is well known that buffer performance depend on first and second order statistics (i.e. pdf and autocorrelation function) of the source feeding the buffer. So the target is to build an MMPP with specific pdf and autocorrelation function. This is the well known inverse eigenvalues problem [3] which has been solved in the continuous time and discrete time [7] cases. We will refer in the following to the approach for the continuous time case. To this purpose, the SMAQ tool [6] can be employed to obtain a Markov chain by the probability density function and the autocorrelation functions of the process. To apply this strategy, we observe that, when assuming $N_{C_i} - 1$ identical and incorrelated sources, the pdf can be obtained as the convolution of the pdf of the emission processes of the composing sources and the autocorrelation is $N_{C_i} - 1$ times the autocorrelation of the individual emission process.

$$\begin{array}{l} R_{[C_i-TS.E,C_i-TS.E]}(\tau) = (N_{C_i} - 1) \cdot R_{[TS.E,TS.E]}(\tau) \\ Pdf_{[C_i-TS.E,C_i-TS.E]}(\delta) = \otimes_{N_{C_i}-1} Pdf_{[TS.E,TS.E]}(\delta) \end{array} \tag{7}$$

where $\otimes_{N_{C_i}-1}$ denotes $N_{C_i} - 1$ times convolution of the pdf of the emission process at the individual TS.

The state of this process is denoted as $S^{(C_i-TS.E)}(t)$ and the state space is identified as $\Im^{(C_i-TS.E)}$ where $\Im^{(C_i-TS.E)} = \{0, \dots K^{(C_i-TS.E)}\}$ having identified as $K^{(C_i-TS.E)}$ the maximum number of states of this process. The state transition rate matrix, $\mathbf{Z}^{(C_i-TS.E)}$, and the array of the emission rates, $\mathbf{\Lambda}^{(C_i-TS.E)}(t)$, can be identified. Observe that the emission rates array depends on the criticality level of the area where the cluster is located. More specifically,

$$\lambda_j^{(C_i-TS.E)} = \begin{cases} \lambda_0^{(C_i-TS.E)} \text{ if } S^{(C_i.ACL)}(t) = 0 \\ \lambda_1^{(C_i-TS.E)} \text{ if } S^{(C_i.ACL)}(t) = 1 \end{cases} \tag{8}$$

Finally the array of the steady state probabilities can be calculated by considering that the state transition rate matrix and the steady state probability array should satisfy a system similar to the one in eq. (1).

3.5 System Process

Let us now model the cluster system. In particular we are interested in modeling a finite buffer of maximum length q_{max} fed by the traffic generated by multiplexing a tagged source TS and the aggregate of the other sources $C_i - TS$ belonging to the same cluster C_i. To this purpose we will model the system as an MMPP/MMPP/1/q_{max} process.

The state of the system MMPP, which has been estimated given that the number of clusters in the system exhibiting a high criticality level NC_{C_1} is l, is denoted as $S^{(\Sigma_l)}(t)$, and can be written as

$$S^{(\Sigma_l)}(t) = (S^{(Q)}(t), S^{(\Sigma_l - Q)}(t)) \quad where \quad \Sigma_l = \{\Sigma | NC_{C_1} = l\} \tag{9}$$

The state space $\Im^{(\Sigma_l)}$, consequently, could be written as the Cartesian product of the state spaces of the server process Q and the complimentary system, except for the server, i.e. $\Sigma_l - Q$. Accordingly $\Im^{(\Sigma_l)} = \Im^{(Q)} \times \Im^{(\Sigma_l - Q)}$.

Observe that the process denoted as $\Sigma_l - Q$ is obtained by exploiting the processes identified before. More specifically, by saying $S^{(\Sigma_l - Q)}(t)$ the state of the $\Sigma_l - Q$ process, this state could be described through a transition rate matrix $\mathbf{Z}^{(\Sigma_l - Q)}$ and an array of the stationary state probabilities, $\mathbf{\Pi}^{(\Sigma_l - Q)}$.

$$S^{(\Sigma_l - Q)}(t) = (S^{(TS)}(t), S^{(C_i - TS.E)}(t), S^{(C_i - TS.CL)}(t), \tag{10}$$

According to the theory of MMPPs [10], the transition rate matrix of the Σ_l MMPP could be represented as the Kronecker sum of the matrices of the underlying processes Q and Σ_l. More specifically $\mathbf{Z}^{(\Sigma_l)} = \mathbf{Z}^{(Q)} \oplus \mathbf{Z}^{(\Sigma_l - Q)}$.

Now let us focus on the buffer which is one of the components of the overall system. We assume a "Late arrival system with immediate access" [2]. Observe that the service rate is not constant but varies as a function of the criticality level of the area where is located the cluster to which the considered TS belongs. Also, the service rate depends on the criticality level of the other clusters in the overall system, i.e. $\mu = \omega(C_i.ACL, S.CL)$ where $S.CL$ is the process representing the criticality of the other clusters in the overall system other than the considered one to which the TS belongs. More specifically, as will be thoroughly discussed in the following, the buffer management procedure uses a threshold such that, when the queue length is lower than this threshold, no packets are dropped. Then, only packets with low criticality level will be discarded; finally, upon reaching the maximum queue length, all packets, independently of their criticality level, will be dropped. Such a server process could be represented by a transition rate matrix $\mathbf{Z}^{(Q)}$ and a steady state probability array $\mathbf{\Pi}^{(Q)}$ which should again satisfy a system analogous to the one in eq. (1).

If we identify as $s_1^{(\Sigma_l - Q)}$ and $s_2^{(\Sigma_l - Q)}$ two states of the $\Sigma_l - Q$ process and $s_1^{(Q)}$ and $s_2^{(Q)}$ two states of the server process, the state transition rate matrix $\mathbf{Z}^{(\Sigma_l)}$ of the overall system can be written as: $z^{(\Sigma_l)}_{(s_1^{(\Sigma_l - Q)}, s_1^{(Q)}),(s_2^{(\Sigma_l - Q)}, s_2^{(Q)})} =$

$$= \begin{cases} z^{(Q)}_{s_1^{(Q)}, s_2^{(Q)}} & s_2^{(Q)} = s_1^{(Q)} + 1,\ s_1^{(\Sigma_l - Q)} = s_2^{(\Sigma_l - Q)} \\ z^{(Q)}_{s_1^{(Q)}, s_2^{(Q)}} & s_2^{(Q)} = s_1^{(Q)} - 1,\ s_1^{(\Sigma_l - Q)} = s_2^{(\Sigma_l - Q)} \\ \left[z^{(\Sigma_l - Q)}\right]_{(s_1^{(\Sigma_l - Q)}, s_2^{(\Sigma_l - Q)})} & s_1^{(Q)} = s_2^{(Q)},\ s_1^{(\Sigma_l - Q)} \neq s_2^{(\Sigma_l - Q)} \\ -\sum [z^{(\Sigma_l)}]_{(s_1^{(Q)}, s_1^{(\Sigma_l - Q)}),(s_2^{(Q)}, s_2^{(\Sigma_l - Q)})} & s_1^{(Q)} = s_2^{(Q)},\ s_1^{(\Sigma_l - Q)} = s_2^{(\Sigma_l - Q)} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

Observe that matrix $\mathbf{Z}^{(\Sigma_l)}$ is the following block matrix:

$$Z^{(\Sigma_l)} = \begin{bmatrix} \mathbf{Z}'^{(\Sigma_l - Q)} & \mathbf{Z}'^{(Q)} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{Z}''^{(Q)} & \mathbf{Z}'^{(\Sigma_l - Q)} & \mathbf{Z}'^{(Q)} & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{Z}''^{(Q)} & \mathbf{Z}'^{(\Sigma_l - Q)} \end{bmatrix} \tag{12}$$

where

- matrix $\mathbf{Z}'^{(\Sigma_l - Q)}$ is the transition rate matrix of the process $\Sigma_l - Q$ where elements on the main diagonal are substituted with the opposite of the sum of the elements of matrix $Z^{(\Sigma_l)}$ per each row.
- matrix $\mathbf{Z}'^{(Q)}$ is a diagonal matrix where elements on the main diagonal are the transition rates of the process Q from state i to $i+1$.
- matrix $\mathbf{Z}''^{(Q)}$ is a diagonal matrix where elements on the main diagonal are the transition rates of the process Q from state i to $i-1$.

By solving the usual set of equations in eq. (1), the value of the stationary state probabilities can be obtained. However, observe that solution of such a system could be very complex due to the excessive number of states. Accordingly, we exploit the Neuts theory [10] to solve the problem by reducing the complexity of a factor proportional to the maximum queue size, i.e. q_{max}. More specifically,

$$\begin{cases} \mathbf{\Pi}_0^{(\Sigma_l)} = \mathbf{\Pi}^{(\Sigma_l - Q)} \cdot [\sum_{q \in \Im^{(Q)}} V_q]^{-1} \\ V_1 = Z_{0,0}^{(\Sigma_l)} \cdot [Z_{1,0}^{(\Sigma_l)}]^{-1} \\ V_q = \left(-V_{q-2} \cdot Z_{q-2,q-1}^{(\Sigma_l)} - V_{q-1} \cdot Z_{q-1,q-1}^{(\Sigma_l)}\right) \cdot [Z_{q,q-1}^{(\Sigma_l)}]^{-1} \end{cases} \tag{13}$$

where the array of the stationary state probabilities for the states of the system depends on the stationary state probabilities of the process identified as $\Sigma_l - Q$.

Also, an emission rate array $\Lambda^{(\Sigma_l)}$ can be used to describe the rate of emissions depending on the criticality state of the cluster. Now, we could finally calculate the array of the steady state probabilities for the entire system:

$$\begin{array}{l} \Pi_y^{\Sigma} = \sum_{l=1 \ldots NC^{max}} \Pi_y^{(\Sigma_l)} \cdot \Pr\{NC_{C_1} = l\} \quad where \\ Pr\{NC_{C_1} = l\} = \binom{NC^{max}}{l} Pr\{\text{Cluster}_\text{C}\}^l \cdot (1 - Pr\{\text{Cluster}_\text{C}\})^{NC^{max}-l} \end{array} \tag{14}$$

and $Pr\{\text{Cluster}_\text{C}\}$ is the probability of the criticality state of the cluster and NC^{max} is the maximum number of clusters in the area.

4 Performance Analysis

In this section we investigate the loss probability and the delay distribution.

Loss Analysis. The buffer at the gateway can both queue or drop packets. To this purpose, the following scheme is employed for buffer management:

- The buffer queues packets independently of their criticality level as soon as the queue length remains below an appropriate threshold q_{Th};
- When the buffer queue length is between q_{Th} and q_{max}, low criticality level packets are dropped with probability 1;
- When the queue length is equal to the maximum q_{max}, all packets are dropped with probability 1, independently of their criticality level.

According to the above mentioned scheme for buffer management, the loss probability at a gateway buffer can be calculated as the sum of the probability to drop packets when the queue length is in $[q_{Th}, q_{max}]$ and when the queue length is the maximum possible q_{max}. More specifically,

$$P_{Loss} = \sum_{s^{(\Sigma)} \in \Im^{(\Sigma)} \text{ s.t.} s^{(TS.CL)}=0, q_{Th}<s^{(Q)}<q_{max}} [\Pi^{(\Sigma)}]_{s^{(\Sigma)}} + \\ + \sum_{s^{(\Sigma)} \in \Im^{(\Sigma)}, s^{(Q)}=q_{max}} [\Pi^{(\Sigma)}]_{s^{(\Sigma)}} \quad (15)$$

Delay Analysis. In order to characterize the delay, we apply the definition of average delay, i.e. $E\{\Delta\} = \int_0^{+\infty} \delta \cdot f_\Delta(\delta) d\delta$, where $f_\Delta(\delta) = \frac{\partial}{\partial\delta} F_\Delta(\delta)$. Consequently we will derive the cumulative distribution of the delay at a gateway node. More specifically, the probability that the delay Δ is lower than δ could be written as:

$$F_\Delta(\delta) = Pr\{\Delta \leq \delta\} = \mathbf{\Pi}^{(C_i.ACL,Q)} \cdot e^{-\mathbf{Z}^{(C_i.ACL,Q)} \cdot \delta} \cdot \mathbf{v} \quad (16)$$

with $\Pi^{(C_i.ACL,Q)} = Pr\{S^{(C_i.ACL)}(t) = s_{(C_i.ACL)}, S^{(Q)}(t) = s_q\}$,$\mathbf{v} = [1, 0, \dots 0]^T$ and $\mathbf{Z}^{(C_i.ACL,Q)}$ is the transition rate matrix whose generic element represents the transition rate from state $(s_{C_i.ACL_1}, s_{q_1})$ to $(s_{C_i.ACL_2}, s_{q_2})$.

5 Model of the System for the Different Approaches

In this section we will explicit the values of some system parameters depending on the addressed approach.

PA Approach. When considering this approach, the behavior of the sources does not depend on the criticality level. Accordingly, at the gateway packets will only be dropped based on a drop tail scheme when the maximum of the buffer length q^{max} is reached. More specifically, processes $C_i.ACL$, and $TS.CL$ will not be needed to characterize the system. So, the array of the emission rates at the TS source can be written as $\lambda^{(TS)}(t) = f_{State}(S^{(TS.E)}(t)) \cdot f_{TS\_Share}(N_{C_i})$.

Moreover, $C_i - TS.CL$ reduces to a process having a single state since all sources in the cluster will be characterized by having the same criticality level. The emission process at other sources in the cluster other than TS reduces to a single emission rate $\lambda^{(C_i-TS.E)}$. Finally, the state of the process $\Sigma_l - Q$ reduces to $S^{(\Sigma_l-Q)}(t) = (S^{(TS)}(t), S^{(C_i-TS.E)}(t))$.

Once the process $\Sigma_l - Q$ has been characterized, each term of the stationary state probability array can be written as $\mathbf{\Pi}^{\Sigma} = \mathbf{\Pi}^{\Sigma_l}$. The loss probability can be thus calculated as $P_{Loss} = \sum_{s^{(\Sigma)} \in \Im^{(\Sigma)}, s^{(Q)}=q_{max}} [\Pi^{(\Sigma)}]_{s^{(\Sigma)}}$ The average delay can be written as $E\{\Delta\} = \mathbf{\Pi}^{(Q)} \cdot \mathbf{T}^{-1} \cdot \mathbf{H}''' \cdot \mathbf{T} \cdot \mathbf{v}$, where $\mathbf{H}'''$ is defined as follows:

$$\mathbf{H}''' = \begin{bmatrix} -\frac{1}{h'_1} & 0 & \dots \\ \dots & \dots & \dots \\ 0 & \dots & -\frac{1}{h'_B} \end{bmatrix} \quad (17)$$

and h'_i with $i \in [1 \dots B]$ is the generic eigenvalue of matrix $\mathbf{Z}^{(Q)}$.

PDA Approach. In this case, the behavior of the sources depends on the individual criticality level. Accordingly, at the gateway, packets will be dropped with a policy which depends on their critical level. When considering this approach, process $C_i.ACL$ is not needed to characterize the system. So, the array of the emission rates at the TS source can be written as $\lambda^{(TS)}(t) = f_{State}(S^{(TS.E)}(t)) \cdot f_{TS\_Share}(N_{C_i}, S^{(TS.CL)}(t))$. Finally, the state of the process $\Sigma_l - Q$ reduces to $S^{(\Sigma_l - Q)}(t) = (S^{(TS)}(t), S^{(C_i - TS.E)}(t), S^{(C_i - TS.CL)}(t))$. Once the process $\Sigma_l - Q$ has been characterized, each term of the stationary state probability array can be written as $\mathbf{\Pi}^{\Sigma} = \mathbf{\Pi}^{\Sigma_l}$. The loss probability can be thus calculated as $P_{Loss} = \sum_{s^{(\Sigma)} \in \Im^{(\Sigma)}, s^{(Q)} = q_{max}} [\Pi^{(\Sigma)}]_{s^{(\Sigma)}}$. The average delay can be written as $E\{\Delta\} = \mathbf{\Pi}^{(Q)} \cdot \mathbf{T}^{-1} \cdot \mathbf{H}''' \cdot \mathbf{T} \cdot \mathbf{v}$, where $\mathbf{H}'''$ is defined as eq. (17) with h'' instead of h'.

ACA Approach. When considering this approach, the behavior of the sources depends on the criticality level of the area where the cluster is located. Accordingly, buffer management at the gateway will be the same ad in the PDA cases. In this case, process $TS.CL$ is not needed to characterize the system since it automatically comes from the aggregate criticality of the cluster. So, the array of the emission rates at the TS source can be written as $\lambda^{(TS)}(t) = f_{State}(S^{(TS.E)}(t)) \cdot f_{TS\_Share}(N_{C_i}, S^{(C_i.ACL)}(t))$. Moreover, $C_i - TS.CL$ reduces to a process having a single state since all sources in the cluster will be characterized by having the same criticality level. The emission process at other sources in the cluster other than TS reduces to a single emission rate $\lambda^{(C_i - TS.E)}$. Finally, the state of the process $\Sigma_l - Q$ reduces to $S^{(\Sigma_l - Q)}(t) = (S^{(TS)}(t), S^{(C_i - TS.E)}(t), S^{(C_i.ACL)}(t))$. Once the process $\Sigma_l - Q$ has been characterized, each term of the stationary state probability array can be written as from eqs. (1). The loss probability can be thus calculated as $P_{Loss} = \sum_{s^{(\Sigma)} \in \Im^{(\Sigma)}, s^{(Q)} = q_{max}} [\Pi^{(\Sigma)}]_{s^{(\Sigma)}}$. The average delay becomes: $E\{\Delta\} = \mathbf{\Pi}^{(C_i.ACL,Q)} \cdot \mathbf{T}^{-1} \cdot \mathbf{H}'' \cdot \mathbf{T} \cdot \mathbf{v}$, where $\mathbf{H}''$ is defined as eq. (17), where h'_i with $i \in [1 \dots B]$ is the eigenvalue of matrix $\mathbf{Z}^{(C_i.ACL,Q)}$.

6 Performance Results

We consider a scenario where $NC^{max} = 10$ clusters share an overall satellite capacity equal to 2 Mbps. In accordance to DVB-RCS communication standard, data sent by the gateways is transmitted in frames of 188 bytes containing a header of 4 bytes and 184 bytes of payload. We assume that all the sources in the clusters maintain active an audio call and use the GSM half rate encoding scheme, based on the *Vector Self-Excited Linear Predictor* (VSELP) codec at bit rate of 5.6 kbit/s. Also we assume that in a given area, a critical event occurs on the average each 10 hours and that the average duration of such critical event is one hour. Regarding the criticality level process, we assume that

- when the area criticality is low, on the average, each source becomes critical after 30 minutes and remains such for 10 minutes.
- when the area criticality is high, on the average, each source becomes critical after 10 minutes and remains such for 30 minutes.

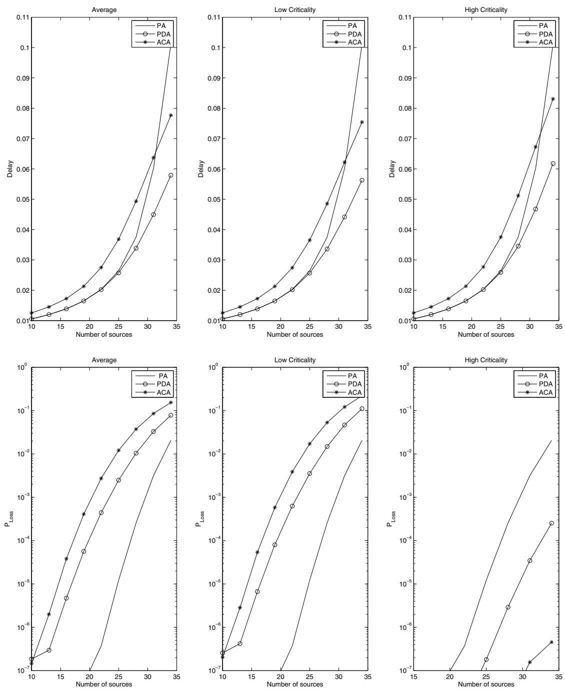

Fig. 1. Delay and loss probability as a function of the number of sources in the a) Average case b) Low criticality case c) High criticality case

Data packets coming from the sources at each gateway are buffered in a queue that can accommodate a maximum of $q_{max} = 30$ packets. In Figures 1, we show the average delay (up) and the packet loss probability (down) versus the number of sources in each cluster. The curves in the above figures have been obtained assuming that the threshold of the buffer size above which low criticality packets are dropped is set equal to $q_{Th} = 15$. In each figure we have three different plots:

– *Left plots* show the average delay and the packet loss probability for any packet (independently of its criticality).
– *Central plots* show the average delay and the packet loss probability for low criticality packets.
– *Right plots* show the average delay and the packet loss probability for high criticality packets.

Finally, observe that in each plot we provide three curves: one for each of the approaches analyzed. In Figure 1 we observe that the average delay increases as the number of sources increases and that PDA approach achieves the lowest delay, always. This is because it drops low criticality packets when the buffer size is higher than 15 and therefore, on the average, packets find smaller buffers; which results in shorter delays. Instead, by comparing ACA and PA we observe that the delay achieved by ACA is higher for most of the values of the number of

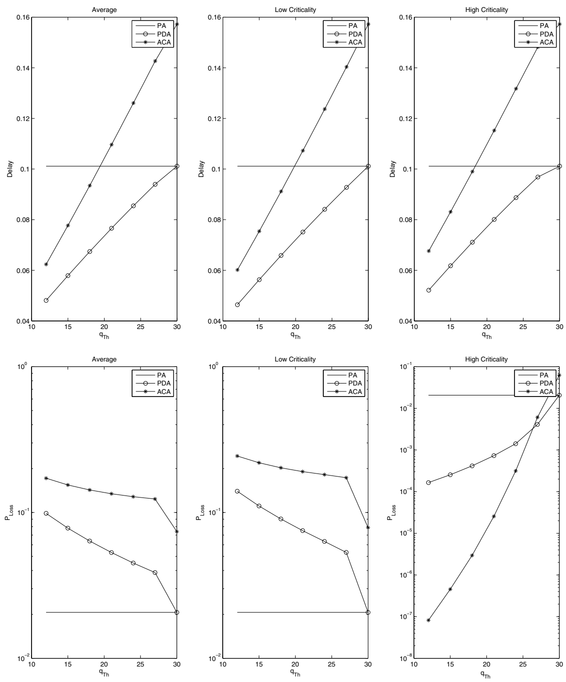

Fig. 2. Delay and loss probability as a function of the buffer threshold q_{Th} in the a) Average case b) Low criticality case c) High criticality case

sources. This is because ACA assigns more capacity to high criticality areas even when such additional capacity is not needed. When the number of sources in each cluster is high (and the traffic load is high as well) ACA achieves lower delay than PA. This is because the additional capacity given to high criticality areas is fully utilized and the delay is lower because ACA drops low criticality packets when the buffer size is higher than 15. In the left plot of Figure 1 we observe that the PA approach achieves the lowest packet loss probability. This is because the other approaches drop packets when the buffer size is higher than 15. However, in the right plot of the same figure we observe that the loss probability of high criticality packets is much lower in the case of the PDA and ACA approaches.

Similar observation can be drawn by observing Figures 2 where we show the average delay and the loss probability versus the threshold q_{Th} assuming that the number of sources in each cluster is equal to 34.

7 Conclusions

Satellite communications provide support to situation-awareness applications. However, areas where such services are deployed are usually equipped with only few gateway ports which should be shared among operators. Such operators usually move in groups, so clustering could be a natural solution to allow gateway sharing. However, due to the high cost of the satellite link especially when users request support of bandwidth demanding applications, bandwidth allocation should be adaptive and dynamically take into account the dynamics of the clusters and the importance, i.e. criticality, of the traffic being issued by the sources. In this paper, differently from the previous literature in the field, we provided a mathematical model of an aggregate of cluster sources and studied the dynamics of the bandwidth allocated based both on the criticality of the individual sources and of the area where the sources are located.

Acknowledgment. This work was partially supported by the European Commission in the framework of the FP7 Project IMSK (contract n. 218038).

References

1. Aweya, J., Ouellette, M., Montuno, D.Y.: A Control Theoretic Approach to Active Queue Management. Elsevier Computer Networks 36(2-3) (July 2001)
2. Bae, J.J., Suda, T., Simha, R.: Analysis of individual packet loss in a finite buffer queue with heterogeneous Markov modulated arrival processes: A study of traffic burstiness and a priority packet discarding. In: IEEE Infocom 92, Florence, Italy (May 1992)
3. Chu, M., Golub, G.: Inverse eigenvalue problems: Theory, Algorithms and Applications. Oxford Scholarship (September 2007)
4. Galluccio, L., Licandro, F., Morabito, G., Schembra, S.: An Analytical Framework for the Design of Intelligent Algorithms for Adaptive-Rate MPEG Video Encoding in Next Generation Time-Varying Wireless Networks. IEEE Journal on Selected Areas of Communications 23(2) (February 2005)
5. Galluccio, L., Morabito, G., Schembra, G.: Transmission of Adaptive MPEG Video over Time-Varying Wireless Channels: Modeling and Performance Evaluation. IEEE Transactions on Wireless Communications 4(6) (November 2005)
6. Li, S.-q., Park, S., Arifler, D.: SMAQ: A Measurement-Based Tool for Traffic Modeling and Queuing Analysis Part I: Design Methodologies and Software Architecture. IEEE Communications Magazine 2(7) (August 1998)
7. Lombardo, A., Morabito, G., Schembra, G.: A Discrete-Time Paradigm to Evaluate Skew Performance in a Multimedia ATM Multiplexer. IEEE/ACM Trans. on Networking 7(1) (February 1999)
8. Mahadevan, I., Sivalingam, K.M.: Quality of Service Architectures for Wireless Networks: IntServ and DiffServ Models. In: Fourth International Symposium on Parallel Architectures, Algorithms, and Networks, I-SPAN (June 1999)
9. Narula-Tam, A., Macdonald, T., Modiano, E., Servi, L.: A dynamic resource allocation strategy for satellite communications. In: Proc. of IEEE Milcom 2004 (October-November 2004)
10. Neuts, M.F.: Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. J. Hopkins University Press, London
11. Xu, J., Shen, X., Mark, J.W., Cai, J.: Adaptive Transmission of Multi-Layered Video over Wireless Fading Channels. Transactions on Wireless Communications 6(6) (June 2007)