Multipoint Communications in a Beyond-3G Internetwork

Elias C. Efstathiou and George C. Polyzos
Mobile Multimedia Laboratory, Department of Informatics
Athens University of Economics and Business
Athens 104 34, Greece

Abstract - *We consider the problem of supporting IP-based quasi-reliable mobile multipoint communications in a Beyond-3G internetwork which combines the current Internet (including all connected home users and wired or wireless LANs) with the soon-to-be-deployed 3G networks and the future digital broadcast networks. We summarize existing IETF protocols such as IP multicast, Mobile IP, and Cellular IP, which we believe help in solving the problem. We refer to filtering techniques useful for delivering media to mobile users. We discuss the Mobile IP and IP multicast constraints as well as their inter-operability problems, and we present existing alternatives. We also offer our own perspective by outlining requirements for future mobile multicast protocols; by describing fundamentals of the mobile multicast problem; by mentioning our own approach to the problem, which involves IETF protocols and current trends; and by presenting the Beyond-3G internetwork we envisage.*

1 Introduction

Packet switching technology is quickly replacing the circuit-switched backbones of telecommunication carriers. The Internet is experiencing its biggest expansion yet, with 3G operators worldwide deploying their new IP-based networks. Meanwhile, digital broadcasters are also joining the IP bandwagon, with DVB (Digital Video Broadcasting – the emerging digital TV standard) providing IP support and over 30 Mbps of shared bandwidth. Older technologies that support IP over various link layers include Ethernet networks, wireless IEEE 802.11 networks, as well as PSTN, ISDN, xDSL and Cable connections. Although the "All-IP" goal has not yet been reached, it is clear that we are approaching it fast. Each technology is traditionally associated with a particular service, e.g. bidirectional one-to-one audio (voice) for cellular telephony, unidirectional one-to-many video for DVB. The ideal all-IP internetwork must support each one of these services, integrate them, and "enhance" them with features such as higher interactivity and mobility. We use the term *Beyond-3G internetwork* to describe this model.

The IP suite was originally designed with one-to-one communications in mind. However, the advantages of one-to-many and many-to-many communications are numerous [10, 28] and we will not dwell on them here. Networks that provide native broadcast are better suited to support multipoint communications. Examples, such as Ethernet, 802.11, and satellite, terrestrial or cable broadcast networks, allow transmitted frames to be received efficiently by all hosts in the local subnet. Other network technologies offer point-to-point links in the *user plane* [19], while supporting broadcast in the *control plane* (GSM/GPRS, UMTS). Some technologies (PSTN, ISDN, DSL) only provide native support for point-to-point communications.

IP currently supports efficient, local, network layer, many-to-many communications, wherever link level broadcast is provided. However, extending this support beyond a local subnet is non-trivial. Also, to support many-to-many communications in an internetwork topology that changes dynamically while IP hosts (and even entire subnets) are moving is even more challenging. This paper focuses on *IP-based quasi-reliable mobile multipoint communications*. IP multicast [10], which is the basis of our analysis, is a best effort multipoint communication protocol. Reliable, sequenced delivery extensions do exist [13]. However, applications don't always require this reliability, since they can adapt in their own more advanced ways. This, in particular, is the case with *streaming media* services (especially real-time media).

In a "Beyond-3G internetwork," IP hosts should be able to change their network point of attachment with minimal disruption of ongoing communications. We can assume that one or more of the following may occur while an IP host is moving: (1) the host may temporarily disconnect from the network, (2) the host's IP address may change, (3) in a wireless scenario, a horizontal or vertical handoff may occur, and (4) handoffs may occur between cells belonging to different administrations. In this paper we offer our perspective on the issues involved in combining multipoint communication capability with host mobility in a Beyond-3G environment. We do this assuming the existence of a fixed, wireline routing infrastructure. Also, whenever we discuss IP protocols, such as IP multicast and Mobile IP, our focus is on IPv4. We are aware that the move to IPv6 is happening, albeit slower than expected, so IPv6 is taken into account when it helps with our analysis.

The remainder of this paper is organized as follows. Section 2 shows how early Internet assumptions influenced present day protocols and made it harder to solve the mobility problem. Section 3 outlines the IP multicast protocol. Section 4 presents two IP mobility protocols, Mobile IP and the less well-known Cellular IP. Section 5 mentions the usefulness of transcoding filters to the problem. Section 6 deals with mobility and IP multicast coexisting. In Section 7, we present our own perspective on the problem; we mention requirements for future protocols; we outline our own solution; and we describe the Beyond-3G environment we envisage. Section 8 concludes the paper.

2 Fixed, Wireline Internet Legacy

The Internet was originally built to support packet based data communications among pairs of stationary IP hosts. This assumption influenced the design of most protocols in the IP suite [26]. Although the overall design of the Internet protocols is "layered," in practice, it is often found that interlayer dependencies exist and become apparent when attempting to port existing services to newer network technologies.

We will use TCP as an example of the dependencies, design limitations and assumptions that cause service disruption when a TCP application like Telnet or FTP is used in a mobile environment:

Interlayer dependency - Each TCP connection is identified by a unique pair of *sockets*, with each socket being a <*Host_IP_address*, *TCP_port*> pair. The TCP connection ID is, therefore, the 4-tuple <*Client_IP_addr*, *Client_TCP_Port*, *Server_IP_addr, Server_TCP_port*>. It is obvious that lower layer identifiers (IP addresses) are used as part of higher layer ones. If either end-host changes IP address, the TCP connection will break because, not only does TCP rely on lower layer identifiers, but also, it assumes that they will not change during the lifetime of a TCP connection [26].

Design limitations - "Classic" TCP (RFC 793, including the 1988 Jacobson refinements) is designed to interpret packet losses as a sign of congestion along the router path between the two end-hosts. TCP adapts by lowering its rate and allowing router queues to drain [17]. However, if the packet loss was caused by the increased BER (Bit Error Rate) of a wireless link, lowering the rate only decreases TCP performance. The underlying assumption is that TCP hosts are stationary and communicate over wired links with low BER [30].

3 IP Multicast

IP multicast is a many-to-many communication protocol. The *host group* service model defines its requirements: let H be the set of all IP hosts. Let E_G be a subset of H. Set E_G forms a *multicast group with address G*, if and only if, (1) members of H may join and leave E_G at any time, and (2) members of H can communicate unidirectionally with all members of E_G, using only identifier G.

The definition above does not specify how to satisfy these requirements. As IP hosts may belong to more than one group, mechanisms are needed to (1) associate hosts with groups, (2) track group membership, and (3) route data to all group members [28]. Ideally, IP multicast packet delivery should emulate the exactly-once semantics of packet delivery in a traditional (non-switched) Ethernet [2]. Since, in an Ethernet, all transmitted frames are received by all attached interfaces, exactly-once delivery comes for free. In an Ethernet, the only additional mechanism needed is the mapping of multicast group identifiers to MAC addresses. In this way, Ethernet hosts know which frames to process and which to discard. (Switched Ethernets must also provide support for multicast frames.)

The IETF set aside all legal IPv4 class D addresses to be used as multicast group identifiers, underlining the fact that IP multicast is a network layer protocol. The IPv4 class D address *224.0.0.1* always identifies the link-local *all-hosts* multicast group, which all multicast IP hosts are required to be a part of [9]. The mechanisms needed to implement multicast packet delivery are divided into *global* and *local* [28].

3.1 Global Mechanisms

With IP multicast, the burden of delivering multiple copies of packets falls on the network. The sender need only transmit one copy of each packet addressed to a multicast group identifier. Multicast enabled routers will forward the packet as needed, replicating it onto more than one of their outgoing interfaces only when paths towards the destination group members diverge. The global multicast routing protocols deliver a group's packets to multicast routers that have expressed interest in receiving packets for the particular group. This interest, in turn, is triggered by hosts in the router's local IP subnets that declare their wish to join the specific group.

Proposed multicast routing protocols include the *Distance Vector Multicast Routing Protocol* (DVMRP), *Multicast Open Shortest Path First* (MOSPF), *Core Based Trees* (CBT), *Protocol Independent Multicast-Sparse Mode* (PIM-SM), and *Protocol Independent Multicast-Dense Mode* (PIM-DM). All these protocols attempt to build a multicast delivery tree of routers for each multicast group. DVMRP and MOSPF are less scalable (they construct one tree per sender per group) than CBT or PIM (they construct one tree per group, which senders share). CBT and PIM are also independent of the underlying unicast routing protocols. PIM-SM is optimized for *sparse* multicast groups and PIM-DM for *dense* multicast groups. PIM works by choosing a *Rendezvous Point* (RP) when it constructs the multicast delivery tree for a group, where multicast senders can "meet" multicast receivers. With all protocols, there is an associated *graft* delay when a multicast router joins an existing tree because the tree has to be adjusted. All these protocols implicitly assume stationary hosts.

3.2 Local Mechanisms

The global protocols we described above are concerned with multicast senders, multicast routers and multicast groups, but not with the individual multicast listeners. These listener hosts are "hidden" behind their local multicast router. The router can be thought of as the interface between the local and the global mechanism [28]. Based on a local group management protocol, this router builds a list with all the different multicast groups its hosts have joined (on each one of the IP subnets it may serve). Only this aggregate list is exposed to the global mechanism.

The local group protocols are the *Internet Group Management Protocol* (IGMP) [12] for IPv4 and the *Multicast Listener Discovery* protocol (MLD – derived from IGMPv2) [11] for IPv6. IGMP assumes the existence of link-level native broadcast (e.g. Ethernet) and is designed around the soft-state principle which traditionally leads to robust Internet protocols. IGMP works as follows: every *querying_period* the multicast router responsible for a local subnet sends out queries to the all-hosts group. Its objective is to refresh the group list it exposes to the global mechanism. All local listeners receive the query and respond (after a small *random_report_delay*) with a *group_report*, one for each of the groups they participate in. Each report is actually sent to the multicast address for the group reported so that all interested local listeners may learn of this fact.

The soft state principle in IGMP dictates that queries are router-initiated and that no explicit *leave_group* message is needed when a listener leaves a group: the router will discover it in the next query cycle. Extensions for unsolicited listener messages exist [12]. For example, to lower the *join delay* (the local equivalent of the global graft delay), a host joining a group not already present on the local link may send an unsolicited *group_report*. If not, it could very well wait up to *querying_period* + *random_report_delay* + *graft_delay* before multicast packets start arriving.

4 IP Mobility Protocols

4.1 Mobile IP

The goal of *Mobile IP* (M-IP) [24] is to allow internetwork host mobility in a manner transparent to the transport layer. This can be achieved by assigning a permanent IP address to every *mobile host* (MH). This IP address is called the MH's *home address*. If the MH changes its point of attachment and moves to a *foreign* IP subnet, a *Correspondent Host* (CH) may still deliver packets to the MH using the MH's home address as the destination. This is possible because a new entity at the MH's home subnet - a router known as the *Home Agent* (HA)- intercepts these packets and *tunnels* them (reroutes them using *IP-in-IP encapsulation* [23]) to the MH's current network address.

The MH's current network address is known to the HA as the result of a registration procedure: when the MH first moves to the foreign subnet, it communicates back to its assigned HA its new *Care-of Address* (CoA). This new address is either a *co-located* CoA or a *Foreign Agent* (FA) CoA. In the co-located case, this new address is usually obtained through DHCP on the foreign subnet. In the FA case, this address corresponds to a special router - the FA - that resides on the visited subnet. In both cases, this address is communicated back to the HA (either by the MH or the FA), where a new address *binding* is created, associating the MH's home address with its current CoA.

In the co-located case, *reverse tunnelling* (decapsulation of packets) is carried out by the MH, whereas in the FA case, reverse tunnelling is the responsibility of the FA which will then forward each packet to the MH (FAs and MHs are assumed to have link-layer connectivity). In both cases, IP-in-IP from the HA will deliver a packet with an outer destination address matching the CoA and an inner destination address matching the MH's permanent home address.

The aforementioned mechanisms describe how CHs can always send packets to MHs. For the reverse path, the standard IP mechanism can be used: the MH will send packets addressed to the CH's address and place its home address in the source address field. This "trick" has no adverse effect since standard unicast IP routing depends only on the destination address.

The M-IP communication mechanism results in an inefficiency called *triangle routing* [28]. Extensions [25] to the original M-IP RFC allow CHs to make address bindings of their own which subsequently gives them the ability to discover the MH's new address and route packets directly to the MH bypassing its HA and home subnet. This feature can only be used if the CHs are also M-IP-aware.

M-IP is built around the soft-state principle which dictates that all registrations with the HA (as well as all local registrations of MHs with FAs) should have an associated lifetime after which they expire. MHs need to refresh the bindings periodically. There is no explicit deregistration procedure. When an MH moves to yet another subnet or when it disconnects completely, the old FA (if one existed) and the HA in charge will soon learn of this fact. FAs maintain a list of all the visiting MHs they serve and HAs maintain a list containing the bindings of all the MHs for which they are responsible.

4.2 Cellular IP

M-IP was designed with "slow" mobility in mind. The protocol incurs several delays, the most important one being the time MHs take to discover an FA in the foreign subnet, register with it, acquire a CoA, and register with their HA. This *agent discovery time* can be significant. If a moving MH traverses many *microcells*, each one controlled by a different FA (co-located with the cell's base station), the signaling required during every handoff may inhibit the normal reception of user packets. In the case of a fast moving host, the *mobility assumption* [3] may be violated: the total registration delay may be more than the time an MH spends inside a cell. As a result, no user packets will find the time to get rerouted to the MH.

Cellular IP (C-IP) [6] is a best effort *micromobility* protocol designed with campus-wide, partially-overlapping microcells in mind. C-IP can rely on M-IP for *macromobility* support (for example, an MH arriving at a campus internetwork and registering this fact with its HA). For intracampus mobility and local handoff control, C-IP takes over and uses its own internal network of packet forwarders (which may be co-located with base stations). As far as the MH's HA is concerned the whole campus internetwork is controlled by only one FA (M-IP has to be used in FA mode for this to work).

A C-IP *gateway* can be co-located with M-IP FAs. This gateway is the interface between the M-IP and C-IP routing infrastructures. Base stations within a C-IP network are fixed, interconnected IP forwarders with at least one wireless interface. To forward packets to MHs, they employ non-standard IP routing techniques: MHs are only identified by their M-IP home addresses and, unlike M-IP, no tunnelling is employed [6]. The route learning mechanism resembles the MAC bridge autolearning mechanism and it follows the soft state principle: routes have to be refreshed periodically. This happens whenever MHs send regular user data packets towards the gateway.

Certain key notions at the heart of C-IP are based on existing mechanisms used in cellular telephony networks. These include the notions of *active* and *idle* nodes and the notion of *paging* [6]. To minimize signaling overhead, MHs only update their location information when they are active, i.e. engaged in communication. Since C-IP is connectionless, MHs are deemed idle after a certain period of inactivity. When they are idle, the fixed base station infrastructure has no knowledge of their whereabouts if the soft state routes have also expired. When new incoming packets have to be delivered to an MH, paging is used to discover its exact cell: the C-IP forwarders flood all their outgoing interfaces (unless a cached route entry exists), except the one they received the packet from. C-IP supports smoother handoffs by allowing MHs to send *route_update* packets as soon as they detect that they have moved to a different cell.

5 Filters and Media Stream Quality

There exist software components, collectively called *media filters* or *transcoders*, which, given an input media stream, can produce a different, but similar, output media stream. Their purpose is to help maintain a level of *Percepted Quality of Service* [14] as the media stream traverses many internetwork links of varying bandwidth. These filters can be divided into *smart* and *simple* [14]. Smart filters do not have to decode the media stream (e.g. an MPEG-2 video stream) to a raw format and then re-encode it. Rather, they operate on the encoded stream directly. Smart filters differ from simple ones in that they require more processing power, but, generally, because they take advantage of a previous encoding stage they are also faster and produce better perceived output than their simpler counterparts.

Filters can be used to reduce the bit rate of a media stream before it is injected into bandwidth limited links [22]. Filters can support multipoint wireless communications by allowing wireless receivers that cannot receive the source stream at its original bit rate to receive a "cut-down" version of it. Usually, this means lower resolution, fewer frames per second and smaller frame-size. It can also mean lower color-depth, less smooth movement and the tweaking of a variety of parameters that depend on the media encoding format.

Layered coding and multiresolution layered coding [15] can be used to separate a media stream into more than one streams, each carrying, progressively, more information. Listeners can receive as many of the streams their incoming bandwidth allows. This can be used in conjunction with filtering. A novel proposal also suggests using different multicast groups to transmit each substream [20]. Listeners may then "tune into" as many groups as possible.

Filters become very useful to multimedia communications in wireless environments. A natural location to place such filters is at the boundary between the wired and the wireless part of the internetwork. There, they can modify the passing stream based on actual measurements of the conditions on the wireless link. In a multicast scenario, different receivers may require different filters. In the case where many receivers at the same multicast subtree request the same transcoding, filters can propagate upstream (towards the subtree root, closer to the stream source), where they can combine into one [21]. In the general case, an *active* router (one that allows third-party code uploading and execution) is a natural location for these mobile filters to reside in.

6 Combining IP Multicast and Mobility

The IP multicast protocol was designed to bring multipoint communication capability to the Internet. At the same time, Mobile IP and Cellular IP were designed to allow transport-layer-transparent mobility. Trying to integrate the protocols into a coherent IP framework for multicast mobility exposes several fundamental issues. Existing approaches [16, 18, 24, 27] offer some functionality, but further testing is needed before the scalability of each proposed solution can be judged.

6.1 "Fixed Host" Assumption - IP Addressing Issues

We already mentioned that the "fixed host" assumption influenced the design of several protocols in the IP suite. When considering mobility, we need to take into account that mobile devices can be very different from fixed hosts connected to an Ethernet network: (1) limited battery life dictates that unnecessary operations should be avoided, so, protocols that rely on constant monitoring of network traffic are impractical, (2) protocols that assume high bandwidth, low latency connectivity may become inoperable, (3) protocols that assume low BER and no disconnections are faced with a hostile wireless environment, (4) the notion of handing-off the connection to a different cell, neighboring or overlaid, is non-existent in the wired world, (5) protocols that rely on link-local broadcast capability are not always easy to port to a wireless environment because the techniques used in several wireless networks cannot be easily modified to support link-level multicast. GSM's time division multiple access, CDMA's power control, and the fact that, in most current cellular systems, the base station is not an IP-layer router but a lower layer entity, make such attempts difficult.

IPv4 address shortage is also a problem now. GPRS operators usually have to rely on *Network Address Translation* (NAT) and are forced to assign private IP addresses (usually over PPP) to the mobile devices that register with their networks. All the usual NAT limitations affect IP multicasting although work-arounds do exist [31]. Also, some cellular operators have their own interpretation of 2.5G and 3G multicasting [1] which is more limited than the IP-based multicasting envisaged for a Beyond-3G environment.

6.2 IETF Mobile IP Multicast Support

The current IETF proposed standard for Mobile IPv4, RFC 3220 [24], devotes no more than a single page to multicast packet routing in conjunction with Mobile IP. Two methods are mentioned, which are referred to by [8] as *Remote Subscription* (MIP-RS) and *Bi-directional Tunnelled Multicast* (MIP-BT). Both methods are only relevant when MHs are visiting a foreign subnet. We only describe them with M-IP operating in FA mode, since this is the preferred mode of operation both for wireless communications and for interoperability with C-IP.

Remote subscription - This may be used only when a multicast router is present in the visited subnet (this router may be co-located with the FA). The MH can use IGMP to subscribe to any number of groups using this router.

Bi-directional tunnelled multicast - Another option is for the MH to setup a bi-directional tunnel to its HA (this HA must also be a multicast router). The HA will join groups on the MH's behalf. The tunnel is used to send IGMP messages and receive multicast packets. In this case, double encapsulation is required: first, the HA has to encapsulate the multicast packet inside another packet addressed to the MH's home address, and then, encapsulate once more as specified by Mobile IP. This means that the MH must be able to decapsulate the multicast packet even if it uses an FA for the standard Mobile IP decapsulation procedure.

Both approaches have their disadvantages. MIP-RS requires that the MH re-subscribes to a potentially large number of multicast groups after every subnet move. This delay will cause packet losses for the MH and will require rearrangements of the multicast tree with each move. MIP-RS assumes that a multicast router will exist in the visited subnet. Also, MIP-RS generally causes *get-ahead* and *lag-behind* effects (terms borrowed from [6]). These may happen when MHs register with a new multicast router and resubscribe to multicast groups. Because of the complex nature of a multicast router tree, some edge routers may lag behind others in their reception of multicast packets. One solution to the get-ahead problem is for the host to accept the loss of some packets and allow higher layer protocols to deal with it. More sophisticated solutions involve buffering at the multicast routers. The lag-behind problem may be solved locally at the host, where higher layer protocols can discard already received packets, or at the router if somehow the router is signaled to drop packets before forwarding them to its internal subnet (which is important if the subnet is a wireless, bandwidth limited one).

With MIP-BT, the multicast tree does not have to be rearranged, since the MH's HA will be a stationary multicast receiver. Also, if the HA buffers multicast packets, no get-ahead or lag-behind problems will exist for the MH. MIP-BT, however, is associated with the three following unwanted phenomena: (1) If an HA supports many MHs, all visiting the same foreign subnet and all subscribing to the same groups, then multiple point-to-point channels will have to be setup between the same HA and the foreign subnet, each one carrying the same information. The duplicate packets will strain the (potentially wireless and bandwidth limited) visited subnet. (2) If, somehow, the HA detects this and sends only one packet copy per foreign subnet, then a problem known as *tunnel convergence* may still occur if *many* HAs have MHs, all visiting the same foreign subnet and all subscribing to the same group. This will also lead to unnecessary packet duplication. (3) Another way for unwanted duplicate packets to appear on the foreign subnet is if a host, local to that subnet, already subscribes to one or more of the groups some MHs wish to subscribe to. Since MHs will receive multicast packets encapsulated in unicast packets addressed to their home address (see MIP-BT description), it would not be easy for the FA or some multicast router in the foreign subnet to detect this duplication and stop the unnecessary transmission.

6.3 Extensions to the IETF Approach

The following examples are alternatives or extensions to the basic Mobile IP and IP multicast interoperability approaches proposed by the IETF. Each one of these examples improves on the basic mechanisms but we believe that further research is needed.

Mobile Multicast (MoM) *Protocol*

MoM [8, 16] is based on MIP-BT and its key extension is the use of a *Designated Multicast Service Provider* (DMSP). DMSPs attempt to solve the tunnel convergence problem (see section 6.2). A DMSP for a given multicast group is an HA chosen by the visited subnet's FA out of the many HAs that forward packets for the specific group to the visited subnet. MoM supports choosing more than one DMSP for redundancy and it also supports DMSP handoff, which is necessary when a DMSP has no more MHs of its own in the visited subnet that require packet tunnelling.

MoM-specific algorithms are executed everytime (1) MHs arrive at a foreign subnet, (2) MHs depart from a foreign subnet, (3) MH registrations with the FA time out, (4) unicast or multicast packets from the HAs arrive at the FA. The subnet's FA keeps track of HAs, MHs, and multicast groups, so it always has enough information in order to choose a DMSP and to perform DMSP handoffs whenever required. MoM is the most cited alternative to MIP-BT and MIP-RS.

Mobile Multicast with Routing Optimization (MMROP)

MMROP [18] is based on MIP-RS and its key extension is the introduction of the *Mobility Agent* (MA) entity, which attempts to solve the get-ahead problem due to handoffs (see section 6.2). This is done to "ensure routing efficiency and no packet losses from roaming" [18]. MAs are FAs that route missing packets (via tunneling) to neighboring subnets. MMROP works as follows: let FA1 and FA2 be an MH's old and new foreign agents respectively. Let's assume the MH was subscribed to group G through FA1. The MH will attempt to resubscribe to group G through FA2, at which point FA2 will start buffering packets. Upon joining the new subnet, the MH will look at the sequence numbers of the packets for G and decide whether or not it should ask for cached packets from FA2. (MMROP assumes packets are somehow numbered.) If FA2 cannot supply these packets, it will request FA1 to continue transmitting packets to the MH through a tunnel between FA1 and FA2, until FA2 and the MH are synchronized, at which point FA2 will start delivering packets to the MH through its own multicast subscription.

Constraint Tree Migration Scheme (CTMS)

CTMS [7] is an attempt to design a new global multicast routing protocol that would improve on CBT [5] when it comes to highly dynamic multicast configurations, such as those found when multicast listeners are mobile. CTMS "automatically [migrates multicast trees] to better ones, while maintaining the QoS guarantees specified by mobile users" [7]. CTMS uses fewer resources per multicast tree and, as a result, packet losses due to reconfigurations and join delays are reduced. CTMS is a good alternative to existing mobile routing protocols but its adoption will be difficult, considering most multicast routers still run DVMRP.

Multicast Scheme for Wireless Networks (MobiCast)

MobiCast [27] is based on MIP-RS and its key extension is the introduction of the *Domain Foreign Agent* (DFA) which serves many small adjacent wireless cells. A hierarchy is introduced, with small cells being organized into one *Dynamic Virtual Macrocell* (DVM). Micromobility is thus hidden from the global multicast mechanism, which does not require reconfiguring when handoffs occur within the same DVM.

6.4 IGMP Mobility Support and IGMP Assumptions

IGMP was designed with Ethernet networks in mind. Its basic functionality assumes link-level native multicast. Also, its soft-state timers require that multicast listeners repeatedly announce all the multicast groups to which they are subscribed. This happens every time the subnet's designated multicast router issues a query. We present IGMP's two main problems with respect to mobility support:

IGMP is not suitable for point-to-point links – If the local multicast router is not only connected to an Ethernet subnet, but also has interfaces that connect to *point-to-point* links, then IGMP queries have to be issued to each one of these interfaces [28]. The IGMP replies will not be heard by all other participants unless the router specifically multi-unicasts them to every one of the point-to-point links it supports. This increases delay, data traffic and state information needed at the router. Instead of just the group list which the global multicast mechanism requires (see section 3), the multicast router must record per host information [28]. Indeed, many cellular networks currently offer only point-to-point links for user data. The PPP protocol that is usually used on these networks (as well as on most Internet home connection technologies) to support IP packet transfer does not have the same semantics nor does it use the same techniques as the shared Ethernet protocol.

IGMP is not suitable for mobile hosts – We already mentioned (section 6.1) that mobile hosts don't have the luxury to constantly monitor network traffic. That would place a burden on their battery and processing power. Also, mobile hosts should not be forced to keep resending unnecessary information if this can be avoided. The IGMP soft state timers, although they are simple to implement and they contribute to IGMP's robustness, force the hosts to keep repeating the same data for as long as each one of their group subscriptions is active. A solution proposed in [29] suggests using explicit *join_group* and *leave_group* messages which would require from hosts (which are only multicast listeners) to send significantly fewer packets. This scheme may interoperate with "traditional" IGMP.

7 Our Perspective on Mobile Multicast

7.1 Multicast Semantics and Mobility

An extension to the issues raised in section 6.1 is that multicast semantics need to be reexamined in the presence of mobility. We present a simple example based on ideas raised on [8] that exposes this problem. Let's assume X and Y are two Ethernet-based IP subnets. Let's also assume that MHs with home addresses in subnet X are designated Xi and that MHs with home addresses in subnet Y are designated Yi. Some Xi MHs are visiting subnet Y and some Yi MHs are visiting subnet X. If an IP multicast packet addressed to the link-local all-hosts group 224.0.0.1 is directed towards subnet X, then there are three possibilities, according to [8], about what should happen: (1) the packet should only be delivered to MHs in subnet X, regardless of whether they belong to the Xi or Yi set, (2) the packet should only be delivered to all MHs in subnet X that belong to the Xi set, (3) the packet should only be delivered to all MHs belonging to the Xi set, irrespective of their current location.

Obviously, there exist techniques for each one of the three possibilities. But which one is semantically correct? There is no right or wrong answer. This depends on the service protocol that originated the link-local all-hosts multicast. We refer to three service examples, each one assuming a different interpretation: (1) an advertisement for public network printers that are present on a specific floor, (2) an advertisement for available high-quality color photo printers, to be used as part of subnet X's core business, (3) an administrator's advertisement, describing the new authentication procedure for subnet X's SMTP server.

Currently, there is no clear IPv4 mechanism that would help mark the advertisement packet and allow the mobility protocol to make an informed decision. IPv6 can, however, differentiate between *link local*, *site local* and *organizational local* multicast scopes.

7.2 Mobile Multicast Requirements

Some general issues that should affect all mobile multicast solutions are the following:

Significant vs non-significant moves - Let MH be a host subscribed to a set of multicast groups. If the move of MH to a new subnet causes the subnet's multicast router to subscribe to new multicast groups, then the move is *significant*. Otherwise, if due to existing subscriptions packets addressed to the MH's set of groups are already being transmitted on that subnet, then the move is *non-significant*. Mobile multicast protocols will have to differentiate quickly between these two types of moves. Ideally, non-significant moves must have no effect at all on the global mechanisms. All the necessary information to identify non-significant moves can be found within the local subnet and at the subnet's multicast router. In addition, significant moves should appear identical to non-significant moves from a user's perspective.

Multicast packet buffering – Although, in theory, IP multicast is a best effort protocol, in practice, if mobile multicast schemes are to work efficiently, packet buffering should happen at the IP multicast layer. For example, with MIP-BT, the HA may buffer packets before tunnelling them to the MH. This is necessary in order to achieve smooth handover when the MH moves to a new subnet and reestablishes the bi-directional tunnel. With MIP-RS the situation is more complicated. Depending on how MIP-RS is used, both the FA and a local multicast router on the visited subnet are candidates for buffering packets. The main problem with buffering is the following: buffer packets until when? In a wireless environment, with significant and non-significant moves, disconnections due to handoffs, disconnections due to physical layer problems and disconnections due to user intent, it will not be easy to judge how long buffering should go on. When multiple entities buffer simultaneously complexity increases. Soft-state timers must adapt based on a number of parameters and on input from both lower and higher protocol layers.

Mobile subnets - Ships, planes, trains, and even cars can be thought of as mobile subnets, each one with several local mobile hosts. Dealing with these as one logical entity will greatly assist routing protocols, ease tunnel convergence problems and minimize state information kept throughout the internetwork.

Roaming - The problem of global roaming between different administrations is very difficult to solve, even for point-to-point communications (the same applies to vertical handoffs). Even if mobile multicast routing protocols are simulated and tested, true mobile multipoint communications will ultimately rely on sophisticated authentication mechanisms and pricing schemes.

7.3 Cellular IP and Mobile Multicast

We are conducting research on integrating efficient multicast mechanisms to Cellular IP (C-IP) and using C-IP in conjunction with Mobile IP (M-IP) in a hierarchical manner similar to the MobiCast scheme (see section 6.3) but based on C-IP and M-IP interoperability ideas developed within the IETF. In our scheme, a MobiCast DFA is a C-IP gateway and a MobiCast DVM is a C-IP subnet. Introducing multicast support to C-IP is relatively straightforward considering that the basic C-IP forwarding mechanism is simple. However, more simulations are needed in order to test the scalability of our proposed architecture.

We chose C-IP and a method based on MIP-RS because we took into account not only the current evolution of the Mobile IP and Cellular IP specifications but also the real network configurations that people deploy. These include campus-wide 802.11 internetworks, UMTS cells and the future DVB-T macrocells. It is our position that the MIP-BT based tunneling schemes (although friendlier to current multicast routing protocols) are simply not scalable enough. The current *Content Delivery Network* (CDN) trends strengthen our belief that content and services need to be pushed as much as possible to the edges of the Internet and that hosts should first try to exploit whatever resources they have available in their immediate environment (i.e. follow the *locality* principle) and only when this fails should hosts try to access more distant resources.

7.4 The Internet Beyond-3G

Having completed our discussion of the basic protocol interoperability issues, we describe the Beyond-3G internetwork that we envisage will support improved versions of all the aforementioned protocols. It will be made up of many, superimposed, cellular technologies. If we ignore the satellite component, which is usually the least cost-effective way to create wireless cells, we still have many promising technologies with which to build this hierarchical cell structure. DVB-T may be used for metropolitan size (1–100 Km) macrocells, 3G systems, such as UMTS, may be used for neighborhood size cells, and the 802.11 variations, wherever available, for local microcells. Although DVB-T is unidirectional extensions do exist that provide bidirectional functionality. Even without these extensions, 3G networks are perfectly suited to provide the necessary return channel [4].

DVB-T, with 5-30 Mbps of shared bandwidth and excellent support for mobility within a DVB-T macrocell, could become the technology of choice for delivering IP multicast traffic, bypassing most of the problems we described in previous sections. Going down the cellular hierarchy, we would have overlapping 3G cells (running versions of Mobile IP) and, then, localized 802.11 networks (running versions of Cellular IP). This structure needs to be augmented with umbrella cells that can handle traffic for fast moving receivers in trains or in cars. Mobile subnets could also be supported by installing Cellular IP gateways on ships and on trains, which would support micromobility within the moving subnets.

Devices with multiple interfaces are starting to appear. Common offerings include support for 802.11, combined with either DVB or GPRS. As long as IP and Mobile IP are accepted standards we can expect that all future devices and networks will support them. For unicast applications, alternatives to TCP will be used most of the time and improved variations of TCP will provide backwards compatibility. Transcoding functions and filter mobility protocols will be standardized and they will be put to use by all wireless network providers.

8 Summary and Conclusion

We examined the multipoint communications problem assuming a Beyond-3G environment with mobile hosts, a fixed network infrastructure, and a best effort network layer. We focused on reusing IETF protocols where possible. IP multicast and Mobile IP were obvious choices. We showed how existing transcoding techniques may be used. However, we saw that the proposed IETF IP multicast and Mobile IP interoperability solutions are not perfect and that they require extensions. We presented a number of additions and alternatives to the basic scheme. We mentioned our own approach involving Cellular IP and we offered our perspective on additional multicast mobility issues. Finally, we described the Beyond-3G environment we envisage.

Still a lot of functionality needs to be added to the basic IP multicast and Mobile IP offerings before infra-structureless networks, strong reliability and security are also supported. Ultimately, global roaming and pricing agreements will be necessary to complete this vision of the ideal Beyond-3G Internet.

References

[1] 3GPP TSG-RAN Working Group 2, “Definitions of Multicast service and requirements,” *Public Document, www.3gpp-.org/ftp/tsg_ran/WG2_RL2/TSGR2_02/Docs/pdfs/R299075.pdf*, March 1999.

[2] A. Acharya, A. Bakre, and B.R. Badrinath, “IP Multicast Extensions for Mobile Internetworking,” *Proceedings of 1996 IEEE INFOCOM*, pp. 67-74, San Francisco, CA, March 1996.

[3] A. Acharya and B.R. Badrinath, “A framework for delivering multicast messages in networks with mobile hosts,” *ACM/Baltzer Journal of Mobile Networks and Applications*, 1(2):199-219, October 1996.

[4] Ad hoc Group DVB-UMTS, “The Convergence of Broadcast & Telecomms Platforms,” *Public Document, www.dvb.org*, March 2001.

[5] A. Ballardie, J. Crowcroft, and P. Francis, “Core based Trees (CBT) – An architecture for scalable inter-domain multicast routing,” *Computer Communications Review*, 23(4):85-95, October 1993. (Proceedings of the ACM SIGCOMM’93.)

[6] A.T. Campbell, J. Gomez, and A.G. Valko, “An Overview of Cellular IP,” *IEEE WCNC*, New Orleans, September 1999.

[7] K. Chen, N. Huang, and B. Li, “CTMS: A novel constrained tree migration scheme for multicast services in generic wireless systems,” *IEEE JSAC.,* 19:1998-2014, October 2001.

[8] V. Chikarmane, C.L. Williamson, R.B. Bunt, W. Mackrell, “Multicast Support for Mobile Hosts Using Mobile IP: Design Issues and Proposed Architecture,” *ACM/Baltzer Mobile Networks and Applications*, 3(4):365-379, Jan. 1999.

[9] S. Deering, “Host Extensions for IP Multicasting,” *RFC 1112*, August 1989.

[10] S. Deering, “Multicast Routing in a Datagram Internetwork,” *Ph.D. Thesis, Department of Computer Science, Stanford University*, 1991.

[11] S. Deering, W. Fenner, and B. Haberman, “Multicast Listener Discovery for IPv6,” *RFC 2710*, October 1999.

[12] W. Fenner, “Internet Group Management Protocol, Version 2,” *Internet RFC 2236*, November 1997.

[13] S. Floyd, V. Jacobson, and S. McCanne, “A Reliable Multicast Framework for Light-weight Sessions and Application Level Framing,” *IEEE/ACM Transactions on Networking*, 1995.

[14] G. Gardikis, E. Pallis, A. Kourtis, “Beyond 3G: A Multi-Service Broadband Wireless Network with Bandwidth Optimisation,” *IST Project “Multi-Services Management Wireless Network with Bandwidth Optimisation,” Public Document*, *www.openmux.-com/mambo/Public/Beyond 3G.pdf*, 2000.

[15] J.K. Han and G.C. Polyzos, “Multi-Resolution Layered Coding for Real-Time Image Transmission: Architectural and Error Control Considerations,” *Real-Time Imaging*, 4(4):275-298, Academic Press, August 1998.

[16] T. Harrison, C.L. Williamson, W.L. Mackrell, and R.B. Bunt, “Mobile Multicast (MoM) Protocol: Multicast Support for Mobile Hosts,” *Proceedings ACM MOBICOM*, Budapest, Hungary, September 1997.

[17] V. Jacobson, “Congestion Avoidance and Control,” *Proceedings ACM SIGCOMM*, CA, USA, August 1988.

[18] J. Lai, W. Liao, M. Jiang, and C. Ke, “Mobile Multicast with Routing Optimization for Recipient Mobility,” *Proceedings IEEE ICC2001,* pp. 1340 - 1344, June 2001.

[19] Y. Lin, “A Multicast Mechanism for Mobile Networks,” *IEEE Communications Letters*, 5(11), November 2001.

[20] S. McCanne, V. Jacobson, and M. Vetterli, “Receiver-driven Layered Multicast,” *Proceedings ACM SIGCOMM*, Stanford, CA, August 1996.

[21] J.C. Pasquale, G.C. Polyzos, E.W. Anderson, and V.P. Kompel-la, “Filter Propagation in Dissemination Trees: Trading off Bandwidth with Processing in Continuous Media Networks,” *NOSSDAV, D. Lecture Notes in Computer Science 846*, Shepherd et al., eds., Springer-Verlag, Berlin, Germany, pp. 259-268, 1994.

[22] J.C. Pasquale, G.C. Polyzos, E.W. Anderson, and V.P. Kompella, “The Multimedia Multicast Channel,” *Internet-working: Research and Experience*, 4:151-162, 1994.

[23] C. Perkins, “IP Encapsulation within IP,” *Internet RFC 2003*, October 1996.

[24] C. Perkins, editor, “IP Mobility Support for IPv4,” *Internet RFC 3220*, January 2002.

[25] C. Perkins and D.B. Johnson, “Route Optimization in Mobile IP,” *Internet Draft draft-ietf-mobileip-optim-11*, September 2001.

[26] A.C. Snoeren, H. Balakrishnan, and M.F. Kaashoek, “Reconsidering Internet Mobility,” *Proceedings HotOS-VIII*, May 2001.

[27] C.L. Tan and S. Pink, “MobiCast: A Multicast Scheme for Wireless Networks,” *ACM MONET*, 5(4):259-271, 2000.

[28] G. Xylomenos and G.C. Polyzos, “IP Multicast for Mobile Hosts,” *IEEE Communications*, 35(1):54-58, January 1997.

[29] G. Xylomenos and G.C. Polyzos, “IP Multicast Group Management for Point-to-Point Local Distribution,” *Computer Communications,* 21(18):1645-1654, Elsevier Science, December 1998.

[30] G. Xylomenos, G.C. Polyzos, P. Mahonen, and M. Saaranen, “TCP Performance Issues over Wireless Links,” *IEEE Communications*, 39(4):52-59, April 2001.

[31] “How Does Multicast NAT Work on Cisco Routers?” *Technical Note*, *www.cisco.com/warp/public/105/multicast_nat.html*, 2001.