# The Multicast element in SATIN: from services to implementation

M. Karaliopoulos[1], I. Andrikopoulos[2], P. Henrio[3], P. Philippopoulos[4], B. Evans[1]

[1]Centre for Communication Systems Research, University of Surrey, Guildford, GU2 7XH, UK
Email: M.Karaliopoulos@surrey.ac.uk
[2]Space Hellas S.A., Athens, Greece, [3]Alcatel Space Industries, Nanterre, France
[4]OTE Consulting S.A., Athens, Greece

**ABSTRACT**

The advantages of satellites in multicast and broadcast service provision has attracted a lot of interest within the satellite community and motivated a re-examination of the role of mobile satellite systems with regard to terrestrial networks. This paper reports on the approach taken by the IST project SATIN towards the definition of a suitable architecture for the delivery of multicast/broadcast (MC/BC) services, making use of S-UMTS packet mode.

## I. INTRODUCTION

The multimedia concept, strongly embedded within UMTS, introduced a new potential for satellite systems, as the collaborative parts of terrestrial UMTS (T-UMTS) rather than as stand-alone systems. The steady convergence of broadcast technology and mobile satellite systems has brought in front the undoubted advantage of satellites in supporting broadcast and multicast services and has motivated new research ventures in the satellite communications community.

This fact seems to be well established within the Advanced Satellite Mobile Systems Task Force (ASMS TF) [1]. The Digital Multimedia Broadcast (DMB) concept attracts a lot of interest within the European satellite community (and elsewhere) and a quite wide combination of network technologies, radio interfaces and system configurations are candidates for its implementation. To put things into a perspective, SATIN (Satellite UMTS IP-based Network) focuses on one of the available alternatives, featuring reduced terminal complexity as its main advantage.

Having established that the overall system success requires much more than technology availability and respective capabilities, SATIN[1] attempts to determine such a role for satellites in UMTS networks and service delivery, that can lead to a successful business case. In order to achieve this, it was deemed necessary to:

1) Perform a market and business analysis and identify realistic service scenarios for S-UMTS
2) Propose (on the basis of identified service scenarios) an efficient system architecture that is closely integrated with T-UMTS, considering the cost/complexity of the individual system components and spectrum issues.
3) Define the access scheme in a packet mode, namely functions and respective component interfaces.
4) Evaluate the performance of a set of system techniques (i.e. key issues) through simulation.

This paper describes the multicast role within SATIN. The outcome of the first step and the system architecture scenarios selected accordingly are reviewed in section II A short summary of the current standardization effort on multicast/broadcast within 3GPP is given subsequently, before the requirements for transport of MC/BC services in the UMTS network, under two different assumptions about the core network, are discussed in section IV. The respective functionality required at the Satellite Radio Access Network (S-RAN) and the terminal is also identified in the same section. The paper concludes by presenting the decisions made by SATIN regarding the reference architecture to be considered further for the access scheme definition.

## II. SERVICE AND ARCHITECTURE CONSIDERATIONS

The main outcome of the study performed within [2] can be summarized into three major points: first, it confirmed that the only viable market approach for S-UMTS is to target the mass, consumer market rather than competing with exiting systems for the niche markets. Secondly, it showed that multimedia multicast and broadcast services (e.g. streaming services) can be the basis of this mass market, under the condition of a ‘close co-operative’, rather than competitive, relation to the terrestrial UMTS, which will allow S-UMTS to address a wider subscriber base. Thirdly, it demonstrated that it is feasible to build a viable business case upon the ‘MC-BC users’ that will normally use S-UMTS for MC/BC services, even though they move in areas where T-UMTS or 2G(+) networks are available. The required subscriber base was estimated to be lower than the respective base for ‘direct users’, corresponding to the niche markets, but still high enough (1.15millions) to strengthen the importance of the ‘close co-operative’ approach.

Therefore MC/BC type of services were selected as the basis for the service portfolio to minimize potential implications due to spectrum and secure higher Average

[1] http://www.ist-satin.org

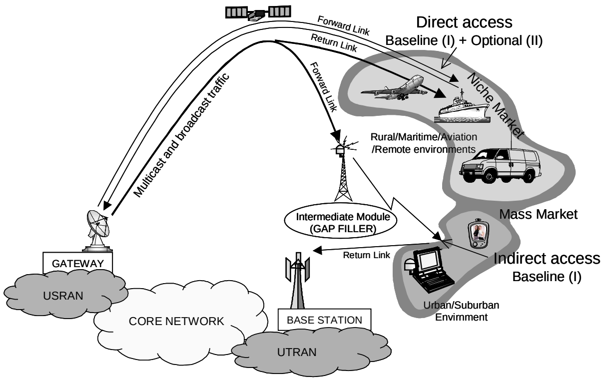


Figure 1: SATIN reference architecture

Revenue per User (ARPU); nevertheless other services implying a considerable market potential are also considered but to a lesser extent.

In order to overcome the inability of satellite-only systems to offer in-building and in-urban areas coverage (where the mass market resides) and support the envisaged, moderate and high bit rate, MC/BC services, it was deemed mandatory within SATIN to introduce a ground component (IMR -Intermediate Module Repeater). The last, but not least, consideration in this service-/market- driven architecture definition was the requirement for a less complicated, cost-effective terminal (the 'user cost metric') and this was yet another point favoring a close integration with T-UMTS.

Two architectural scenarios, a *baseline* and an *optional*, were selected to address the aforementioned requirements (Figure 1). In the baseline scenario, a handheld mobile terminal, receives data through the satellite and/or the intermediate module that features one-way repeater functionality. The satellite path would be the preferred communication link, but if the user's satellite path were blocked, the communication link would be sustained via the IMR stations. The return path is provided via the T-UMTS network (*baseline* case). Such a terminal may well serve a given subset of services (pure broadcast, and non-highly interactive multicast). Alternatively, the terminal may also support direct transmission (in the return path) to the satellite (*optional* case), leading to the more conventional system configuration that allows a stand-alone system to be built at the expense of a more expensive and complex terminal.

## III. MULTICAST AND 3GPP

The broadcast/multicast case in media with inherent broadcast capabilities is a significantly different case from the conventional implementation of broadcast/multicast over networks consisting of point-to-point links. A great deal of the complexity and difficulty arising in multicast support in the wireline networks has to do exactly with the non-broadcast nature of the media. The routing of packets to the end hosts in an efficient manner is not a simple problem. In broadcast media, and even better in the case of satellites, this problem is solved a priori, since all hosts within the earth station (satellite) coverage can be reached in a single-hop. The problem in the wireless broadcast networks, and this mainly but not exclusively corresponds to the multicast case, is rather the engineering of a scheme (subscription registration, activation and respective security associations, group management) that will allow the reception of information only from those that have subscribed to it, while avoiding inefficient use of network resources.

### *A. Multicast in Release 99 and Release 4*

Multicasting and broadcasting are areas where not much has been done so far within 3GPP. Although four categories of point-to-multipoint services had been advertised initially, only two of them are maintained in Release 99 and Release 4 documents.

The first one, the only one standardized within Release '99, is the Cell Broadcast Service (CBS) allowing for low bit-rate data to be transmitted to all subscribers in a set of given cells over a shared broadcast channel [3].

The second one is an optional IP-Multicast service that allows mobile subscribers to receive multicast traffic. The support of this service is not mandatory for the GGSN. The GGSN duplicates the incoming multicast packets and relays them to the already active tunnel endpoint identifiers (TEIDs). These TEIDs are those of MSs that have joined a multicast group. This service, whose name is rather misleading, does not allow for multiple subscribers to share radio or core network resources and as such does not offer any advantages as far as resource utilization within the PLMN and over the radio access network are concerned.

### *B. Multimedia Broadcast Multicast Services*

The standardisation process of point-to-multipoint services carries on within the Release 5 framework. The respective architecture is named Multimedia Broadcast Multicast Services (MBMS) and it seems that it will not be based on CBS, i.e. it does not draw from the Radio Interface for Broadcast/Multicast services adopted in Release 99 [3].

MBMS is split into broadcast and multicast modes. The multicast mode orientation is to make use of IP service platforms to maximise the availability of applications and content so that current and future services can be delivered in a more resource-efficient manner.

Among the general architecture guidelines set in the respective specification [4] are:

- MBMS architecture shall support external data sources in both modes. MBMS shall support both IP multicast and IP unicast sources.
- MBMS architecture should re-use, to the extent possible, existing 3GPP network components and protocol elements thus minimising necessary changes to existing infrastructure and providing a solution based on well-known concepts.
- MBMS shall be a point-to-multipoint bearer service for IP packets in the PS domain.
- MBMS shall be interoperable with IETF IP Multicast.

One difference envisaged between the broadcast and the multicast mode lies in the fact that the latter generally requires a subscription to the multicast subscription group and then activation of the service. On the contrary, for the broadcast mode, it is expected that charging data will not be generated for the end user.

## IV. MULTICAST IN SATIN

The support of IP multicast in SATIN is mainly foreseen in:

- Taking benefit of the advantages of the User Datagram Protocol (UDP) connectionless, datagram service for broadcast/multicast transport of applications and leaving acknowledgement processing at the application level (reliable multicast transport techniques).
- Targeting minimum acknowledgement of multicast transmission and retransmission needs.
- Optimising the content distribution by means of broadcast/multicast data servers and techniques such as web caching and mirroring, that are not necessarily located in the SATIN gateway and perform:
  - Routing to build multicast/broadcast IP streams of multimedia content (use of different multicast addresses, each corresponding to a service offerg to the users in terms of content type and associated quality of service and security requirements) associated with content element segmentation, possibly QoS based routing (terrestrial versus satellite segment), scheduling as well as security features and reliable multicast transport techniques (FEC, retransmission).
  - Content serving to assign a service descriptor to each multimedia content; this descriptor being used all along the distribution chain to perform optimum routing, scheduling, and subsequently filtering, cache management as well as presentation to the user.

The implementation of both these functions will be based on open standards such as those devised within IETF or other fora.

The way multicast will be supported in SATIN (and more generally in any S-UMTS configuration) is heavily dependent on the level of IP penetration in the UMTS Core Network (CN) and its role in the macro-mobility support.

While it is agreed that it is not easy for IP-derived solutions/protocols to cope with the strict requirements of the UMTS micro-mobility functions, hence these functions are left to the native UMTS protocols, there are, macroscopically, two approaches for the UMTS macro-mobility support:

The first is the solution currently implemented, up to Release 5, relying on the conventional SGSN, GGSN nodes and the GTP tunnels throughout the CN till the RAN edges. The second draws heavily from the IP-based solutions and promises better integration with the Internet. The standard GPRS network is replaced by (compressed into) a UTRAN/IP gateway, which is attached to a backbone of routers running pure IPv6, while Mobile IPv6 is charged with the macro-mobility task. The latter approach is strongly expressed within the IST WineGlass project [5] and is considered to be a mandatory step towards the realization of fourth-generation (4G) networks.

In the following, the implications of each one of these approaches regarding multicast support are explored.

### *A. Multicast in an all-IP CN*

The adoption of Mobile IPv6 in the CN makes the application of IP-derived solutions for multicast support more straightforward (or even mandatory):

- Multicast capable routers can be deployed at the CN for more efficient multicast transport
- The Internet Group Management Protocol (IGMP) can/must be used for group management purposes.

Support of IP multicast in SATIN has to address mainly the *scaling* problem; that is the standard IP multicast architecture implies a significant overhead of signalling/control messages, given the number of potential hosts per spot beam. These messages can generally be either multicast routing messages exchanged between the multicast-routing capable entities of the network or IGMP messages. Within the SATIN context the problem is related to the IGMP messages. IGMP capable routers detect the presence of group members by sending IGMP queries, to which hosts answer with IGMP report messages. The messages are timer-driven and may constitute a significant portion of the network load, effectively reducing its available capacity for data traffic.

Nevertheless, there are two features of SATIN (and more generally satellite networks) that have to be noted and can be exploited for a more efficient support of multicast services.

#### *A.1 The tree-like network topology and the 'IGMP proxying' principle*

The aforementioned signalling load and the respective resource consumption can be avoided in certain topologies. This is the main reason why the 'IGMP proxying' (IGMP-based Multicast Forwarding) technique was conceived. The specification of this mechanism is still in a draft state [7] but some of the ideas contained therein seem to fit well the considered multicast scenarios.

With respect to their position in the multicast spanning tree, the router interfaces can be divided into downstream interfaces (DI) and upstream interfaces (UIs).

There can only be one UI for an IGMP proxying device. DIs are in the direction of hosts while UI is in the direction of another router (Figure 2). This differentiation is introduced since, depending on its type, a different role in the protocol is played by the interface.

In the proxying technique, DIs run the so called *router portion* of the IGMP protocol, in other words, on each

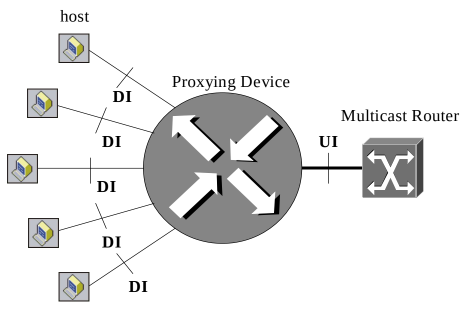


Figure 2: The 'IGMP proxying' interfaces

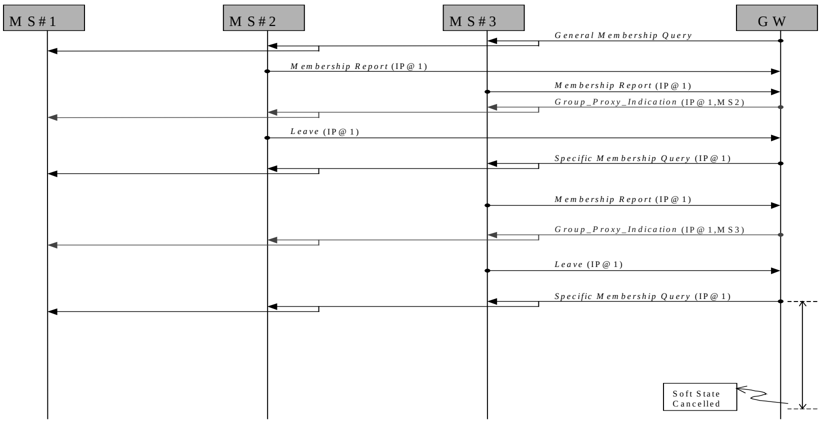


Figure 3: Example of IGMP proxying with overhead reduction

interface, the normal IGMP operations are performed, maintaining in a separate way, a membership database.

These databases are then merged to obtain a *global membership database* that accounts the memberships on each interface.

UI runs the *host portion* of the IGMP protocol, so it has to send IGMP membership reports when it receives a query message, and has to send unsolicited reports or leaves when database changes occur.

As far as the forwarding technique is concerned, when a router (or proxying device) receives a multicast packet, it builds a record in a *forwarding database* consisting of a list of the interfaces (UI and DIs) where there is a subscription to the group except for the interface from which the packet arrives. Then it forwards the packet to those interfaces. This operation can be made simpler if the forwarding database is used as a cache, so that the creation of a record in the database is made once for all the packets belonging to the same group. This simplification comes however at the cost of updating the cache every time the situation in the membership changes.

In SATIN it is the S-RNC (Gateway) and potentially the UTRAN-IP gateway (physically they might be the same) that has to play the role of the proxying device(s).

*A.2 The LAN-like nature of the network*

Rather than implying a strict resemblance with a Local Area Network (LAN), the term "LAN-like" refers to the capability of all the hosts within a beam to receive all transmissions destined for this beam. This capability can be exploited to reduce the number of exchanged IGMP messages over the air interface. Rather than letting every mobile host (MH) send reports back to the gateway, which implements the IGMP querier functionality, one of the multicast group members is elected as the group representative for IGMP proxy[2] functions for the whole group. The other hosts trigger a timer whenever they see a report from the designated group proxy and only send their own report when this timer expires. The underlying principle is that the gateway does not have to be aware of the exact number of MHs participating in a given group but rather whether there is one or more MH(s) in a specific beam so that a copy of the message is forwarded to this beam. The requirement for the *Max Response Time* field is to be higher than the roundtrip time but reasonably low so as to reduce the number of membership report messages sent after the receipt of the *General Membership Query.*

An instance of IGMP message exchanges, when both the aforementioned optimisations are adopted, is shown in Figure 3. The figure illustrates the case of three Mobile Stations (MS) associated with a given Gateway (GW) which acts as querier. The difference in comparison with the fixed broadband access system case lies in that MS terminals are not connected to a CPE (Customer Premises Equipment, a device with layer 3 functionality in this case), which would take on the role of the IGMP proxy for them.

A periodic *General Membership Query* (GMQ) message is broadcast to the cluster by the GW, which contains the selected *Max Response Time*.

Upon receipt of the GMQ, the MS sets a delay timer for each group of which it is member. Timers are set to a random value selected from the range (0, *Max Response*]. In our example, MS#2 and MS#3 wish to receive traffic sent to the multicast address IP@1, while MS#1 has no active memberships. The first *Membership Report* message that is received comes from the MS#2, which, according to the overhead reduction protocol, becomes the elected group proxy.

The GW broadcasts a Network signalling message, namely the *Group_Proxy_Indication* to inform all MSs in the cluster that MS#2 has been elected the group proxy for the address IP@1. From now on, all members of the group IP@1 except MS#2 can suppress membership report and leave messages.

At the end of the session, MS#2 cancels its subscription to group IP@1 by sending a *Leave* message to the GW. As in standard IGMP, the latter sends a *Specific Membership Query* to make sure that no other member of the group is active in the cluster. In our example, MS#3 is the remaining member of the group IP@1, therefore it will send a *Membership Report* to the GW, and will become the new group proxy for address IP@1. When MS#3 - last and single member of the group IP@1 - finally leaves this group there is no reply to the *Specific Membership Query* sent by the GW. When the timer for the subscription to the group IP@1 expires, the GW cancels the relevant soft state.

[2] In the remainder of the paper the term 'group proxy' refers to the IGMP signalling overhead reduction, while the 'IGMP proxy' term refers to the IGMP-based multicast forwarding.

The extra difficulty, when applying the second principle (IGMP signalling overhead reduction technique) in the case of mobile hosts, featuring no proxy device in front of them, is that modifications can no longer be transparent to the end hosts. Hence, it is necessary to modify the IGMP 'client' software at all hosts, while in the fixed satellite systems with end-hosts in a LAN behind a router, it would be enough to modify the latter.
For the *baseline scenario*, where the return link is provided by the T-UMTS, a solution for overcoming the unidirectional nature of the satellite link is provided by the Link Layer Tunneling Mechanism (LLTM), standardized in the IETF UniDirectional Link Routing (UDLR) WG [6].

*A.3 The UDLR LLTM*

In the baseline scenario, it is necessary to come up with a solution to the problems posed to the IGMP by the unidirectional nature of the satellite link. The IGMP, much like the IP routing protocols, has been designed and engineered assuming a bi-directional link. Since this does not exist in the SATIN baseline scenario, it has to be emulated somehow over the T-UMTS link.
Such problems have mainly been addressed in the context of fixed satellite networks, where the unidirectional link is a satellite broadcast link (e.g. DVB-S) and there is a return terrestrial channel (e.g dial-up line, PPP) that allows some form of interaction between the end-user and the provider/network operator. The IETF UDLR WG concluded the first part of its activities with the specification of a link-layer tunnelling mechanism, which effectively allows the emulation of a bi-directional link over a unidirectional link.
Within the SATIN context, UDLR feeder/hub functions are required in the GW and UDLR receiver/host functions are required in the terminals.

*B. Multicast in a GPRS-based CN*

The implementation of multicast in this case seems to be a different case. The main reason for this is the different business paradigm of the two networks, namely the current, best-effort Internet and the UMTS.
In the former, there is intensive, time-based, signalling at the edges of the network, between the hosts and the closest, multicast-capable router because there is no detailed state at the router upon the exact number or addresses of the hosts that receive multicast content. In other words, the only thing required by the router is to know whether there are one or more hosts that want to participate in a multicast session. In order to facilitate this, the end hosts are subject to this frequent IGMP message exchange. In effect, the trade-off between signalling overhead and router complexity is determined in favour of the latter, the underlying assumption being that at the edge of the network the luxury of wasting some bandwidth on additional signalling is feasible.
On the contrary, in a mobile, wireless network like UMTS, there is generally much more information for the end user available at the network nodes. This information is available anyway in order to support the user mobility and AAA (Authentication, Authorization and Accounting) functions.
The additional capability in the GPRS networks is the capability to use this information when 'routing' each packet in pre-established tunnels that are created during the multicast Packet Data Protocol (PDP) contexts. Therefore it is feasible for the SGSN to route traffic between the two access networks (T- and S- UMTS), since it is only necessary to set up the tunnels initially and make the respective bindings. This is not feasible, at least on the basis of the standard datagram routing paradigm, in a WineGlass like, all-IP CN.
Therefore the IGMP-related issues become of less (or even no) relevance in this case, since the information provided by IGMP is 'there' and, most significantly, can be used for routing packets to the users.

## V. CONCLUSIONS

Regarding the multicast implementation, the goal initially pursued in SATIN is smooth integration into UMTS (3GPP Rel.5 is the reference although the MBMS architecture specifications are likely to be included in Rel.6). Therefore the architecture features a satellite RNC interfaced to GPRS backbone with some provision for evolving with the possible IP penetration in the 3GPP core.
For the SATIN baseline architecture the support of prime SATIN Multicast Services (i.e. no real-time interaction of the return and forward links required) favours the following main architecture features:

- Home environment requirements of MBMS stage 1, inter alias Multicast Area.
- IP multicast capability in GGSN.
- Use of GTP (i.e. GTP-U protocol in the User plane and GTP-C signalling), in the terrestrial path and in the satellite path till the satellite RNC**.**

The provision of the multicast services implies the following main phases:

- *Multicast service delivery preparation* (service advertising, subscription, and alerting or discovery).
- *Multicast session initiation* by either the network (push-like services predominating SATIN portfolio of services) or the UE, that encompass PDP context creation and IGMP / IGMP-like exchanges.
- *Multicast PDP context activation* and subsequent Radio bearer creation / adaptation (hence dynamic radio resource management at the satellite gateway).

## REFERENCES


[1] ASMS TF, 'The vision of the Advanced Satellite Mobile Systems Task Force', September 2001.
[2] SATIN project, 'S-UMTS IP specific service requirements', Deliverable No. 2, 2001
[3] UMTS, 'Radio interface for broadcast/multicast services', 3GPP TR 25.925, version 3.3.0, Release 1999.
[4] UMTS, 'MBMS: Architecture and functional description', 3GPP TR 23.846, Release 5, v0.0.1.
[5] L. Dell' Uomo, E. Scarrone, 'A 4G IP-based mobile network enabling QoS mobility management', IST Mobile Communications Summit, Barcelona, September 2001.
[6] IETF UDLR WG, www.ietf.org/html.charters/udlr-charter.html.
[7] B. Fenner, 'IGMP-based multicast forwarding (IGMP Proxying)', Internet draft (work in progress), November 2001.