On Supporting Handoff Management for Multi-Source Video Streaming in Mobile Communication Systems

Tarik Taleb*, Tomoyuki Nakamura, and Kazuo Hashimoto
Graduate School of Information Sciences, Tohoku University
*talebtarik@ieee.org

***Abstract*— Video streaming to mobile users is gaining momentum within the communities of both industrial and academic researchers. In a previous research work, the authors proposed a multi-source streaming method for video streaming to mobile users. The focus was on video fragmentation and packet scheduling to avoid packet reordering and packet redundancy. As a continuation to the work, this paper presents a handoff management method to support users' mobility and to guarantee continuous and smooth playback of video data for users while they are on move. For performance evaluation, some simulation results are presented.**

I. Introduction

The way people use the Internet is rapidly changing from a mere browsing of the Internet to peer-to-peer (P2P) networking and real-time multimedia streaming. Advanced multimedia services have thus become the de-facto service bundle for both network operators and service providers. Among multimedia services, streaming services (both browser-based and P2P-based) are strongly emerging on the telecommunications scene and are conquering most of today's Internet traffic. The popularity of YouTube is a strong indication for the rapidly growing success of streaming services.

The recent and on-going advances in portable computing technologies, coupled with the need for ubiquitous infrastructures, have given birth to a number of new mobile communication systems with significantly broadband bandwidth features (e.g., 100Mbps to 1Gbps as download bandwidth in case of 4G networks and large scale IEEE 802.11-based networks). To efficiently exploit the unique features of these networks, there is a need for a killer application. Being the primary time-spending activity of almost all Internet users and the most desirable service, streaming high quality video contents to mobile users is indeed an excellent candidate to accelerate the popularity of these emerging mobile networks.

For multimedia streaming (not based on multicast), four different approaches can be envisioned:

- Single-source single-path streaming
- Single-source multi-path streaming [1]
- Multi-source single-path streaming [2]
- Multi-source multi-path streaming

For a thorough discussion on the advantages and pitfalls of each approach, interested reader is referred to [3]. In this paper, we consider the delivery of multimedia applications from multiple sources, with replicated video contents, to mobile users via single paths (third approach). In a previous research work conducted by the authors [3], we developed a novel multi-source streaming strategy with the following features:

- Efficient rate allocation mechanism that ensures efficient and fair use of network resources
- Efficient video fragmentation mechanism that prevents redundant transmissions of data packets
- Context-aware server selection mechanism that ensures fast data playback for users
- Accurate video packet scheduling mechanism that prevents packet reordering and ensures smooth playback

As a continuation of the work, in this paper we present a handoff management method that can support users' mobility and guarantee seamless streaming of video data to mobile users.

It should be noted that while there have been few attempts in using multi-source streaming in wireless local networks [4], to the best knowledge of the authors, no previous research work has considered the multi-source streaming concept in the context of next-generation mobile communication systems. We are thus not aware of any handoff management mechanism that is specifically tailored to support multi-source streaming in mobile systems. We believe that the findings in this paper may help in the realization of interesting systems such as a reliable and fast mobile blogging system [5] where mobile users can access different streaming servers to upload or download informative video clips about particular areas of interests. The approach can also play a key role in the construction of content centric networks [6].

The remainder of this paper is structured as follows. Section II highlights some research work pertaining to handoff management, and multi-path and multi-source streaming. Section III presents the key components of the envisioned network architecture. Section IV describes the envisioned handoff management scheme, specifically designed to support multi-source streaming in the considered network architecture. Section V portrays the simulation environment and discusses the results. The paper concludes in Section VI.

II. Related Work

As there has been no prior research work on multi-source streaming in mobile networks, this section describes the main post-standard improvements that have been devised in recent literature "to realize multi-source streaming" and "to support users' mobility in mobile networks" in two separate subsections.

978-1-4244-2413-9/08/$25.00 ©2008 IEEE

A. Multi-source streaming

Streaming services have been the focus of researchers in both industry and academia. The traditional method of streaming consists in the streaming of video contents from a single server to a client via a single route. For the sake of high reliability, researchers have investigated simultaneous use of multiple independent routes for communication. Adequate data transmission protocols have been devised (e.g, Parallel TCP (pTCP) [7] and Stream Control Transmission Protocol (SCTP) [8]) and analytical models are developed [1]. To cope with the issue of packet reordering, new scheduling mechanisms have been also designed [9].

While the multi-path approach increases the throughput, it does not solve the issue of start-up delay. As a remedy to this issue, Content Delivery Network (CDN) operators (e.g., Akamai Inc.) considered the deployment of servers at the network edges. In this way, users are served from nearby servers, the service delay is shortened and the overall network congestion can be minimized. However, to increase the system's scalability, other approaches have suggested the involvement of multiple edge servers in the data streaming. In such approaches, instead of connecting to a single streaming server, a client connects to several video servers to receive different, yet complementary, packets via different paths with uncorrelated loss processes.

The work in [2] proposes a receiver-driven protocol for simultaneous video streaming from multiple senders with forward error correction (FEC) for packet-switched networks. The approach employs a rate allocation algorithm that determines the sending rate for each server and a packet partitioning algorithm that fragments the video stream into several sub-streams, each delivered by a different server via a different path. Whilst the approach considered the case of wired clients, its use in mobile communication systems is not practical due to a number of reasons related to the operations incorporated in the approach. Firstly, the approach assumes constant Round Trip Time (RTT) between the endpoints. This assumption does not hold in case of mobile communication systems as users are on the move. Secondly, the approach is receiver-driven, in other words, most operations of the approach are initiated by the receiver side. This incurs some computation load at the receiver, such as the exchange of a number of signaling messages with the servers and their processing load. The receiver also monitors the dynamics of each route to each server in order to reallocate the data transmission rates among servers. Applying such an approach to mobile users will definitely drain up the scarce energy of mobile terminals. Furthermore, a receiver-driven approach usually makes selfish decisions as receivers do not take into account the efficiency of the overall network resources and fairness in their usage. Another issue pertains to the number of signaling messages. Indeed, by exchanging signaling messages between each client and servers, a storm of signaling messages will be generated when the number of subscribers increases. This shall affect the scalability of servers and waste the network resources. In [3], the authors described a multi-source streaming approach that is specifically tailored to mobile communication systems. As a remedy to the above mentioned issues, the authors substitute the receiver-driven approach by the introduction of newly defined network elements (dubbed decision makers or domain managers) that compute the sending rates of each server on behalf of all clients while maintaining an efficient and fair utilization of network resources.

B. Handoff management

To ensure seamless communication, efficient mobility management schemes are required. There are two types of mobility schemes: end-to-end based and network-infrastructure based. In the former, the mobility issue is resolved by adding adequate enhancements to end hosts and keeping the network unchanged. Notable examples of such schemes are SCTP and the session initiation protocol (SIP) [10]. Via the introduction of new states, TCP-R [11] and Migrate [12] are two other end-to-end mobility management schemes that enable the end-to-end handling of TCP connections. A major drawback of these schemes is that they do not represent a complete end-to-end mobility solution for various applications and their performance is limited under numerous mobility scenarios (e.g., simultaneous movement).

For network-infrastructure based mobility management techniques, they can be classified into two categories: Micro-mobility and Macro-mobility. In the former, handoffs are handled locally without any involvement of Home Agents (HAs). Notable examples are Cellular IP [13] and Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [14]. Cellular IP is specifically designed to support handoff for frequently moving hosts. It is applied on a local level and can inter-work with Mobile IP (MIP) to support mobility among Cellular IP networks. The HAWAII protocol divides the network into hierarchies based on domains. The functioning of HAWAII hinges on the assumption that users' mobility is local to domains. For each host, the HA and any Correspondent Node (CN) are unaware of the node's mobility within the host domain. Each domain has a gateway, called the domain router, and each host has an IP address and a home domain. In HAWAII, host based forwarding entries are installed in gateways using a set of specialized path setup schemes. These entries help to reduce both the data path disruptions and the number of binding updates. A major credit of micro-mobility management techniques consists in their reduction of handoff signaling delays.

In macro-mobility, when a mobile node roams to a different network area, the node solicits for a new Care-of-Address (CoA). A Binding Update (BU) message is then sent to the HA. The major issue with macro-mobility pertains to the significant handoff signaling delays for users roaming far away from their home networks. These delays disrupt active connections each time a handoff to a new attachment point of the network is performed. To reduce handoff-signaling delays in macro-mobility, a large body of prior work was proposed. The central theme in these pioneering studies pertains to the adoption of hierarchical management strategies using local agents. Hierarchical MIPv6 (HMIPv6) [15] and TeleMIP for Cellular IP [16] are notable examples.

To cope with packet losses that may occur during handoffs due to the broken data path from the source to the destination, a set of mobility management techniques has been proposed in recent literature. They can be classified into two categories: *caching-based* and *smooth handoff* techniques. In the first category, when a handoff occurs, the old Access Router (AR) caches and forwards the packets to the new AR based on a request to forward the packets. The most pioneering example that uses this technique is Fast Handovers Mobile IP [17]. In the second category, packets are routed to multiple nearby ARs around the mobile node to ensure delivery of the packets to the node. For instance, multi-path smooth handoff scheme [19] and multicast mobility support [20] use this technique.

As a hybrid of end-to-end based and network-infrastructure based mobility schemes, Guo *et al*. proposed in [21] a mobility management scheme for roaming across heterogeneous wireless networks. The concept consists of a connection manager that makes handoff decision and virtual connectivity that performs the mobility management. The handoff decision is made based on an intelligent sensing of both MAC and physical layers. This sensing operation detects the network conditions (e.g., network type, signal strength, bandwidth availability and expected delay) and bases the handoff decision on that. The concept of virtual connectivity consists of three major operations: i) peer negotiation during which two end hosts agree on items that will be required for a secure and accurate mobility management, ii) local connection translation which defines a unique identifier of the connection (i.e., not a global unique home address), unchanged during the connection time, iii) application level subscription/notification service administrated by a server that is responsible for direct notification of changes in the address of nodes among their corresponding peers. This is to cope with the limitations of existing end-to-end mobility management approaches in handling simultaneous movement.

III. Architecture Description

The architecture and its major components are conceptually depicted in Fig. 1. The figure portrays a number of wireless network domains interconnected to a number of streaming servers via a backbone network (e.g., Internet). Each wireless domain is formed according to the geographical proximity and the density of end-users. It consists of the coverage areas of a number of access points (APs) linked to the backbone network via a gateway (GW). Each wireless domain is administrated by a domain manager (DM, referred to as decision maker as well) and an authentication authorization accounting (AAA) server. The latter is used to verify whether mobile users are authorized to access the wireless network. Whereas, the former carries out the whole service management. The domain managers are assumed to acquire the ability to manage a continuous and high request arrival rate, above all, while meeting real-time demands. They are also assumed to have knowledge on the

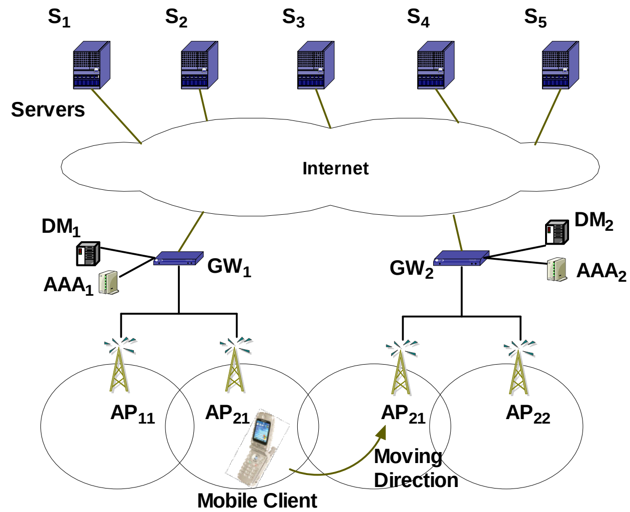

Fig. 1. The envisioned architecture for multi-source video streaming to mobile users.

entire network topology and server contents. As in [3], some of the major operations of domain managers consist in the selection of adequate servers for each mobile user roaming in their domains, estimation of RTT and real time assessment of bandwidth availability of each path to each server, computation of the streaming rate of each server for each mobile user based on the estimated RTT and route bandwidth, and periodic notification of these streaming rates to the servers. To guarantee fast inter-domain handoff management for mobile users, exchange of profile information on users between neighboring DMs is also envisioned.

On the other side of the architecture, a number of servers with replicated video contents are deployed. Whilst at the time being, it might be seen not practical or costly to replicate the same video contents in many servers, it will become highly required in the near future. Indeed, the traffic generated from streaming services is already dominating more than 48% of the entire Internet's traffic. As the number of both users and the provided services are increasing, a large scale service will not be possible unless network operators increase their resources by deploying more servers. In the absence of such a strategy, many users will not be able to access the services and the lucky ones may get the quality of their services degraded during the service. This shall inevitably affect the revenues of the network operators or the service providers. The policy of deploying many servers with duplicate contents has been widely accepted by several major incorporations (e.g., Akamai, Bandai networks). Servers are assumed to have the ability to stream temporarily scattered video packets. For this purpose, they use the information sent via control packets from the domain managers and run the same program to split the contents of a video data [3].

At the user side, terminals are assumed to have a sufficient memory for packet buffering to deal with packet reordering and the associated jitter. They are also assumed to have the ability to simultaneously receive video data from multiple servers. To ensure efficient streaming of video data to users, it is assumed that there is no bandwidth bottleneck at the last hop. This assumption still holds given the recent advances in wireless communication technologies. Indeed, 4G mobile users will be able to download data at rates as high as 100Mbps to till 1Gbps. The video playback rate at the end user is assumed to be known to domain managers. It is also assumed to be coverable by the aggregate bandwidth of all paths to the involved servers.

IV. Proposed Handoff Management Scheme

As mentioned earlier, our multi-source streaming strategy described in [3] incorporates a set of approaches that coordinate among the participating servers, synchronize between them and schedule the transmission of video data. Details on the working of these approaches can be found in [3].

In addition to these mechanisms, the work outlined in this paper attempts to find an adequate method to support the mobility of users over the mobile network and to smoothen handoffs. Before delving into details about the mobility management method, we first describe some operations to help domain managers :i) to reduce the frequency of handoff occurrences and ii) to predict the handoff occurrence time.

A. Preliminaries

For any kind of service, it is vital to assure that the communication is seamless, in other words the application layer at the mobile terminal should be unaware of the handoff event and the associated procedures.

To reduce the impact of such handoffs on the streaming quality, the domain manager can adopt a context aware server selection method when choosing the servers for a particular mobile node entering its coverage area. Indeed, in the area of a given domain, using a cross layer design similar to that proposed by the authors in [22], the application layer of a mobile node can refer to a set of tools to sort out the access points which the mobile node is most likely going to be connected to during the streaming service. The application layer may use history on the user's mobility pattern to predict the access points. Referring to a spatial conceptual map, along with the user's personal information, its current position, and its velocity heading (e.g., vehicles), the application layer can make an accurate prediction of the most probable future access points [23]. Prior knowledge on the topology of the wireless network [24] and the type of the application [25] can further increase the accuracy of the prediction. After this operation, the domain manager is informed of the list of access points that the mobile node is most likely going to be connected to during the streaming service time. Note that this operation is performed only at the beginning of the service or when a mobile client enters a new domain and shall incur no significant overhead at the client.

Using the mobility pattern of a mobile node, the domain manager selects the most appropriate servers for the mobile user in a way that guarantees that the video streaming to the mobile user traverses the minimum average number of hops during the entire service time. This would prevent the delay variations and jitter, alleviate network congestion, and ensure seamless and prompt service to the user. For instance, given the mobility direction of the user in Fig. 1, if a pair of servers is to be selected, the mobile client ought to be connected to Servers S_3 and S_4, rather than to Servers S_1 and S_2, despite the fact that the mobile user is geographically closer to the latter servers at the time of the connection establishment. It should be stressed out that in the absence of the aforementioned cross layer design, the selection of the servers can be arbitrary and dynamic.

Another important aspect in mobility management consists in defining efficient proactive mechanisms that can assess network conditions and user's preferences to make accurate handoff decision and accurate estimation of the handoff occurrence time. Regarding the former operation, [21] provides some good insights about how intelligent sensing of MAC QoS conditions and monitoring of the physical layer can help in making accurate roaming decisions in heterogeneous mobile networks. Concerning the second operation, it can be achieved by referring to users' profiles and adding adequate intelligence to domain managers, as will be explained below.

First of all, two types of end-hosts can be envisioned: terminals equipped with technologies (e.g., GPS) that enable them to locate their current positions and their velocity vectors (i.e., both direction and norm), and terminals without such features. A typical example of the former and the latter are GPS-equipped vehicles and simple mobile phones, respectively. Note that in case of a user using a cellular phone while driving his/her car, information on the car speed, its location, and its moving direction can be transmitted via short range technologies such as Bluetooth to the cellular phone, discoverable by the Bluetooth following different neighbor discovery protocols [26]. With prior knowledge on the topology of their service areas, domain managers can easily estimate the handoff occurrence time for mobile terminals capable to retrieve their current locations and to compute their moving speeds. In case of terminals without such ability (i.e., a mobile user walking while receiving a video stream on his/her mobile phone), prediction of handoff occurrence time can be achieved by adding context awareness to domain managers. Indeed, for each particular location (i.e., overlapping area between two or more adjacent access points), domain managers develop a statistical profile of users' behavior over time. For instance, at a particular location, (e.g., near to a train station), the time elapsed since a user enters the overlapping area between the coverage areas of two adjacent access points till the actual handoff occurrence time during rush hours is different than that during less busier times (e.g. late in the night). With this spatial and temporal aware user behavior profiling mechanism, domain managers will be able to roughly estimate the actual time a user may perform handoff for each location and each time. Similar in spirit to this operation, engineers from Azalea Networks Solutions have developed a system for distributed roaming in wireless mesh networks where cells are able to track a user's movement across cells and to predict the user

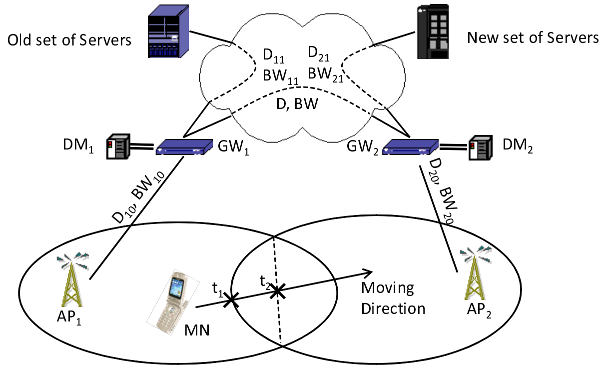

Fig. 2. A schematic model of the handoff operation.

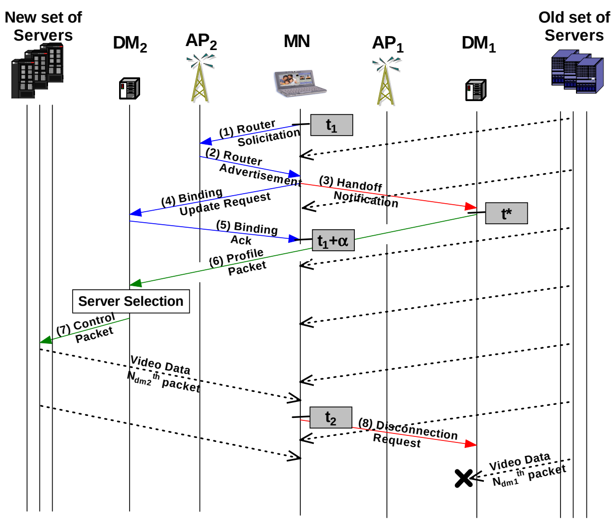

Fig. 3. Sequence of packets exchanged during the handoff operation.

’s next movement [27].

B. DM-driven handoff management scheme

Two types of handoff can be envisioned:

- Intra-domain handoff: handoff between two access points within a particular domain.
- Inter-domain handoff: handoff between two adjacent access points belonging to different domains.

The handling of intra-domain handoffs can be performed by a set of mobility management techniques that have been proposed in recent literature (as described in the related work section). The inter-domain handoff can be modeled as shown in Fig. 2. Let t_1 denote the time at which MN enters the coverage area of the new AP (i.e., overlapping area between AP_1 and AP_2). Let t_2 denote the actual handoff occurrence time. While t_2 may depend on both MAC and physical layers [21], for the sake of simplicity, t_2 is assumed to be the time when the mobile node passes the middle line of the overlapping area and switches to AP_2.

Fig. 3 shows the sequence of signaling packets exchanged during the handoff operation. When a MN enters the overlapping area between the coverage areas of two access points and receives a router advertisement message from the new AP, it notifies DM_1 (via AP_1) of an imminent handoff occurrence informing it of the next AP. Simultaneously, the mobile node initiates the handoff registration with the new AP. Let α denote the time required for the binding registration. It should be noted that the value of α largely depends on the underlying mobility management protocol (e.g., Hierarchical MIPv6) and the used handoff management protocol. In general, α should be smaller than the difference $(t_2 - t_1)$. Otherwise, disruption in the communication will be inevitable. It should be also observed that the difference value of $(t_2 - t_1)$ depends on the surface of the overlapping area and the velocity vector of the MN. In the envisioned network architecture, ignoring the medium access delay (which is in the order of micro seconds), α can be estimated as the round trip time from AP_2 to DM_2; $(\alpha = 2 \cdot D_{20})$. A more accurate estimate of α can be obtained by using history records of the MN's mobility and averaging the values using the Exponentially Weighted Moving Average (EWMA) method.

In response to the “handoff notification” message, DM_1 sends a “profile packet” to DM_2 containing the following information: current time t^*, session ID (video title), MN's home address, its play-out rate R_p, IDs of the involving servers, and an estimated sequence number N_{dm2}. This estimated sequence number N_{dm2} refers to the sequence number of the packet that should be reaching the mobile node by $(t_1 + \alpha)$, the time the handoff operation is successfully registered. Let N^*_{dm1} denote the sequence number of the packet to be forwarded at time t^*, the estimated sequence number N_{dm2} can be computed as follows:

$$N_{dm2} = N^*_{dm1} + \lfloor \frac{\alpha \cdot R_p}{L_p} \rfloor \tag{1}$$

where L_p denotes the average video packet length. As explained earlier, in case of mobile nodes able to estimate their moving speeds and their geographical locations (e.g., vehicles equipped with GPS), and with prior knowledge of the architecture topology (e.g., coordinates of access points and their transmission ranges), the actual handoff occurrence time can be roughly estimated. In this case, the sequence number N_{dm2} can be then computed as follows:

$$N_{dm2} = N^*_{dm1} + \lfloor \frac{MAX\{(t^*_2 - t_1); \alpha\} \cdot R_p}{L_p} \rfloor \tag{2}$$

where t^*_2 denotes the predicted handoff occurrence time. In response to the “profile packet” transmitted from DM_1, DM_2 searches for adequate servers, carries out the rate allocation mechanism as in [3], and transmits control packets to the new set of selected servers[1]. Upon receiving the control messages, the new servers carry out the scheduling mechanism as proposed in [3] and starts streaming data to the new location of the mobile node starting from the packet with the sequence number N_{dm2}.

At a later time t_2, the mobile node sends a “disconnection request” to DM_1 via AP_1. Upon receiving the disconnection request message, DM_1 checks the sequence number of the first packet that should be discarded. Let N_{dm1} denote that

[1]This set of servers can include all, some, or none of the old servers.

sequence number. DM_1 compares between the two sequence numbers N_{dm2} and N_{dm1}. Three cases can be envisioned:

- $(N_{dm1} - N_{dm2}) < 0$: In this case, $(N_{dm2} - N_{dm1})$ packets will be missed. Similar in spirit to many fast handoff management schemes such as FMIPv6 [17] and FH-MIPv6 [18], DM_1 immediately encapsulates this number of packets and forwards them to DM_2 so they can be transmitted to the mobile node via AP_2. Intuitively, the objective of this operation is to reduce the packet loss due to the error in the estimation of N_{dm2}. In this case, the MN buffer should help in accommodating the delay in the transmission of these packets. It should be noted here that unlike fast handoff approaches, in the proposed system DM_1 transmits only a portion of "in-flight" packets.
- $0 \leq (N_{dm1} - N_{dm2}) \leq \epsilon$: ϵ indicates the number of replicated packets that the system can tolerate. ϵ can be computed as follows:

$$\epsilon = \frac{(t_2 - t_1 - \alpha + D) \cdot R_p}{L_p} \tag{3}$$

  This equation shows the number of packets that would have been received by the mobile node since the handoff registration $(t_1 + \alpha)$ till the actual handoff occurrence time t_2 in addition to the time required for DM_1 to notify DM_2. In other words, in case of setting ϵ as in Equation 3, even if DM_1 notifies DM_2 of the possible number of replicated packets, it would be too late as the packets have been already sent. For this reason, DM_1 does nothing in this case.
- $\epsilon < (N_{dm1} - N_{dm2})$: In this scenario, a high number of redundant packets will be (or has been) transmitted. DM_1 will immediately notify DM_2 of the event. DM_2 will consequently forward only packets with sequence numbers exceeding N_{dm1}.

During handoff, out-of-order and/or duplicate packets may inevitably occur. This issue can be augmented by buffering capabilities. Indeed, a small buffer is typically required to ensure coherent reception, to remove the jitter added by the network, and to recover the original timing relationships between the media data. At the transport layer, mobile nodes are assumed to acquire a small buffer for holding a small number of frames before playing them. This small buffer is responsible for buffering and reordering all the incoming packets. It is also responsible for filtering out the duplicate packets that may occur during handoff before delivering them to the decoder at the application layer. In case of a loss detection, the small buffer should wait for the lost packet for a certain time interval. If the packet does not arrive within the time interval, the buffer delivers its content to the decoder with the missing packets. The time interval should be set in a way that avoids user level disruption during handoff, and keeps the buffer size and play-out delay small. Throughout the paper, this time interval is referred to as play-out delay and is denoted as Δ.

It should be emphasized that there are several implementation issues that must be resolved when applying the proposed mobility management scheme to practice. For instance, domain managers should be capable of reading up to the sequence number of packets. This violates the IP-Security (IPSec) semantics according to which packets must be processed by only the end-hosts; no third party is allowed to look at/alter the payload. Despite such violations, a number of mobility management schemes requiring the same ability have been proposed in the literature. For instance, in [28], a MIP based mobility management approach is proposed. The scheme addresses the packet ordering issues during handoffs by applying a double buffer technique. The technique consists in buffering the packets at both the old point of attachment and the new point of attachment; the re-ordering is then achieved by the new point of attachment before forwarding to the mobile terminal in the right order.

TABLE I

SIMULATION PARAMETERS

Factor	Parameters and range of value
Buffer size B (pkts)	100
Packet size L_p (kB)	1
Playback rate R_p (pkts/s)	100
Monitoring interval δ (s)	0.5
Video size (pkts)	1940

V. PERFORMANCE EVALUATION

The remainder of this section verifies how the proposed handoff management strategy is efficient in smoothly handling handoffs. The performance evaluation relies on computer simulation, using Network Simulator (NS) [29]. For the sake of simplicity, the coverage areas of two access points, AP_1 and AP_2, managed by two different domain managers, DM_1 and DM_2, are considered. The coverage radius of each access point and the distance between the two access points are set in a way that the maximum overlapping distance is equal to D_{max}.

A population of T_m mobile nodes is simulated and is randomly scattered over the coverage area of AP_1. The moving directions of mobile nodes are simulated in a way that mobile nodes perform handoff to AP_2 at different times. The moving speed of mobile nodes is, thus, deliberately derived from a uniform distribution. Considering users in both an urban scenario (e.g., downtown) and a highway scenario, the minimum and maximum values of the distribution are set to a slow node moving speed, $18Km/h$, and a high node moving speed, $108km/h$, respectively. To ensure a certain level of stability in the streaming of video data from AP_1, all nodes remain immobile for a short period of time from the commencement of the simulation. To avoid any possible confusion between stream disruption due to packet drops (due in turn to network congestion) and that due to handoffs, no background traffic is simulated.

On every handoff of a mobile host, statistics such as time of entrance to the overlapping area t_1, actual handoff occurrence time t_2, handoff time α, sequence number of the first packet received via AP_2 and its reception time are collected. All results are an average of seven simulation runs.

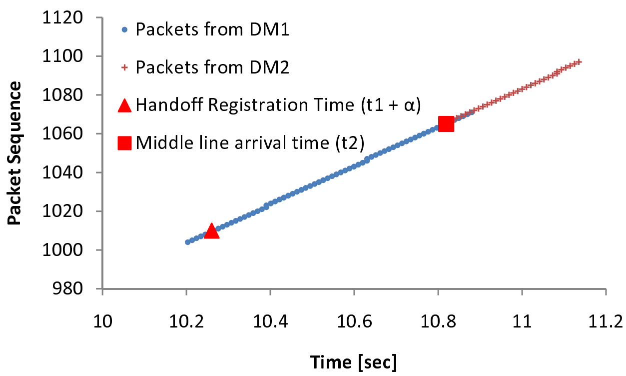

Fig. 4. Packet sequence number variation during handoff period for a mobile node traveling between DM_1 and DM_2 at a slow speed, 18 km/h. ($D_{max} = 10m$)

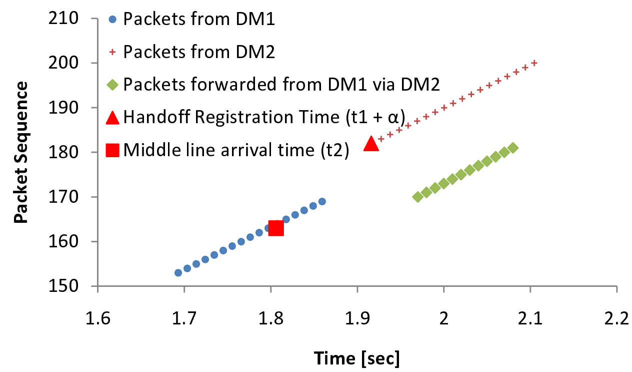

Fig. 5. Packet sequence number variation during handoff period for a mobile node traveling between DM_1 and DM_2 at a relatively high speed, 46 km/h. ($D_{max} = 10m$)

A handoff is considered to be successfully handled if all packets could be played in order at the MN's display. To minimize the effects of packet drops on the system performance, numerous experiments conducted in [19][14] have recommended the setting of the play-out delay to values larger than $100ms$. From Table I, this is equal to the time required to display 10 packets. To demonstrate the efficiency of the system in smoothing handoffs, the following parameter is defined

$$\Phi = \frac{N_s}{N_h} \cdot 100 \tag{4}$$

where N_h and N_s denote the total number of mobile nodes that performed handoff during simulation time and the number of successful handoff operations, respectively.

As explained earlier, upon an inter-domain handoff occurrence, the old domain manager notifies the new manager of the range of frames that the mobile node may need. The prediction of this range of necessary frames may be inaccurate and results in the transmission of duplicate packets. To evaluate the system efficiency in terms of the number of duplicate packets, we define the transmission efficiency of the system, Ψ, as the ratio of the number of no redundant packets to the total number of transmitted packets averaged over the number of handoffs, N_h. For each mobile node k, $\{k = 1, 2....N_h\}$, let N^k_{total} and $N^k_{duplicate}$ denote the total number of packets received by the mobile node k and the total number of duplicate packets received during the residual time of the mobile node in the overlapping area, respectively. The system efficiency, Ψ is expressed as

$$\Psi = \frac{1}{N_h} \cdot \sum_{k=1}^{N_h} (1 - \frac{N^k_{duplicate}}{N^k_{total}}) \cdot 100 \tag{5}$$

Figs. 4 and 5 plot the sequence numbers of data packets received, during the handoff period, at two mobile nodes traveling between domains DM_1 and DM_2 at two speeds, $18km/h$ and $46km/h$, respectively. The two figures consider a scenario where the maximum overlapping distance D_{max} is set to $10m$. Fig. 4 illustrates the case of a smooth handoff operation where the mobile node received packets with no disruption in communication and with only few duplicate packets. In contrast, Fig. 5 shows the case of a mobile node moving at a relatively high speed which makes the residual time of the mobile node in the overlapping area significantly short, resulting in the actual occurrence of the handoff before the end of the handoff management operation. This makes the mobile node miss a number of packets, which are forwarded later by DM_1 to DM_2 and then to the mobile node. The figure demonstrates that the disruption time (between the reception of the last packet from DM_1 and the reception of the first packet forwarded from DM_1 via DM_2) is less than $110ms$, a value that can be accommodated by the node's buffering. This indicates the good coordination that takes place between the two managers to enable mobile nodes to recover from missing packets.

Fig. 6 plots the handoff success rate Φ and the streaming efficiency Ψ for different sets of mobile nodes. For all the simulated number of mobile nodes, the figure demonstrates the robustness of the proposed scheme in handling handoff as the handoff success rate remains always in the vicinity of 100%. This indicates a successfully smooth playback of data at most user terminals. This good performance, however, comes along with low values of Ψ. This is most probably due to the inaccuracy in the estimation of the handoff delay α and the actual handoff occurrence time t_2^* which in turn affects the estimation of N_{dm2}.

To investigate the impact of the surface of the overlapping area on the working of the proposed handoff management scheme, we plot Φ and Ψ for different values of D_{max} in Fig. 7. The total number of nodes is set to 100. The figure shows that the handoff success rate increases with the increase in the overlapping area surface. This is due to the fact that with wide surface areas mobile nodes have residual times long enough to recover from the handoff. At the same time, these long residual times lead to the transmission of higher number of duplicate packets, and thus result in small values of the streaming efficiency Ψ. With this regard, it should be noted that it is worthier to secure smooth playback of video data at terminals even if this generates a number of duplicate packets. The streaming efficiency can be improved if the system can accurately predict the exact handoff occurrence time t_2. This

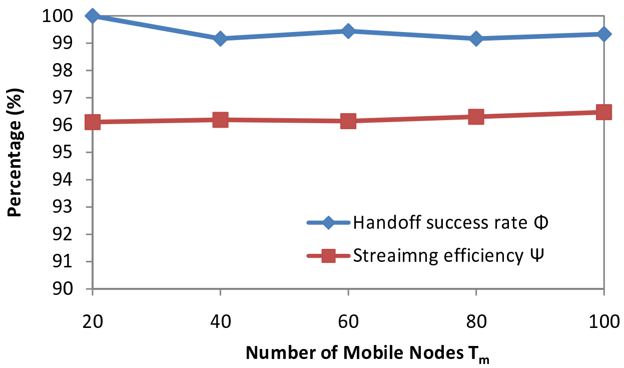

Fig. 6. Performance in terms of Φ and Ψ for different sets of mobile nodes.

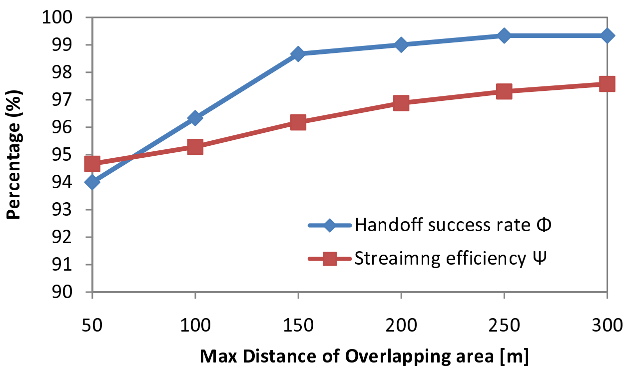

Fig. 7. Performance in terms of Φ and Ψ for different values of D_{max}.

deserves further investigation and is left for future research works.

VI. Concluding Remarks

In a previous research work of the authors [3], a multi-source approach for video streaming to mobile users is proposed. As a continuation of the work, this paper discussed the impact of handoff events on the approach and devised an adequate handoff management scheme. Simulations were conducted and results are discussed.

While the presented results are satisfactory, the working of the performance of the proposed handoff management scheme can be highly improved if it incorporates a mechanism that can predict the actual handoff occurrence time with high accuracy. Designing such a mechanism defines the future work of the authors in this particular area of research.

References

[1] B. Wang, W. Wei, Z. Guo, and D. Towsley, "Multipath Live Streaming via TCP: Performance and Benefits", in Proc. IEEE INFOCOM, Anchorage, Alaska, USA, May 2007.

[2] T. Nguyen and A. Zakhor, "Multiple Sender Distributed Video Streaming," IEEE Transactions on Multimedia, Vol. 6, No. 2, pp. 315-326, Apr. 2004.

[3] T. Taleb, T. Nakamura, and K. Hashimoto, "Multi-Source Streaming in Next Generation Mobile Communication Systems," in Proc. IEEE ICC'08, Beijing, China, May 2008.

[4] H. Man and Y. Li, "A multipath video delivery scheme over diffserv wireless lans," in Proc. SPIE-IS&T Electron. Imag., San Jose, USA, Jan. 2004.

[5] S. Yeh, C. Chen, and R. Liang, "A Context-aware Blogging and Learning Service in Mobile IPv6 Networks," in Proc. IEEE WCNC, Hong Kong, China, Mar. 2007.

[6] Palo Alto Research Center (PARC), "Content Centric Networking: A Self-Organizing Network That Meets Information Needs," Media Backgrounder, Nov. 2006.

[7] H. Hsieh and R. Sivakumar, "pTCP: An End-to-End Transport Layer Protocol for Striped Connections," in Proc. of 10^{th} IEEE Int. Conf. on Network Protocols (ICNP 2002), Paris, France, Nov. 2002.

[8] R. Stewart, Q. Xie, K. Morneault, et al., "Stream Control Transmission Protocol," RFC 2960, Oct. 2000.

[9] D. Jurca and P. Frossard, "Video Packet Selection and Scheduling for Multipath Streaming," IEEE Transactions on Multimedia, Vol. 9, No. 3, pp. 629-641, Apr. 2007.

[10] J. Rosenberg et al., "SIP: Session Initiation Protocol," RFC 3261, Jun. 2002.

[11] D. Funato, K. Yasuda, and H. Tokuda, "TCP-R: TCP mobility support for continuous operation," in Proc. IEEE ICNP, Atlanta, GA, pp. 229-236, Oct. 1997.

[12] A. C. Snoeren and H. Balakrishnan, "An end-to-end approach to host mobility," in Proc. Mobicom, Boston, MA, USA, Aug. 2000.

[13] A.G. Valko, *"Cellular IP: A new approach to Internet host mobility,"* ACM SIGCOMM Comp. Commun. Review, Vol. 29, No. 1, pp. 50-65, Jan. 1999.

[14] R. Ramjee, K. Varadhan, L. Salgarelli, S. Thuel, S.Y. Wang, and T. La Porta, *"HAWAII: A Domain-Based Approach for Supporting Mobility in Wide-Area Wireless Networks,"* IEEE/ACM Trans. Networking, Vol. 10, No. 3, pp. 396-410, Jun. 2002.

[15] H. Soliman, C. Catelluccia, K. El Malki, and L. Bellier, *"Hierarchical mobile IPv6 mobility management (HMIPv6),"* Network Working Group, RFC 4140, Aug. 2005.

[16] S. Das, A. Misra, and P. Agrawal, *"TeleMIP: Telecommunications-enhanced mobile IP architecture for fast intradomain mobility,"* IEEE Pers. Commun., Vol. 7, No. 4, pp. 50-58, Aug. 2000.

[17] R. Koodli, *"Fast Handovers for Mobile IPv6,"* Network Working Group, RFC 4068, Jul. 2005.

[18] H. Jung, H. Soliman, S. Koh, and N. Takamiya, *"Fast Handover for HMIPv6,""* Network Working Group, draft-jung-mipshop-fhmipv6-00.txt, Oct. 2005.

[19] Y. Pan, M. Lee, J.B. Kim, and T. Suda, *"An End-to-End Multi-Path Smooth Handoff Scheme for Stream Media,"* IEEE J. Select. Areas Commun. Vol. 22, No. 4, pp. 653-663, May 2004.

[20] A. Helmy, *"A Multicast-Based Protocol for IP Mobility Support,"* ACM SIGCOMM 2^{nd} Int. Workshop on Networked Group Communication, pp. 49-58, Nov. 2000.

[21] C. Guo, Z. Guo, Q. Zhang, and W. Zhu, "A seamless and proactive end-to-end mobility solution for roaming across heterogeneous wireless networks," IEEE JSAC, Vol. 22, No. 5, pp. 834-848, Jun. 2004.

[22] T. Taleb, K. Kashibuchi, A. Leonardi, S. Palazzo, K. Hashimoto, N. Kato, and Y. Nemoto, "A Cross-Layer Approach for an Efficient Delivery of TCP/RTP-based Multimedia Applications in Heterogeneous Wireless Networks," to appear in IEEE Trans. on Vehicular Technology journal.

[23] N. Samaan and A. Karmouch, "A Mobility Prediction Architecture based on Contextual Knowledge and Spatial Conceptual Maps," IEEE Trans. Mobile Computing, Vol. 4, No. 6, pp. 537-551, Nov./Dec. 2005.

[24] C. C. Tseng, L. H. Yen, H. H. Chang, and K. C. Hsu, "Topology-Aided Cross-Layer Fast Handoff Designs for IEEE 802.11/Mobile IP Environments," IEEE Commun. Mag., Vol. 43, No. 12, pp. 156-163, Dec. 2005.

[25] S. Khan, Y. Peng, and E. Steinbach, "Application-Driven Cross-Layer Optimization for Video Streaming over Wireless Networks," IEEE Commun. Mag., Vol. 44, No. 1, pp. 122-130, Jan. 2006.

[26] J. Beasley, D. Dombrowski, M. Kuiken, W. Mergenthal, and S. Stephens, "Wireless base station neighbor discovery in a communication system, such as a system employing a short-range frequency hopping scheme," U.S. Patent Num. 7,251,489, Jul. 2007.

[27] W. Wang, X.F. Wang, F. Zhao, and F. Wu, "Method and system for wireless roaming services for mobile users on wireless mesh networks," U.S. Patent Application 20060234701, Oct. 2006.

[28] A.A. Akintola, G.A. Aderounmu, and M.O. Adigun, "Performance analysis of double buffer technique (DBT) model for mobility support in wireless IP networks," in J. Network and Computer Applications, Vol. 29, No. 4, pp.217-234, Nov. 2006.

[29] UCB/LBNL/VINT: Network Simulator - ns (version 2). http://www.isi.edu/nsnam/ns/