Mobility Support for a QoS Aggregation Protocol

A. Kaloxylos^, D. Vali*, S. Paskalis+, G. Panagiotou^, I. Gonianakis^, E. Zervas #
^ Department of Telecommunications' Science and Technology, University of Peloponnese, Greece
* OTE Research, Hellenic Telecommunications Organization - OTE S.A,, Greece
+ Department of Informatics and Telecommunications, University of Athens, Greece
#Department of Electronics, TEI Athens, Greece

Abstract - The end-to-end QoS support for multimedia applications is an important issue. However, existing proposals either end up in a coarse-grained QoS support or they are simply not scalable. To address this issue, dynamic aggregation of individual reserved sessions into common classes is recommended.
In end-to-end communications it is important to carefully select the starting and ending point for an aggregation. The goal is to minimize state information and processing load in the routers. Moreover, the capability of the users to switch their active connections/sessions from one point to another, and even from one operator to another, places extra requirements in the management of the aggregated information. Our work discusses the performance of existing protocols and builds upon BGRP, a scalable protocol that is designed to efficiently aggregate resources. We propose the addition of mobility extensions to BGRP and assess their performance through some simple metrics.

I. INTRODUCTION

QoS support is a well-studied issue and several proposals and standards have been issued during the past years. A number of protocols are proposed to handle QoS in local networks or even end-to-end ([1], [2]). The IETF "Next Steps in Signaling" working group ([3]) is currently working to standardize IP signaling protocols with QoS signaling as the first use case. In [4], a protocol for the routing and transport of per-flow signaling messages is presented, while in [5] the NSIS Signaling Layer Protocol (NSLP) for signaling QoS reservations in the Internet is described. These protocols do not follow the coarse-grained QoS support of Diffserv [7], but provide similar functionality to RSVP [6]. Obviously, special care has been taken to support additional features such as sender based reservation as well as receiver based reservations. From these efforts it is clear that in order to dynamically control end-to-end QoS schemes, signaling has to travel from one end to the other each time a new session is to be established. However, this may raise scaling issues, especially for the core network routers. To ameliorate this burden, aggregation of signaling information is required. The scheme that one should use to minimize the processing load and the signaling information stored in the routers remains as an open issue.

Another challenge is to study how these aggregation schemes will perform with mobile terminals. Future scenarios expect users to be able to vertically handoff their connections from one radio access technology to another or even dynamically select to switch from one Operator, or Internet Service Provider, to another (e.g., handover from/to WLAN to/from WiMAX, UMTS etc). In these cases it is possible that the new end-to-end path, although sharing a large segment with the old one, will not include the previous aggregation point. Thus, resources need to be re-established in an end-to-end fashion, despite the fact that a portion of the previously established path could be re-used.

To tackle this issue our proposal builds on BGRP [8], an existing inter-domain protocol for aggregating resources. With minor modifications we have designed an extended version of it, called Mobile BGRP (MBGRP) that efficiently handles the movement of terminals.

This paper is organized as follows. In Section II, we discuss existing protocols designed to aggregate reservations. In Section III, we examine more closely aggregation schemes' scalability issues and discuss the reason why we need mobility support for these protocols. In Section IV, we present the details of MBGRP and perform a simple quantitative evaluation of the new protocol. We conclude the paper in Section V.

II. RELATED WORK

There are several alternatives for aggregating information. The first one is described in [9] and suggests the use of a single RSVP reservation to aggregate other RSVP reservations across a transit routing region. With this technique, an aggregation region is defined as a contiguous set of systems that can perform aggregations along any possible route throughout this region. Obviously, this solution manages to reduce the number of states and signaling exchange inside these areas. However, it does not specify a selection mechanism for the appropriate placement of aggregators and de-aggregators in the end-to-end path. It merely specifies as appropriate selections the ingress and the egress routers of the aggregation region. Thus, we believe that if used in end-to-end schemes this solution will result in a significant number of aggregates for the nodes located inside the core network.

Another proposal is described in [10]. The DARIS (Dynamic Aggregation of Reservations for Internet Services) architecture assumes the existence of a central resource management entity (similar to the bandwidth broker approach) inside each DiffServ domain that has a complete knowledge and control of the resources inside the domain. It also avails the inter-domain BGP routing table. DARIS enables the creation of an aggregate between two arbitrary domains as soon as a threshold of active common reservations between the two domains is exceeded. In this case, all intermediate edge routers can substitute the respective per-flow states with a single aggregate state. The DARIS aggregation approach reduces stored states and signaling messages when compared to a per-flow protocol that does not perform aggregation. However, the penalty to pay is the need for this central entity and the mechanisms to keep it updated with routing

and QoS information. Also, the support of nested aggregates adds to the complexity of the protocol.

A QoS signaling protocol specifically designed for inter-domain usage between heterogeneous domains (AS - Autonomous Systems) is the Border Gateway Reservation Protocol (BGRP) [8]. BGRP operates in end-to-end mode only between domain border routers. It aims mainly at aggregating reservations between administration domains improving thus, scalability. BGRP uses the sink-tree aggregation approach and performs reservation aggregation by building a sink tree for each destination domain. Reservations from different source domains that are targeted towards the same destination domain are aggregated along the path forming a sink-tree. The root of this tree is the destination's domain edge router. BGRP's basic functionality includes a PROBE message sent by the source domain towards the receiver to determine resource availability and to record the reservation path. The destination domain edge router terminates the PROBE and sends back a GRAFT message along the reverse path. This message performs the actual inter-domain reservation and triggers the intra-domain QoS mechanisms in transit domains. The GRAFT messages follow the same path with PROBE message since record route information is gathered during the probing phase. The destination domain edge router is the de-aggregation point for the reservation aggregate, while the source domain edge routers are the aggregation points. By performing sink tree based aggregation of reservations towards each destination domain, BGRP results in storing per-destination domain QoS states in border routers. This is a significant contribution to scalability when compared to the per-flow RSVP. Analysis results presented in [8] show that BGRP maintains fewer reservation states than RSVP and that the BGRP message rate is significant lower than the respective RSVP message rate, resulting in lower message processing, storage burden and link bandwidth.

The Shared-segment based Inter-domain Control Aggregation Protocol (SICAP) ([11]) is another approach for supporting aggregate inter-domain reservations between Autonomous Systems (ASs). SICAP combines the shared-segment aggregation and the tree-based aggregation approaches to create tree-based reservation aggregates that do not necessarily extend up to the destination domains. Apart from the destination domain de-aggregator point, Intermediate De-aggregation Locations are discovered and created along the path, so that reservation requests that share a common path segment but do not end-up to the same destination domain are aggregated up to one of their common routers along the path. SICAP, similarly to BGRP, performs receiver-based reservations and uses a two-phase setup mechanism for their establishment. Simulation results support that SICAP maintains fewer aggregate states than BGRP. However, SICAP is a more complex protocol than BGRP since an algorithm to elect the most suitable intermediate de-aggregation points is needed. Moreover, the choice of using these points may end up in situations where core routers will have to de-aggregate and aggregate from scratch a significant number of active sessions.

III. SCALABILITY ISSUES

To provide a clearer picture about the state information needed when aggregating resources, we provide a short analysis for [8] and [9]. Aggregate RSVP is currently supported on the ongoing work of NSIS. BGRP is, in our opinion, not only scalable but also the simplest from the existing alternatives. Thus, it is more likely to be considered for adoption in the future. BGRP can be more scalable than aggregate RSVP and moreover it can be easily extended to work in mobile environments.

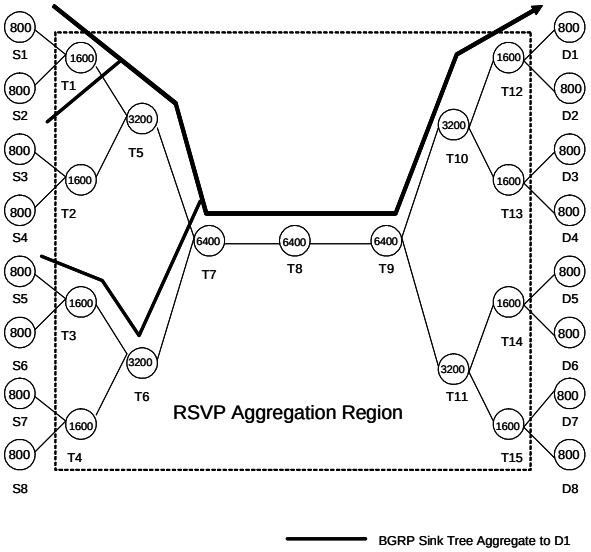

Figure 1: BGRP and RSVP aggregation

As an example consider fig. 1 where a simplified topology in the form of a binary tree structure is drawn. Every circle represents an AS. From every source AS (S1-S8) N unidirectional sessions are active and are equally distributed in every one of the target AS (D1-D8). As an example N equals 800 in fig1. Inside every circle in the transit networks we have also marked the number of connections they serve. Without aggregation we can observe the scalability issue inside the core network.

Suppose that we use aggregate RSVP and define an aggregation region, with T1, T2, T3, T4 as the aggregators and T12, T13, T14, T15 as the de-aggregators. Then the large number of states that had to be stored in T7,T8,T9 is minimized. Each of these nodes will now only to have 16 reservation states, while T5, T6, T10, T11 only 8. However, T1-T4 and T12-T15 will still have a large number of states to support.

In order to find the number of signaling states that we can summarize in the whole network if [9] is used, we define as *L* the tree length. *"a"* and *"d"* are the aggregator's and the de-aggregator's level in the hierarchy. Each leaf of the tree is considered to be in level 1. $V_i(L)$ is the number of connections passing through a node (i.e., AS) in level *i* and is given by $V_i(L) = c*2^{i-1}$. *c* is the number of connections in each leaf of the network (e.g., 800 in the example of Figure 1). $H_i(L)$ is the number of nodes in the binary tree at level L and is given by $H_i(L) = 2^{L-i}$. In the example given, $H_1(L=5)=16$, half of which play the role of source domains and the others act as destination domains.

In a general case the total gain from the number of states that can be eliminated due to aggregation is given from (1):

$$\sum_{i=a+1}^{L} 0{,}5 * V_a(L) H_a(L) + \sum_{i=d+1}^{L-1} 0{,}5 * V_d(L) * H_d(L) - \\ - \sum_{i=a}^{L} 0{,}25 * H_d(L) * H_a(L) - \sum_{i=d}^{L-1} 0{,}25 * H_d(L) * H_a(L) \quad (1)$$

The first two terms calculate the number of states in the levels between the aggregator and the de-aggregator that will be replaced by aggregated states. The last two terms represent the number of the new aggregated states that will be installed in all nodes from the aggregator to the de-aggregator.

Note that the incentive for one to act as an aggregator/de-aggregator is small since this node will have all states of the active connections and in addition the new aggregated states, i.e., it will suffer an "aggregation penalty". Equation (1) holds also if the aggregators and de-aggregators are placed in different tree levels. Studying (1) we see that in order to maximize the gain (i.e., minimize the total number of states in the network) we need to build the longest possible aggregates. However, in this case the core network nodes will have more aggregates states, the number of which is equal to the number of nodes in the aggregators' level multiplied by the number of nodes in the de-aggregators' level (e.g., 16 states for T8 in the example of Figure 2).

In the case of BGRP the corresponding equation for the same topology is given by (2):

$$\sum_{i=2}^{L} 0{,}5 * V_2(L) H_2(L) + \sum_{i=2}^{L-1} 0{,}5 * V_2(L) * H_2(L) - \\ - H_1(L) - \sum_{i=1}^{L-1} 0{,}25 * H_1(L) * H_i(L) - \sum_{i=d}^{L-2} 0{,}5^{L-i+1} * H_1(L) * H_i(L) \quad (2)$$

From (2) we can see that the first two terms are equal to the ones of equation (1) when a=1 and d=1. However, the number of the aggregated states is at most equal to the number of destination domains (8 states in the example of fig 1). Obviously this fact places an upper bound to the last two terms of (2). Using BGRP, one can expect that the source ASs will choose to act as aggregators if they want their users' session to enjoy an end-to end QoS. In this case, the "aggregation penalty" is unconditionally imposed to all the edge domain border routers. From the above elements it is clear that in an end-to-end communication BGRP behaves more efficiently than Aggregate RSVP.

However, what both protocols (i.e., [8] and [9]) do not offer is the support for the cases where a terminal moves from one area to another such that the new branch of the end-to-end path bypasses the aggregator. For example, in case a terminal located in AS S2 performs a handover (e.g., from a UMTS operator to a wireless ISP) to the S3 system then neither BGRP nor aggregate RSVP will have any means to detect that. Thus, end-to-end signaling will be issued to reserve resources that are already reserved. In this case, the moving terminal will face unnecessary delay and the operators will end up with an amount of double-reserved resources due to moving terminals.

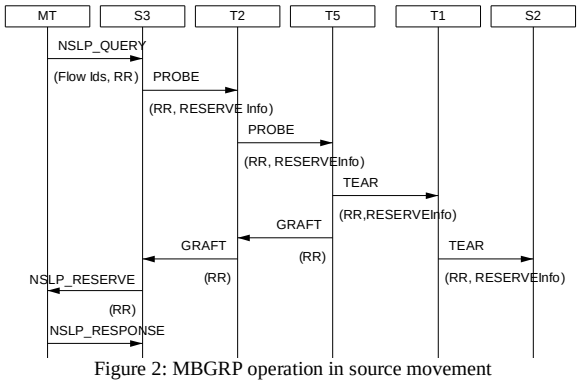

Figure 2: MBGRP operation in source movement

IV. MBGRP DESCRIPTION

MBGRP is an extension of BGRP that takes into consideration the terminals' mobility. It is designed to cooperate with the protocols designed in NSIS. For this, we assume that MBGRP uses GIST [4] to transfer its messages. We also assume that the border routers in the stub domains (i.e., S_i, D_i, where i =1..8 in fig 1) support both NSLP [5] and MBGRP. As with BGRP, MBGRP needs to be placed only in the border routers of the ASs. In addition to BGRP, MBGRP requires that the record route information be communicated to the end terminals, through, for example, NSLP. This information is needed in case of a handover, and is used by the mobile terminal to notify MBGRP routers that resources have already been reserved for a certain path. The information is also used by the MBGRP routers to check if they are crossover routers or not. Another extension to BGRP is the addition of two messages, called R_PROBE and NOTIFY. These messages are needed when terminals located at the root of the MBGRP sink tree change their location.

When a new session needs to reserve resources (receiver initiated mode), then an NSLP QUERY message is issued. This message reaches the border router of the AS (e.g., S1). This router will issue a MBGRP PROBE message that includes the QUERY's message information. The message will eventually reach the border router of the destination AS (e.g., D1). The de-aggregation point will issue a NSLP QUERY message towards the destination host. The destination host will issue a NSLP RESERVE message that will eventually reach the source host. This information will be transferred with GRAFT messages through the MBGRP enabled routers. Finally, NSLP messages will be transferred by GIST, transparently through the MBGRP routers, to the destination host. This way, resources are reserved in an end-to-end fashion but also, the inter-domain signaling states and resources are automatically aggregated as described in [8]. PROBE and GRAFT messages contain the record route of the path. This information is used in order to find the crossover MBGRP node when a terminal changes ASs.

For example, suppose that a terminal moves from S2 to S3. In the BGRP case, end-to-end PROBE and GRAFT messages need to be exchanged. What takes place in the case of MBGRP is shown in fig 2.

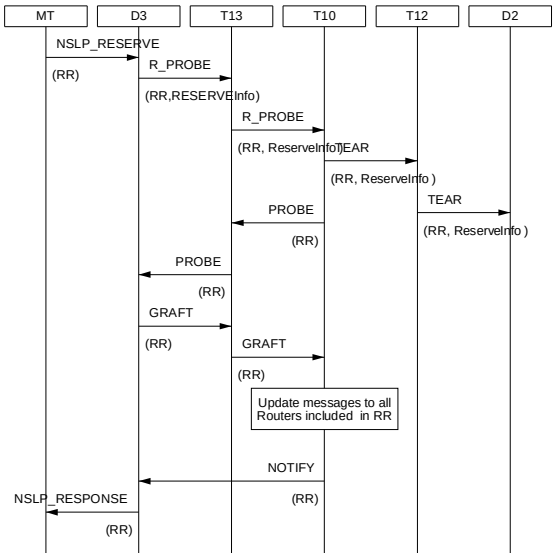

Figure 3: MBGRP operation in destination movement

Firstly, a NSLP QUERY message will be issued (receiver based reservation). The message contains the record route of the active sessions. The first MBGRP router sends a PROBE message that will reach the crossover MBGRP router. This router will recognize that it is a crossover router (its address is in the record route list) and will initiate two new messages: a TEAR to release resources in the old branch, and a GRAFT to the new location of the terminal.

A slightly different procedure takes place when the moving terminal is at the root of a sink tree . Consider the case where the terminal changes its location from D2 to D3. Since it is a receiver-initiated reservation, a NSLP RESERVE will be issued for its active sessions. The first MBGRP router needs to initiate the reservation and aggregation of resources on the new branch. Thus, it sends a reverse PROBE (R_PROBE) message that eventually reaches the crossover MBGRP router. This will tear the resources in the old path and will initiate a PROBE/GRAFT exchange for the new path. Upon successful completion of this procedure it will notify the border MBGRP router of the new AS to send a NSLP RESPONSE back to the terminal. The crossover router has to update in the upstream direction (i.e., towards the source domains) the information stored in the MBGRP routers about the relation of reserved bandwidth and the roots of the sink trees.

The number of signaling exchanges and processing for BGRP in case a terminal moves from one place to another is always:

*Cost = 3 * Distance (src location, dst location)*

This is because end-to-end PROBE/GRAFT and TEAR messages have to be issued. For the cost in MBGRP we distinguish two cases. In the first case, a terminal is moving from one source AS to another:

*Cost = 2*Distance (crossover, new location) + Distance(crossover, old location)*

The first term is related to the exchange of PROBE/GRAFT messages and the second one to the dispatch of TEAR messages.

In the second case, a terminal located in the root of a sink tree changes its location and the related cost is given by:

Cost= 3 Distance(crossover, new location) + Distance (crossover, old location) + Distance (Crossover, src location).*

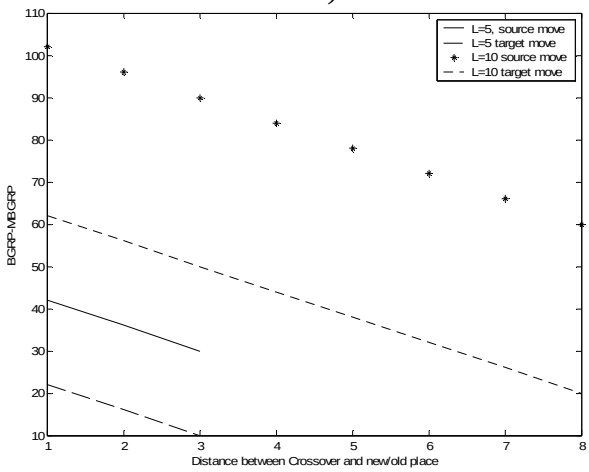

Figure 4: BGRP Vs MBGRP

This cost is higher, since apart from the PROBE, GRAFT and TEAR messages, R_PROBEs need to be issued and UPDATE messages must travel all the way back to the communicating terminals in source ASs.

In fig 4 we present, for illustrative reasons, some simple examples of the cost difference between BGRP and MBGRP, for a binary tree topology where the tree length L=5 for the first example and L=10 for the second. In these examples, we have assumed that the moving terminal has 2 active sessions and it can move only from one leaf to another.

As we can observe, MBGRP needs to exchange far less messages even in the cases where a terminal moves from the first leaf of the tree to the last (e.g., from S1 to S8), or when a terminal located at the root of the sink tree changes its location.

V. CONCLUSIONS

In this paper we have presented arguments on necessity and merits of aggregation signaling, and examined more deeply the scalability issues of two aggregation protocols. After selecting BGRP, based on its performance and its potential to operate in mobile environments, we presented an extended version that can work more efficiently when a terminal's movement obsoletes the old aggregators' location.

By means of simple analysis we have shown that MBGRP results in fewer messages, minimization of the load in core routers as well as the time needed to re-establish the end-to-end QoS supporting path.

REFERENCES

[1] J. Manner, X. Fu, Analysis of Existing Quality of Service Signalling Protocols, RFC 4094, May 2005.

[2] D. Vali, S. Paskalis, A. Kaloxylos, L. Merakos, A survey of Internet QoS Signalling, IEEE Communications Surveys and Tutorials, Fourth Quarter 2004, Vol6, No 4.

[3] IETF Working group, Next Steps in Signalling, http://www.ietf.org/html.charters/nsis-charter.html.

[4] H. Schulzrinne, R. Hancock, GIST: General Internet Signaling Transport, Internet Draft draft-ietf-nsis-ntlp-08.txt

[5] J. Manner, et al., NSLP for Quality-of-Service signalling, Internet Draft, draft-ietf-nsis-qos-nslp-09.txt.

[6] Braden, R., Zhang, L., Berson, S., Herzog, S. and S. Jamin, Resource Reservation Protocol (RSVP) Version 1 Functional Specification, RFC 2205, September 1997.

[7] S. Blake et al., An Architecture for Differentiated Services, IETF RFC 2475, December 1998.

[8] P. Pan, E. Hahne and H. Schulzrinne, BGRP: A Tree-Based Aggregation Protocol for inter-Domain Reservations, Journal of Communications and Networks, Vol. 2., No. 2, June 2000.

[9] F. Baker, C. Iturralde, F. Le Faucher, B. Davie, Aggregation of RSVP for IPv4 and IPv6 Reservations, IETF RFC 3175, September 2001.

[10] R. Bless, Dynamic Aggregation of Reservations for Internet Services, ICTSM 10, Oct. 2002.

[11] R. Sofia, R. Guerin, and P. Veiga, SICAP, a Shared-segment Inter-domain Control Aggregation Protocol, Technical Report, ESE, University of Pennsylvania, October 2002