Cognitive Multicast Routing with QOS Constraints for Multimedia Applications in Integrated Satellite- HAP System Structures

Mohamed M. Elsokkary  
Electronics and Communications  
Cairo University  
Giza, Egypt  
mo_elsokkary@yahoo.com

David Grace  
Communications Research Group  
University of York  
York, United Kingdom  
dg@ohm.york.ac.uk

Mona M. Riad  
Electronics and Communications  
Cairo University  
Giza, Egypt  
mnriad@link.net

Abstract—This paper presents the issue of capacity variable routing in Satellite-HAP system networks with multicasting capabilities, for different traffic types, with the objective of minimizing the short term traffic load oscillations and reacting to the long term changes of the traffic loads. This is realized by learning from the past and using it to influence the current decision. Some of the major QOS constraints (Delay and link Utilization) have been taken into account to meet the criteria of the advance multimedia applications. The proposed routing technique includes the joint utilization and delay metrics which allow the traffic load balancing between the satellite links and HAP links taking the advantage of both segments. The performance of the proposed routing techniques is evaluated using the appropriate simulation models, where the improved performance has been reflected in reducing the number of path re-routing by almost 50 %.

Keywords- Satellite, HAP, cognitive routing, traffic oscillation

I. INTRODUCTION

In the frame of explosive growth of the next generation of communication systems, it is required to have seamless integration of multimedia services over heterogeneous networks (wireline, wireless, terrestrial or space-based).

Satellite networks are playing a crucial role in the global infrastructure of the Internet, due to their global coverage and their capability of sustaining high capacity levels.

Moreover, HAPs (High Altitude Platforms) can provide a better link budget with respect to satellites with shorter communication delay; however, their coverage area is limited, therefore, they offer services on a regional basis. Due to their closer proximity to the ground, HAPs potentially have much higher capacity density and a similar per link capacity to satellite. Hence, satellite and HAP combined system could be enhanced by their joint operations to gain the satellite coverage with the shorter HAP delay and their higher capacity density.

Combined systems of HAPs and satellites have been discussed before where the main issue was to fix the problem of having the minimum spanning tree under some constraints like the delay and bandwidth. The proposed solution was to use a genetic algorithm to reach the optimum solution for this problem.

In this work, we will focus on studying in detail the integrated HAP-satellite scenario, using an integrated cost function with the assumption of taking the delay and utilization as QOS constraints embedded in the cost function. We will use cognitive routing techniques to improve the system performance.

Traffic Class dependent (TCD) routing for LEO systems has been introduced in [6], which reveals that the consideration of link load in the route calculation may cause traffic load congestion. In our work, the same phenomenon has been faced with the integrated HAP-satellite systems, and the concept of cognitive routing has been used to reduce that oscillation, where paths are dynamically routed to the lowest cost route, which unless controlled may immediately suffer congestion.

Cognitive Routing can be defined as the process of perceiving the current network conditions and then plan, decide and act on those conditions to achieve better performance [9]. In this work, this will be realized by taking into consideration the memory effect of the system, which will reduce the short term traffic load oscillations, while still reacting to the long term changes of the traffic loads.

Hence, the work in this paper is arranged as follows: Section 2 describes the proposed scenario using the combined HAP-satellite system. Section 3 describes in detail the traffic oscillation problem. Then in section 4, the system simulation model and the outcome of the simulation are illustrated.

II. INTEGRATED HAP-SATELLITE SYSTEM SCENARIO

The proposed architecture consists of three layers: the Terrestrial Layer, HAP Layer and Geostationary satellite (GEO) Layer, as in Figure (1).

In that system, it is assumed that the HAP and satellite nodes are routers with OBP (On board Processing) capabilities.

Although the links through the satellite encounter high delay, the satellite has wide area coverage and multicast capabilities. HAPs have smaller footprints but suffer less delay than the links passing through the satellite. So
combining both is a good solution to take the advantage from each system and mitigate the drawbacks.

In our proposed scenario, the HAPs split the link between the earth terrestrial station and the satellite in both up- and down-links, so the source nodes have the capabilities of communicating through the HAPs network or via the satellite.

The proposed routing algorithm calculates the best minimum spanning tree paths from the source nodes to all receiving ones.

Figure 1. Integrated HAP-Satellite Proposed Scenario

A. Link Cost Metric for Different types of traffic

In this work, it is assumed three different types of traffic, according to the required QOS as follows.

- Class A traffic representing “delay sensitive traffic” like the transfer of real time audio and video applications.
- Class B traffic, which needs to maximize the throughput like the on demand interactive videos.
- Class C Traffic: representing “non-real time traffic” like the emails, which requires the best effort QOS.

Accordingly, it is proposed to have three different cost functions according to the traffic type as follows:

\[
\text{Integrated Cost function} = \begin{cases} 
\text{Delay (P) / } \alpha (\text{delay (P) – MD}) & \text{for Class A Traffic} \\
\sum_{l_i \in P} \text{utilization}(l_i)/(1- \text{Utilization}(l_i)) & \text{for Class B Traffic} \\
\sum_{l_i \in P} \text{Delay (li)*utilization}(l_i)/(1- \text{Utilization}(l_i)) & \text{for Class C Traffic} \\
\end{cases}
\]

\[
\alpha (z) = \begin{cases} 
1 & \text{if } z \leq 0 \\
y & \text{if } z > 0 \\
\end{cases}
\]

Where, \( y = \{0, 1\} \), and it represents the penalty factor.

\(\text{Utilization}(l_i)\): represents the used capacity for the link \( l_i \).

\( P \): Total end to end path from the source to destination.

\( MD \): Maximum bounded accepted limit for the delay.

B. Multicast node versus Unicast node concept

Multicast is a communication service that allows simultaneous transmission of the same message from one source to a group of destination nodes. Multicast routing creates a tree connecting the sources to all the receivers in the group, where the multicast node duplicates the data and sends it to all receivers. It is obvious the transmission efficiency due to using the multicasting nodes, as illustrated in figure (2).

We assume that the source is sending the data to the core node (i.e. the HAP or satellite) which will distribute the traffic to the receiving nodes, under its footprint. Taking the assumption that the source is utilizing 33 % of link capacity to send the data to each receiver, it can be concluded that

- For the Unicast Scenario: the link utilization between the source and the core node will be equal \((3 * 33 \% = 99 \%)\).
- For the Multicast scenario: the link utilization between the source and the core node will be equally \(33 \%\), since the processing will be in the node itself with no impact on the link utilization between that source and the core node.

In our proposed scenario, it is assumed that the satellite and HAP have OBP where the nodes can regenerate the signal and perform switching, and multiple spot beams (where the satellite/HAPs may send data to one or more antennas directed at different parts of the satellite/HAP coverage area).

C. Centre Based Multicast Distribution Tree

Multicast routing algorithms can handle the distribution of the packets in three ways:

1) Basic methods like (Flooding and Spanning Tree).
2) Source Based Tree, where each source creates its own distribution tree, examples of this kind of algorithms are:
   a) Forward Shortest Path.
   b) (Reverse Path Forwarding (RPF) algorithms).
   c) Steiner Trees (ST).
The main disadvantages of these algorithms is that each router must generate large routing table, since one routing item is needed for each source and each group.

3) Shared Tree, where we have assigned core nodes and if node is aiming to join the one group, it should send a unicast message to the core node.

An example of that type of algorithm is the Centre Based Trees (CBT). In CBT all sources sends the packet to the same tree, and the same tree distributes the packets to all receivers, this can be illustrated in figure (3).

In this work, it is assumed that the shared tree which consists of the core nodes are the satellite and the HAPs. So the sources are sending to this tree which distributes the traffic to all receivers, this can be illustrated in figure (3).

Figure 3. Illustration for the concept of CBT (Centre Based Tree)

III. PROBLEM FORMULATIONS

A. Traffic Oscillation problem

A QoS multicast routing problem usually involves multiple constraints, such as delay, jitter, bandwidth, and packet loss rate. In order to develop a multi-objective QoS multicast routing model, we determine multicast routes satisfying the major constraints of “End-to-end” delay and Bandwidth utilization.

Hence, as explained in section 2, the integrated cost function takes into consideration the above factors (Delay and Utilization), and also it differs according to the type of traffic (Class A, B and C). Solving the multi-objective QOS functions with many constraints (delay, BW) is considered as NP-Complete problem and has been extensively discussed before in [1],[3] with solutions derived using genetic algorithms, which is out of the scope of this work.

Consideration of link utilization in the cost metric causes the traffic load oscillation phenomenon which can be explained as follows. In each routing update, the cost function is checked by all the routers, and the routers can detect that there are some free links in the network where the traffic can be re-routed, so the whole traffic is re-routed to these free links and let us name them as “link group 1”, and then in the next snapshot (next update to the routing table), the routers discover that some links are free, low utilized and have less link cost metric than “link group 1”. Accordingly, another re-routing to the new free links occurs and let us name them as “link group 2”, which will release the traffic from “link group 1” links resulting in reducing their link cost metric and make them the eligible candidates for the next path selections during the routers updates and so on. In our work, it can be assumed that “link group 1” and “link group 2” represent the links that pass through HAPs and satellites respectively.

This traffic oscillation is inconvenient under heavy traffic loads, where the impact of the traffic load is higher than the link delay. Such traffic oscillation affects the quality of service, and the main objective is to reduce such oscillations.

B. Smoothing link Cost Functions

In order to fix the oscillation problem, It is required to build memory into the system so that the system can learn from its past and use it in its current decisions. The main goal of the smoothing link cost function is to avoid the instantaneous fluctuation of traffic which leads to traffic oscillation, and to reduce the traffic path re-routing. Basically a node will remember whether a previous action worked well or became congested. In this work, exponential decaying function has been introduced to weigh the history of the previous paths, Y =aX, Where X represents the snapshot number, a = ] 0, 1[.

Applying this decaying function should help control the system performance as we will take the impact of the past values of the system in the current cost metric decision, which should help to minimize the traffic oscillation as will be illustrated in the simulation result part.

So if we assume that the cost function value of the paths in snapshot X is ‘CFX’, and in the previous snapshot is CFX-1. So we propose to apply the following equation to take the routing decision based on system memory.

\[ CFX=\frac{aCFX + a1CFX-1 + a2CFX-2 + a3CFX-3 + \ldots \ldots + axCFX-x}{(x+1)} \]  

Where a = ] 0, 1[ representing the weighting variable between the sequential snapshots.

Moreover, another tuning factor b is introduced to control the oscillation, i.e., if the difference between the cost function CFX value and CFX-1 is less than that tuning factor b, the existing paths will remain even if the other paths are better in terms of cost function calculation. This should minimize the path re-routing, as will be illustrated in the simulation part.

IV. PERFORMANCE ANALYSIS

A. Simulation Model

The simulation makes use of Matlab and the BIO-informatics toolbox, and we assume the following:

- 3 static continuous traffic sources S1, S2 and S3, where S1 is representing Class B Traffic, S2 representing Class A, and S3 is representing Class C one.
- It is assumed that class A traffic is VIP traffic, which means it has the priority to utilize the best paths in the network.
- The links between all nodes (satellite, HAPs, source generators and receivers) have equal capacity.
- Each source generates traffic which is utilizes 33 % of the link capacity.
- Each source node sends to its separate receivers, which are locating together under the same HAP.
• The network structure can be seen as two parts:
  a) Transportation part, which is a common part between the sources and all receivers. The core multicast nodes are selected from this part. This part includes the satellite and the HAPs (7 Nodes in the simulations). The nodes in this part are considered as routers where they receive the traffic from the source generator nodes and distribute it to the receiving nodes in the network based on certain cost metric as explained in the integrated cost function in section 2.
  b) Access part, which is the end receivers (Earth stations), under the satellite and HAPs footprints (32 nodes). These nodes are considered to be end terminals in our proposed network structure, where they receive the traffic from the source generators via the transportation layers.
• The receiving nodes are assumed to be fixed nodes.
• The satellite footprint covers all the nodes, and the nodes can receive the traffic either via the satellite or HAP routes.
• The traffic generators are located under the footprint of HAPs and satellite and they have the ability to send traffic to any path (satellite or HAPs).
• The simulation will be performed with different values for parameters “a” and “b”, as defined before in section 3.

B. Simulation Results

1) Case of delay sensitive traffic (Class A)

It is assumed that this traffic is high priority traffic, so the routes are chosen from the beginning based on the shortest spanning tree and there is no traffic fluctuation. It is noticed that the paths tend to take the HAP route which has less delay than the satellite ones. This is represented in our simulation by the traffic generated by the S2 source.

This scenario can be realized practically and in the simulation using the smart selection by the receivers using the “First HAP selection” algorithm as explained in detail in [4], where the smart sending node can choose the best paths to carry this high priority traffic. Hence, Source 2 in our simulation generates a continuous static level of traffic which utilizes all the used links by 33 % and leaves the residual capacity to the traffic coming from the other sources S1 and S3 according to certain pre-defined criteria illustrated above in the cost function definition in section 2.

2) Case of Class B traffic

As explained in section 3, Figure (4.a), shows the effects of traffic oscillation between the source node S1, and one of the receiving nodes which is node 28 during different 15 sequential snapshots. This case is similar to the connections between the source node S1 and all the other nodes.

As illustrated in fig (4.a) the cost function value is different for different snapshots as the paths are changing in terms of its utilization. After applying the proposed cognitive routing algorithm, the system performance can be depicted in Figure (4.b) for different values of the decaying parameters “a”. The oscillation has been controlled, under the assumption that b=0.08 which is the fine tuning parameters in the system. Hence, the system reaches its stable stage at certain snapshot (Snapshot 6 in our simulation). This reduces the number of path re-routings in the system, which will improve the system performance. As can be seen that the system can reach its steady state faster when the decaying parameter “a” has a higher value.

As another measure for the proposed routing algorithms, and illustrated in Fig (5) is that of cumulative path re-routings in the system, represented by the path connection between Node 1 and node 28. It is shown that the system reaches its steady state at snapshot 6, after applying the proposed algorithm, which improves the number of path re-routing in these 15 sequential snapshots by almost 50 %.

Figure 4. System Performance before and after applying the routing algorithm

Figure 5. Cumulative Path Re-routing reduction after applying the cognitive algorithm (b=.08, a=0.5)
3) **Case of Class C traffic (Best effort QOS traffic)**

As explained in section 2, that this traffic class represents the non-real time traffic. In this work, we applied the same proposed cognitive routing algorithms for this traffic emerging from traffic source S3 to reduce the number of path re-routings which improves the system performance like the one illustrated in fig.(5).

Finally, after checking the system behavior in detail in each snapshot, Figure (7) can be used to describe the system performance. Figure (7.a) shows that the minimum spanning tree routes the traffic through the HAP links rather than letting the traffic pass through the satellite, meaning that these links will be utilized and the rest will be free. Then in Figure (7.b), the traffic passes through the satellite where the utilization is less for these links. The links through the HAP will then be freed which leads to the traffic being re-routed again through HAP in the following snapshots.

After building the cognitive routing techniques explained in this paper, it can be noticed that when the system reaches steady state, the traffic will be balanced between both the satellite and HAP links simultaneously. This is shown in Figure (7.c), where the source node has multiple routes to balance the traffic which mitigates the path traffic oscillation phenomenon and minimizes the number of traffic re-routings.

![Figure 7. The routing concepts:](image)

- a. Paths passed through HAP links.
- b. Paths passed through Satellite links.
- c. Balanced paths between the Satellite and HAP links.

Our future work will focus on the study of Satellite-HAP system performance using different types of traffic generators (dynamic sources) and their resultant impact on this system. In addition, we will study the multipath routing in this Satellite-HAP systems.

V. **Conclusion**

This paper has introduced a new integrated cognitive cost function which includes the link utilization and delay constraints affecting the QOS for multimedia applications, for satellite-HAP networks. This cost function takes into account history of previous events to better stabilize traffic flows. We have studied the traffic routing in this joint satellite-HAP network under different traffic classes and proposed routing techniques based on obtaining the minimum spanning tree where the HAP or satellite will be the core node to multicast the traffic to all destinations.

It is found that the use of link utilization can give rise to traffic oscillation when no memory of past events is included, which increases the path-re-routing which affects the QOS. When history of previous traffic flows and their paths are taken into account, using a decaying window function it is found that these oscillation effects can be minimized. It can be concluded that the system reaches the steady state when the traffic is balanced between HAP and satellite links. The proposed routing techniques have improved the routing efficiency by spreading the traffic in the network and reduced the heavily loaded links which improves the network throughput.

**REFERENCES**


