An IP Multicast Handoff Scheme with Focus on IGMP Sourced Latency

Buyurman Baykal, Özgür Eser Akdemir, and Özgür Barış Akan
Electrical and Electronics Engineering Department,
Middle East Technical University, 06531 Ankara/Turkey
{buyurman, e077206, e121467}@metu.edu.tr

Abstract

This paper focuses on the latency due to IGMP during inter-subnet handoffs in Mobile IP multicast networks and proposes a new method for this purpose. Pre-informing the neighbor subnets about the expected mobile hosts under the shadow cluster concept and triggering the preparation phase in advance are the fundamentals of the proposal. Performance improvement achieved through the method is discussed in the paper with the particular emphasis on QoS critical applications.

1 Introduction

Various schemes have been employed to enable IP-based networks support mobility. These schemes differ in design according to the type of communication, being either from one source to one destination, i.e., IP unicast, or multiple combinations of sources and destinations, i.e., IP multicast, as well as network topology, application properties and performance expectations. The scope of this paper is the latency experienced during inter-subnet handoff mechanism in mobile IP multicast applications, mainly and specifically due to Internet Group Management Protocol, IGMP. So far studies focus on the efficiency improvements in many tasks involved in the overall handoff process, whereas this paper focuses on the latencies that stems from the necessity to wait for a periodic query to be able to reply it. The paper is, therefore, organized in the following manner. First mobile IP multicast is reviewed with an overall view of IGMP operation. Then the latency sources accompanying the IGMP queries and handoff preparation phase are pointed out. Next, handoff signaling decomposition and EVL employment are presented as new tools for the proposal. Performance considerations and expected improvements are discussed in the following section before we conclude the paper with future work considerations.

2 Preliminary Review

2.1 IP Multicast

In a mobile IP-based network, point to multipoint communication is one of the challenging tasks in terms of mobility support and proper routing. Thanks to the mechanism, known as multicasting, mobile hosts have the ability of joining/leaving dynamically created or modified multicast groups. This process must be paid attention, especially for the construction of multicast routing trees to be able to reflect the changes in group membership information in each subnet. Moreover, mobility brings with further concerns in these issues, such as the necessity for a seamless transition between subnets.

Mobile IP multicast handoff schemes utilize two essential protocols for group management and routing. Multicast routing protocols deal with construction and update of multicast trees to reflect the changes in routing paths, whereas the group management protocols track the changes in the membership information of multicast groups [1]. Two protocols work cooperatively to make the network optimized for the up-to-date traffic conditions and to provide the mobile hosts with seamless handoff operation.

2.2 IGMP Operation

In IGMP [2],[3], one of the most used group management protocols, a multicast router per subnet sends an IGMP Host Membership Query to all stations on its subnet to determine if any hosts on a local subnet belong to a multicast group, and asks them to report back on their host groups membership states.Upon the determination of host group membership information, the multicast tree construction algorithm is invoked in the case of any change in the membership and location changes. In this way, further modifications are performed throughout the network, such as proper configuration of multicast router filter settings and end-to-end QoS messaging for new routes to be used in the new networking environment.

Periodic queries, propagated by the IP subnet router and received by the hosts, are replied as soon as they are detected. The frequency of periodic queries is determined to be at optimum value, which does not overload the network and does not cause the router miss the current membership information by being too high and low, respectively. Despite configurability of group membership query frequency, the periods of IGMP queries is the part that results in severe handoff latencies.

0-7803-7600-5/02/$17.00©2002 IEEE

3 Handoff Latency Sources

3.1 Necessity to Wait for an IGMP Query

After an inter-subnet handoff, a mobile host, which is member of some multicast groups, can continue to receive the multicast traffic in its new subnet only if there exists other active participants of those groups in that network. Otherwise, the mobile host simply waits for the query message sent by the IP multicast router to be able to inform the network about its needs as a reply message. The operation is shown in Fig. 1.

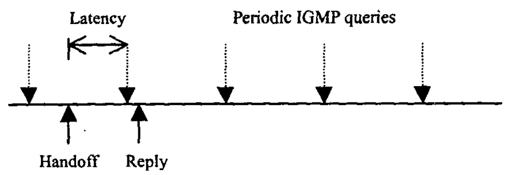

Fig. 1. Illustration of IGMP queries and handoff operation

It is obvious that handoff timing and the duration of IGMP query periods determine the latency experienced during handoff. If the mobile host enters the new subnet just after a missed query, it has to wait nearly a full IGMP query period to be able to inform the new subnet its attendance.

3.2 Handoff Preparation Phase Tasks

When the existence of a mobile host with the need of new multicast group traffic flow is discovered, the necessity of the construction of a multicast routing tree branch arises. This is performed by informing upstream routers and invoking multicast routing algorithms. Building up a new tree branch or tearing down of an existing one may bring with some latency. Moreover, for applications with strict QoS requirements, new subnet router might have to trigger an end-to-end QoS messaging [4] to be able to detect most suitable paths and routers for the requested multicast traffic flow.

Therefore, another latency source is the one accompanying the handoff related tasks performed in the network as illustrated in Fig. 2, where IGMP latency is also reminded.

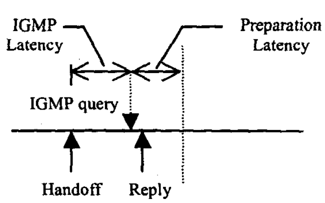

Fig. 2. Latencies experienced during IP multicast handoff

Latencies experienced during inter-subnet IP multicast handoff should be minimized for a seamless operation and continuous service access. In the next chapter, we will present the tools for our proposal, namely handoff signaling decomposition and the expected visitor list employment.

4 Tools of The Proposal

An inter-subnet handoff scenario, as shown in Fig. 3, will be explored including a mobile host in the membership of an IP multicast group and strict QoS requirements of its application.

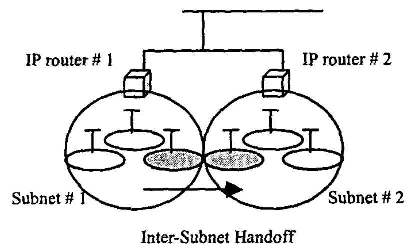

Fig. 3. Inter-Subnet Handoff in Cellular IP

4.1 Decomposed Handoff Mechanism

In conventional IP multicast handoff scheme, handoff-signaling process can be outlined as follows and shown in Fig. 4.

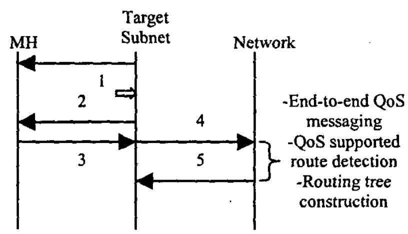

Fig. 4. Conventional IP Multicast Handoff Signaling Mechanism

1. Mobile host enters a new subnet, where no other members of the same multicast group exists, and therefore the requested traffic flow is off-line.
2. New subnet sends periodic queries to learn whether there exists a new mobile entry.
3. Upon the reception of query, mobile host replies with parameters indicating its needs, such as QoS requirements and multicast membership information.
4. New subnet, with these information in hand, constructs/modifies the routing table, determines QoS supported routes after an end-to-end QoS messaging, e.g., Path and Resv messages in the RSVP case.

5. Only after the network is prepared for the mobile host's needs, requested traffic flow is on and handoff is completed.

It would be a clever idea to make possible candidate subnets about the probable handoff requests of mobile hosts in advance. In this way, those subnets may be prepared for the incoming hosts. These preparations include some time consuming processes and they can be classified as follows.

- Detection of the optimum paths for the multicast routing protocol by searching in many alternatives.
- Triggering the end-to-end QoS messaging, especially for multicast applications employing strict QoS requirements.
- Construction of new branches for multicast tree.

These preparations may be performed earlier than the actual handoff instant as follows and shown in Fig. 5.

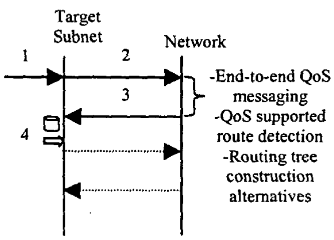

Fig. 5. Decomposed IP Multicast Handoff Signaling Mechanism

1. The probability of the mobile host to request a handoff in near future is computed by the current subnet based on the speed and direction knowledge of the mobile. When it is decided that it is the correct time, current subnet (Home Agent) informs candidate target subnets about the mobile host's possible handoff request with a parameter list indicating its QoS requirements and multicast membership information.
2. Target subnet, learning mobile-specific parameters directly via current subnet, not via IGMP queries and replies, uses them to trigger the network-side preparations in advance. These include end-to-end QoS messaging to detect the most suitable QoS supported paths and determination of alternative multicast routing paths.
3. After the network-side preparations are completed, two alternatives arise;
   a. multicast traffic flow is provided within the target subnet for the incoming mobile host.
   b. preparation results are stored as records to be utilized later on.
4. Whenever the mobile host enters target subnet, it will either find the multicast traffic flow already there (Step 3.a) or wait for the first IGMP query to utilize the information gained through step 2 (Step 3.b).

Alternatives mentioned in steps 3 and 4 differ in their network management approach. Case (a) indicates a more readily handoff handling with eliminated both IGMP and preparation latencies but it also indicates resource consumption before it is actually needed. Making the multicast traffic available for groups the incoming mobile is member of before that mobile host actually enters the new subnet is the reason for resource consumption but it is the worthwhile cost of fast handoff mechanism. Case (b), on the other hand, requires waiting for the next IGMP query at the actual handoff instant, taking only the advantage of time gained through the preprocessing of handoff tasks but not consuming any resources until they are really needed.

Therefore handoff signaling is said to be decomposed into two parts. First one is the transfer of mobile-specific parameters from the current subnet to the most candidate neighboring subnets (Step 1). Since intra-subnet handoffs are likely to be less painful compared to inter-subnet ones, focusing on the adjacent base stations under the control of different subnets is the key point. Accurate direction tracking of the mobile host makes it possible to determine the next possible locations, like the shadow cluster approach [5], and to learn whether they are within the control of other subnet routers. Decomposition of the handoff signaling into two parts and triggering the network-side preparations in advance provides us with the opportunity to get rid of the need to wait for an IGMP query to transfer the handoff parameters. Possible candidate subnets learn these parameters directly from the current subnet and triggers the network-side preparations earlier than the actual handoff instant as explained through steps 2 to 3.

4.2 Expected Visitor List (EVL) Employment

Studying case (b) in the former section, it is seen that results of the handoff preprocessing (alternative routing path determination, awareness of the QoS-support of possible routers) should be stored as records at each subnet routers. Since each subnet preprocesses the handoff tasks of expected mobiles, or visitors, we will name these records as expected visitor entries.

Expected visitor entries are prepared and kept for cell pairs, each of which belongs to different subnets. Therefore, each subnet router keeps a list of expected visitor lists for each of its neighbor cells that are under the control of different subnet routers. Each entry is aimed to be kept updated in terms of validity of the pre-handoff task results. Validity may be subject to disturbance whenever an event occurs in the target subnet or in the rest of the network that makes it necessary to re-process the handoff related tasks. As an example, a change in the QoS requirements of a mobile host in the target subnet certainly disturbs the validity of an expected visitor entry.

Hence, this disturbance results in the re-evaluation necessity of the network-side handoff preparations to detect possibly new reliable routes and update the QoS-support knowledge of each router that will be involved in the communication path, which will be constructed in near future as a result of a handoff request.

A typical expected visitor entry structure is shown in Fig. 6 with mobile-specific parameters on the left branch and results of the network side handoff preparations on the right one.

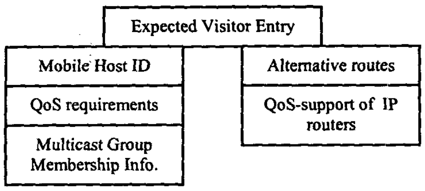

Fig. 6. A typical Expected Visitor Entry Structure with mobile specific parameters and results of pre-handoff tasks

5 Performance Improvements

Removal of the necessity of waiting for the IGMP query results in the faster handoff. A similar study exists in [2] that focuses on reducing the IGMP sourced latency by utilizing mobile agents working as inter-subnet tools that behave as proxies to reply queries in advance before the handoff. In our proposal, without the need for implementing additional entities as proxies, the target subnet is directly informed about the incoming mobile and requested multicast traffic flow is made available before the arrival of the host. This means a direct performance improvement and corresponds to 125 seconds gain at most (default IGMP query interval setting) in handoff latency. This latency is experienced in conventional IGMP handoff scheme if the mobile host enters the new subnet just after a missed query. Considering those lucky mobile host arrivals that can find a query on entrance to the new subnet, it can be said that a mobile host experiences, on the average, around 60 seconds latency due to the necessity of waiting a query. Therefore, the proposal saves that time, which is a valuable amount of gain in latency.

Furthermore, besides the time saved by the removal of the necessity of waiting for the IGMP queries, network-side preparations are invoked earlier than the actual handoff instant. This improves the handoff mechanism since most time consuming processes can be completed long before the possible handoff requests and either requested multicast traffic flow is made available or preparation phase results are stored as records at subnet routers to be utilized later on. Handoff preprocessing is especially important for mobile applications with strict QoS requirements since the end-to-end QoS messaging and determination of QoS supported paths can be really time consuming that contributes to the latency.

However, it is obvious that IGMP-sourced latency is much more higher than the second one, preparation phase sourced latency. This is true since the multicast tree construction and end-to-end QoS messaging latencies may vary depending on the distances between source and target routers, network load and application characteristics that may result in figures that are not worth to compare with IGMP-sourced latency figure, average value of which remains to be as high as mentioned earlier.

6 Conclusions

In an inter-subnet IP multicast handoff scenario with IGMP, implemented as the group management protocol, main sources of latencies are explored in the paper. Using two approaches as the new tools, namely handoff signaling decomposition and the expected visitor list employment, these latencies are aimed to be avoided. By decomposing handoff signaling, necessity to wait for an IGMP query is eliminated, saving us a considerable amount of time. Choosing one of the alternatives arising then, either multicast traffic flow is activated as soon as the network-side preparations are completed or the results of these preparations are stored in units called expected visitor entries. It is the first alternative that provides us with large gains in handoff latency. Second alternative is more suitable for the cases where resources are required to be consumed with up most care.

Every router from the first common one to the one at base station are also involved in this expected visitor entry processing. Although this might seem a drawback of method as it brings processing overhead for intermediate routers, due to the low cost computational resources, this might be tolerable.

References

[1] M. Ramalho, "Intra- and Inter-Domain Multicast Routing Protocols: A Survey and Taxonomy", *IEEE Comm. Surveys,* First Quarter 2000, Vol. 3, No. 1.

[2] J. Wu, "Seamless IP Multicast Receiver Mobility Support", http://www.globecom.net/ietf/draft/draft-jiang-msa-00.html, Internet Draft, April 2000.

[3] V. Johnson, M. Johnson, "How IP Multicast Works", IP Multicast White Paper Series, Stardust Technologies Inc.,http://www.ipmulticast.com/community/whitepapers /howipmcworks.html, 1997.

[4] QoS Forum Publications, "QoS Protocols & Architectures", White Paper, http://www.qosforum.com/ tech_resources.htm, 1999.

[5] D. A. Levine, I. F. Akyildiz, and M. Naghshineh, "A Resource Estimation and Call Admission Algorithm for Wireless Multimedia Networks Using the Shadow Cluster Concept", *IEEE/ACM Trans. Networking,* Vol. 5, No. 1, pp. 1-12, February 1997.