Chapter 8

Handoff optimization for multicast streaming

In this chapter, I propose few optimization techniques that expedite the delivery of multicast stream during handoff in a hierarchically scoped multicast architecture. First, I propose a hierarchically scoped multicast content distribution network, describe the functional components of the architecture and their implementation, introduce the optimization techniques to reduce the join and leave latencies for multicast traffic and finally compare the performance results in the prototype testbed from both optimized and non-optimized handoffs. The previous chapters have focused on fast handoff techniques for unicast traffic. However, in this chapter, I apply some of the optimization techniques that were discussed in Chapter 5 to provide fast delivery for multicast traffic in a hierarchically scoped multicast environment.

8.1 Summary of key contribution and indicative results

Currently, multicast-based content distribution systems lack flexible features such as local and global program management and automatic advertisement insertion. These systems also do not support fast-handoff when the mobile moves between the subnets.
Unlike unicast traffic, multicast traffic is receiver oriented. A mobile receiving multicast traffic is subjected to handoff delay and associated media interruption due to multicast join latency during its movement between layer 2 access points or layer 3 subnets. Multicast join latency is contributed due to periodic IGMP (Internet Group Management Protocol) router query interval and random amount of time the client waits before it can send IGMP client report. This join latency can be as large as 2 minutes in duration and disrupts the streaming media during the mobile’s movement.

I proposed and implemented a hierarchical scope-based multicast streaming architecture that enables local and global program management and real-time advertisement insertion using RTCP(Real Time Control Protocol)-based feedback control information.

In order to reduce the handoff latency for multicast join, I proposed both proactive and reactive triggering techniques. As part of the reactive mechanism, I developed an application layer triggering technique that sends unsolicited RTCP join to join the multicast tree after the mobile hands off to the new network instead of network layer IGMP. The server that receives the RTCP join in turn uses IGMP report to join the upstream router.

As part of the proactive technique, I proposed an application layer proxy and multicast address announcer so that the local server can join the multicast tree on behalf of the mobile as the mobile is impending to handover to the new network. While the multicast proxy and the server join the upstream router using IGMP, the mobile triggers the multicast stream by using RTCP Join to multicast proxy.

A hierarchical scope-based architecture provides the ability to manage local and global program by using local servers in the content distribution network. By using the feedback signal such as RTCP, my proposed mechanism provides the ability to control the advertisement duration without relying on any additional signaling. By using an application layer triggering technique such as RTCP, the mobile does not
need to depend upon layer 3 IGMP router query interval nor does it depend upon multicast support in the kernel. Having the ability to trigger multicast streaming during mobile’s configuration process, the mobile optimizes the operations in parallel. Compared to traditional unoptimized multicast handoff approaches, my proposed proactive optimization techniques can reduce the handover latency by a factor of 10 when the probability of presence of a multicast group is low. Proposed proactive and parallel triggering techniques perform better by a factor 4 compared to the proposed reactive techniques when the probability of presence of multicast group is low (e.g., 0.2).

In the rest of the chapter, I describe the details of the hierarchical scope-based multicast architecture, elaborate the proposed mechanisms that allow local, global program management, advertisement insertion. I describe the experimental testbed where I implemented the architecture and the fast-handoff mechanisms that I proposed. I also compare the results of my proposed fast-handoff mechanisms with that of non-optimized systems.

8.2 Introduction

CDN (Content Distribution Network) distributes the contents from the origin server to the replica servers that are situated closer to the end clients. The replica servers in a CDN store a very selective set of content and only the requests for that set of content are served by the CDN. This mechanism provides reduced access delay for any specific content and consumes lower bandwidth in the core of the network. There are a few commercial content distribution networks (CDN), namely Akamai [Aka], Digital Island [Dig] and Edgecast [Edg] that distribute information from many news media, namely CNN and New York Times. Figure 8.1 shows a sample content distribution network and shows how the local affiliates distribute the global program and local advertisement to the end users.
8.8  Concluding remarks

The proposed hierarchical scope-based multicast streaming architecture provides local control while distributing content over the Internet by using local proxies at the edges of the network. These proxies receive the multicast traffic on a global multicast address and transmits it on a locally scoped multicast address. Having the ability to use locally scoped addresses will alleviate the global multicast address allocation problem or overlapping problem for the multicast receivers in the neighboring networks. Use of the local proxies and real-time localized advertisement insertion mechanism using RTCP feedback provides a good way to manage local program and global program based on the demography and interest of the local community. Application layer triggering techniques avoid the Join delays contributed by IGMP-based router query reports that may take up to 2 minutes during a mobile’s movement across subnets.

Although proxy-assisted proactive technique provides better performance compared to the application layer reactive technique, based on the movement pattern of the mobile, one can either use proxy-assisted proactive multicast or application layer reactive triggering technique or combination of both. For example, a proactive multicasting technique would work the best when the mobile can discover the neighboring
networks ahead of time and can predict the target network it may move to, such as a vehicle moving in a high-way. In a city like environment where the movement of the mobile cannot be predicted easily, application layer reactive triggering techniques will work better. On the other hand, even if the multicast join during mobile’s configuration provides comparable performance as proactive triggering, one needs to extend the configuration protocol such as DHCP and it is also essential that the local streaming server is equipped with DHCP server functionality. Thus, based on a specific performance requirement, network topology, and mobile’s movement pattern, a mobile can adopt either of the proposed optimization techniques.