## Chapter 1: Introduction

Add brief summary of key contribution of each chapter in a tabular form or just write one paragraph for each chapter.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Chapter title</th>
<th>Summary of key contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Introduction to mobility management for multimedia</td>
<td>Comprehensive survey and analysis of several generations of mobility protocols (e.g., 1G, 2G, 4G and 4G) to extrapolate the common abstract functions for a mobility event</td>
</tr>
<tr>
<td>3</td>
<td>System analysis of mobility events</td>
<td>Development of a new synthesis that provides a fundamental taxonomy of handover functions and their relationships. This taxonomy provides a basis for describing and characterizing optimization at each layer. Experimental analysis of the handoff delays from the application layer and network layer mobility protocols based on this handover taxonomy.</td>
</tr>
<tr>
<td>4</td>
<td>Modeling mobility</td>
<td>Data dependency analysis and resource analysis of the handover components based on the taxonomy. Building the mobility model for the handoff processes using Deterministic Timed Transition Petri net (DTTPN) based on data dependency and resource dependency. New mechanisms to predict the systems performance and system anomaly of a handoff system using the Petri net model. New mechanisms to investigate the opportunity for parallelism based on resource modeling</td>
</tr>
<tr>
<td>5</td>
<td>Mobility optimization techniques</td>
<td>Proposed proactive, reactive and cross layer mechanisms to optimize several handoff components as determined as part of Chapter 3. Discovery (5.2) New application layer discovery mechanism that discovers the network elements of the target networks in an access independent manner. By discovering these proactively and caching these, network discovery latency is reduced to almost zero. Authentication (5.3) First network layer assisted layer 2 pre-authentication mechanism that bootstraps layer 2 authentication process in the neighboring networks prior to mobile’s handover. This mechanism reduces the authentication delay to 16 ms for both inter-subnet and inter-domain handover. Configuration(5.4) A new proactive IP address acquisition scheme that reduces the signaling exchange by obtaining the IP address from the target network over a secured tunnel before layer 2 handover. First reactive router assisted duplicate address detection mechanism, where the router multicasts ARP-cache in a periodic interval so that the mobile avoids the ARP checking for duplicate address detection. Security Association(5.5) A new Anchor agent assisted layer 3 security association mechanism that hides the change of network layer identifier address of the mobile and avoids the rekeying process to maintain the security context. First pre-registration-based mechanism that establishes security context</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>5.6</td>
<td>Binding Update</td>
<td></td>
</tr>
<tr>
<td><strong>Prior to mobile’s handoff.</strong> The handover due to security association is completely eliminated. <strong>A new reactive hierarchical binding update mechanism that uses two level hierarchy of addresses and an anchor agent to limit the global signaling update during mobile’s mobility within a domain. This mechanism achieves about 70% reduction in global signaling overhead for a 10 subnets/domain scenario.</strong> <strong>A new proactive binding update mechanism over a secured tunnel that eliminates the binding update delay completely at the expense of the proactive tunnel between the mobile and the target network.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.7</td>
<td>Media Re-routing</td>
<td></td>
</tr>
<tr>
<td><strong>First reactive forwarding mechanism redirects the in-flight data from previous network using application layer mobility proxy in the previous network.</strong> <strong>First mobile controlled buffering mechanism that controls the buffering period dynamically based on handoff duration during proactive handoff.</strong> <strong>A new proactive multicasting mechanism that multicasts the in-flight data to the neighboring networks and reduces in-flight packet loss during handoff.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.8</td>
<td>Route optimization</td>
<td></td>
</tr>
<tr>
<td><strong>First packet interceptor-assisted mechanism that modifies the packets at the end hosts to maintain the direct path between the communicating hosts. This mechanism reduces transport delay by 50% for the large packets.</strong> <strong>First proxy-based packet interceptor that eliminates the trombone routing delay of the signaling traffic by 60% in an IMS environment.</strong> <strong>First binding-cache-based mechanism that minimizes the end-to-end delay by a factor of 5 by localizing the media traffic in case of localized mobility protocol such as ProxyMIPv6.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.9</td>
<td>Media independent triggers</td>
<td></td>
</tr>
<tr>
<td><strong>First to propose a set of cross layer triggers based on abstract primitives that can pass information across layers and expedite handoff related operations independent of the access mechanisms (e.g., CDMA, 802.11).</strong> These triggers are adapted as part of IEEE 802.21 standards.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Optimization using multi-layer mobility protocols</td>
<td></td>
</tr>
<tr>
<td>I have developed first of its kind multilayer mobility management scheme that uses cross layer triggers from data link layers and application layers and optimizes several handoff operations, namely address configuration, layer 3 binding update and media traversal. My proposed mechanism increases the data throughput by 50% under high mobility scenario by reducing the binding update traversal during intra-domain mobility.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Optimization for simultaneous mobility</td>
<td></td>
</tr>
<tr>
<td>First to propose an analytical framework for simultaneous mobility that can predict the probability of simultaneous mobility based on inter handoff time and binding update latency. First to propose solutions for network layer and application layer mobility protocols based on timer-based retransmission, forwarding, redirecting mechanisms, and simultaneous bindings to eliminate the vulnerability of binding update.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Handoff optimization for multicast</td>
<td></td>
</tr>
</tbody>
</table>
| First to propose and implement a hierarchical scope-based multicast streaming architecture that offers local and global program.
streaming media management, real-time advertisement insertion using RTCP feedback and proactive fast-handoff mechanisms using application layer triggering.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>System evaluation</td>
<td>Verification of the experimental results from the optimized mobility systems with the results from the corresponding Petri net models. Behavioral analysis to study the deadlocks and effect of concurrency on handoff operations.</td>
</tr>
<tr>
<td>10</td>
<td>Conclusions and future work</td>
<td>Infers the best current practices for designing a mobility protocol based on the results from mobility taxonomy and systems optimization mechanisms</td>
</tr>
</tbody>
</table>

**Chapter 2: Introduction to mobility management for multimedia**

**Summary of Key contribution and Indicative results:**

**Problem:** Over the last three decades generation of mobility protocols have evolved without any systematic approach and these protocols use ad hoc mechanisms to optimize systems performance. Ad hoc approach to systems optimization cannot help prediction of systems performance.

**My approach:** I analyze the system architecture of each of the available mobility protocols (e.g., 1G, 2G, 3G and several IP-based mobility protocols), describe the respective handoff mechanisms and then compare the handoff mechanisms in terms of their common mobility functions. For example, I extrapolate how discovery, configuration, authentication functions are performed for each of the cellular and IP-based mobility protocols and describe different network parameters that are needed to perform a specific mobility function.

**Key benefits:** There is no prior work that extrapolates these common mobility functions from the existing cellular and IP-based mobility protocols. My comparative analysis and extrapolation of the abstract functions determine the required functions that one would need in order to design a new mobility protocol with certain parameters and design a specific optimization.

**Chapter 3: Systems analysis of mobility events**

**Summary of Key contribution and Indicative results:**

**Problem:** There is no systematic analysis of the handoff processes and the associated sub-processes that are executed in sequences in all the layers during an IP-based handoff in 4G networks. Without this systematic analysis of the handoff processes, it is difficult to determine the data dependency among the handoff processes and determine potential optimization techniques.

**My approach:** Based on the comparative analysis of mobility protocols in Chapter 2, I systematically analyze each of these handoff processes for an IP-based 4G networks. This systematic analysis of the abstract functions highlights how these mobility functions are performed across several layers in an IP-based network. For example, a discovery function at layer 2 implies discovering a layer 2 point of
attachment such as an access point while discovering a layer 3 point of attachment involves discovering a default router. Similarly, authentication, security association, and configuration are some other handoff components that take place along all layers. This analysis also analyzes the most primitive operations that are part of each of these handoff functions. Based on these analysis, I built several un-optimized handoff systems such as MIP-based network layer mobility (MIPv4, MIPv6), and SIP-based application layer mobility and demonstrate experimental results for each of these un-optimized handoff components that show delays associated with each of these mobility components (e.g., discovery, authentication, configuration).

**Key advantages:** From the Systematic analysis (decomposition) of the handover operations and experimental results from some of the indicative un-optimized handoff systems that I have built, it is possible to determine the delays associated with each of the primitive operations associated with the handoff event. Unlike other experimental analysis by Vogt et al. that were meant only for one type of mobility protocol (e.g., MIPv6), my experimental analysis are based on both network layer and application layer mobility protocols and work across access techniques (e.g., CDMA, 802.11) and can thus map the handover delays with the common abstract functions independent of the types of mobility protocols and access mechanism.

**Chapter 4: Mobility modeling**

**Summary of Key contribution and Indicative results:**

**Problem:** In the absence of any formal mechanism to analyze the dynamics of handoff systems, it is difficult to predict or verify the systems performance of un-optimized handover or any specific handoff optimization technique

**My proposal:** I model the handoff-related processes as Discrete Event Dynamic Systems (DEDS) and use Deterministic Timed Transition Petri Net (DTTPN) to build various un-optimized mobility models and their associated optimization techniques.

**Key advantages:** My proposal has the following key results and advantages:

- The proposal analyzes data dependency among handover components in a comprehensive manner.
- Provides resource dependency analysis of various handoff operations from an experimental testbed
- Design of a mobility system model using Timed Transition Petri net based on data dependency analysis and resource dependency. Key benefits of this model are as follows.
  a. This model can predict systems performance for optimized handoff operations
  b. This model can design optimal path for sequence of execution of events based on expected performance and resource constraints
  c. This model can verify systems behavior (e.g., deadlocks) during handover
- Design of various Petri net-based approaches (e.g., Floyd algorithm, RTP-based, Matrix-based solution) to evaluate the mobility models for different handoff components
This system model can investigate parallelism and opportunity for optimization in a specific handoff operation.

Using these models one can predict or verify the systems performance of an un-optimized handover or any specific handoff optimization technique which is otherwise not possible in the absence of any specific handoff optimization technique.

Chapter 5: Optimization Techniques

Key contribution and Indicative results: Developed a series of optimization techniques for different handoff components as analyzed in Chapter 3. Below is a summary of my proposed mechanisms and how these take care of the problems associated with each of these handoff components.

Discovery (Section 5.2):

Problem: Discovery of the network elements and resources during handover in a heterogeneous access network depends upon respective layer 2 discovery mechanism that introduces delays

My proposal: Designed an application layer discovery mechanism that enables the discovery of network parameters and resources of the target network in an access independent manner prior to mobile’s handover to the target network. It uses caching mechanism to cache the discovered network elements that can be used to perform handoff related operations such as configuration, authentication ahead of time.

Key advantages:

1. Ability to discover the network elements of all layers of the target network without depending upon underlying access mechanism

   While the lower layer discovery mechanisms such as IEEE 802.11u allow the mobile to discover the higher layer network parameters, 802.11u mechanisms are limited to 802.11 networks only. My proposed mechanism can discover the network parameters in an access independent manner (e.g., 802.11, CDMA). Gloserv is a prior application layer discovery mechanism that discovers different types of services but not the network elements.

2. Reduce the discovery latency by way of proactive discovery and caching.

   While there are access dependent optimization techniques by Shin et al., Montavont et al., Velayos et al., to reduce the discovery latency of 802.11 networks during mobile’s handover, my proposed mechanism is access independent and optimizes the network discovery operation by proactively discovering the network parameters at all layers (e.g., Access Point, Routers, AAA servers)

Authentication (Section 5.3):

Problem: In general, layer 2 authentication is performed by the respective access mechanisms (e.g., CDMA, 802.11) after the mobile hands over to the new network. Some of the existing layer 2 pre-
authentication mechanism (e.g., 802.11i) is not supported across subnets is limited to intra-subnet handoff only.

**My proposal:** My proposed network layer assisted layer 2 pre-authentication mechanism bootstraps layer 2 authentication process in the neighboring networks before the mobile moves to the new network thereby reducing the layer 2 authentication delay during handoff.

**Key Advantages:** These are the following two key advantages:

1) My proposed mechanism takes care of the shortcomings of the existing layer 2 pre-authentication mechanism (e.g., 802.11i) by supporting pre-authentication across subnets and administrative domains while providing equivalent performance as IEEE 802.11i.

2) By supporting pre-authentication across heterogeneous access networks, this mechanism eliminates the dependence of layer 2 for authentication. Thus, mobile does not need to turn on the secondary interface and saves battery power.

**Layer 3 configuration: (Section 5.4)**

**Problem:** IP address acquisition process and duplicate address detection process are two main components that contribute to layer 3 configuration delays. IP address acquisition process involves signaling exchange between the mobile and server and the mobile waits for a random period of time to detect the uniqueness of an IP address.

**My proposals:** I have proposed two different ways of reducing the IP address acquisition delays and duplicate address detection delays using proactive and reactive mechanisms.

As part of proactive discovery mechanisms, I have reduced the IP address acquisition delays by proactively obtaining the IP address from the target network over a secured handover tunnel and checking the uniqueness of the IP address. My proposed proactive mechanisms reduce the signaling exchange completely by obtaining the IP address over a secured tunnel before the mobile hands over to the target network. This IP address discovery process can work in conjunction with pre-authentication mechanism to securely obtain the IP address.

As part of reactive IP address configuration mechanism, I have used router assisted duplicate address detection where the router multicasts ARP-cache in a periodic interval so that the mobile does not need to do initiate address resolution process. Compared to other available mechanisms (e.g., PDAD), my proposed technique does not need any additional element in the network to detect the uniqueness of IP address.

**Layer 3 Security association (Section 5.5):**

**Problem:** When the IP address of either of the communicating end points changes, a new security context needs to be established that requires generation of new keys. This process results in additional signaling exchanges giving rise to handoff delays during layer 3 security association.
My proposals: I have proposed two different mechanisms that can reduce the delays due to IP security association. These mechanisms can be categorized as reactive and proactive.

1) Reactive: My proposed mechanism maintains the network layer identifier address by introducing an anchor agent in the network. This allows the security context to be maintained by avoiding the rekeying mechanisms during IP address change. Compared to traditional non-optimized versions, my mechanism reduces the handoff delay and packet loss during security association. Existing work cannot operate under heterogeneous access network.

2) Proactive: Using pre-registration with the target network, the mobile establishes the security context ahead of time by obtaining the keys proactively. Unlike other proposals [], my proposal reduces the security risks by avoiding the security

Key advantages: Layer 3 security binding can be maintained by way of establishing security context ahead of time or by hiding the IP address change when proactive means is not feasible.

Binding Update (Section 5.6):

Problem: Longer distance between the mobile node and correspondent node or home agent contributes to the binding update delay resulting in overall handoff delay and packet loss.

My proposals: I have proposed two mechanisms namely hierarchical binding update and proactive binding update that reduce the binding update delay.

1) My proposed hierarchical binding update mechanism is a reactive mechanism that uses two level hierarchy of addresses (e.g., local care of address and global care-of-address) and introduces an anchor agent called mobility agent in the network to limit the global signaling update during mobile’s mobility within a domain. It reduces the global signaling update by 70 percent for a network with 10 subnets per domain. Compared to MIP-RR which was developed around the same time, my proposed mechanism works for both network layer (e.g., MIP) and application layer mobility (e.g., SIP) protocols and offers advanced features such as dynamic load balancing based on placement of the mobility agent.

2) Proactive binding update mechanism sends the binding update before the layer 2 handover over a secured tunnel using the proactively obtained IP address as the new care-of-address. This mechanism reduces the binding update completely as the mobile does not need to send a new binding update after the handover.

Media Rerouting (5.7):

Problem: During media rerouting process transient data may get lost due to handoff operations at several layers or get delayed due to operations such as encapsulation, de-capsulation, tunneling and buffering mechanisms.

My proposals: I list below my proposed mechanisms.
1) **Proactive multicasting**: This proposed mechanism proactively multicasts the in-flight data to the neighboring networks and reduces in-flight packet loss. This mechanism avoids any additional network element in the network at the expense of additional bandwidth in the neighboring networks. Unlike the existing scheme, this mechanism uses a single multicast address and encapsulates the unicast data that gets decapsulated at the mobile.

2) **Reactive forwarding of data from previous network**: This proposed mechanism uses reactive forwarding mechanism to redirect the in-flight data from previous network using application layer mobility proxy in case of longer binding update delay.

3) **Mobile controlled proactive buffering (section 5.9)**: My proposed mobile controlled buffering mechanism provides a per-mobile packet buffer at the edge router that controls the buffering period dynamically based on handoff duration during proactive handoff. My techniques are based on time-limited buffering and explicit buffering approaches.

**Detection of network attachment (5.8):**

**Problem**: Post handoff detection mechanisms at layer 2 and layer 3 points of attachment are independent of each other causing additional delays during layer 3 handover. As part of my proposed mechanisms, I have designed the following cross layer triggers that expedite the network detection and upper layer operations.

1. **Proactive Triggers**: In order to prepare for an impending handoff and perform some of the handoff operations ahead of time I have developed cross layer triggers such as “Link going down”, “Link_Handover_Imminent” that will trigger many of the handover related upper layer operations such as application layer discovery and network layer assisted layer 2 authentication.

2. **Reactive Triggers**: I developed link layer cross layer triggers such as “Link up”, “Link Down” events to expedite upper layer handoff operations in an access independent manner such as handoff between 802.11 and CDMA networks. Unlike other event triggers, my proposed cross layer triggers work across different access mechanisms.

3. **Cross Layer Triggers**: Provide layer 3 related information (e.g., subnet prefix, default router address) during access point discovery. This is accomplished by modifying layer 2 access point beacon and stuffing layer 3 information.

**Media independent handoff triggers (Section 5.10):**

**Problem**: Handoff related functions are spread across different layers of the protocol stack and are executed independently. It is very hard to exchange control information among layers. However, for efficient network communication it is essential for a protocol layer to utilize cross layer information. Thus, it is helpful have some abstract set of primitives that can pass on the information across layers.

I have proposed a set of abstract primitives that can pass information across layers and work independent of the access mechanisms (e.g., CDMA, 802.11). Some of these abstract primitives have been used to develop media independent handover functions that have recently been standardized in IEEE 802.21 standards. Unlike other proposals, these primitives can be applied to support handover
among heterogeneous access networks such as 802.11 and CDMA. These triggers can be categorized as “Information Service”, “command service” and “Event service”. Using these primitives, one can detect the new networks and loss of old networks. Section 5.10 discusses the detailed description of these triggers and different implementation steps.

Chapter 6: Multi-layer mobility protocol

Summary of Key contribution and Indicative results:

Problem: Network layer mobility protocol, application layer mobility protocol and local mobility protocols operates independently without interacting with each other. Each of these mobility protocols has its own pros and cons. For example, network layer mobility protocol needs additional networking element (e.g., home agent) in the middle of the network to support terminal mobility and is thus not most optimized; application layer mobility protocol is best optimized to work for real-time application (e.g., VoIP) and cannot support mobility for TCP-based traffic in its current form. Local mobility protocol such as cellular IP cannot support mobility across subnets or inter-domain. An integrated mobility scheme

My proposal: I have developed a multilayer mobility management scheme that uses cross layer triggers from data link layers and application layers and optimizes several handoff operations, namely address configuration, layer 3 binding update and media traversal. This proposed mechanism uses a policy-based approach based on mobile’s movement pattern and type of application and decides the mobility protocol that is most appropriate to be used under the specific circumstances. This mechanism uses SIP-based application layer mobility to support real-time traffic and MIP-LR-based mobility to support non-real-time traffic during inter-domain movement while it uses the local mobility management protocol (MMP) to support real-time and non-real-time traffic during intra-domain movement.

Key advantages:

- My proposed mechanism increases the data throughput by 50% under high mobility scenario by reducing the binding update traversal during intra-domain mobility and uses the lower layer triggers such as layer 2 beacon id to determine intra-domain and inter-domain mobility
- My proposed mechanism expedites the discovery operation by discovering layer 3 point of attachment while discovering layer 2 point of attachment using the optimization technique by parallelism and reduces the packet loss.
- Using the application layer triggers, my proposed mechanism uses the mobility protocol that is optimized for a specific type of application (e.g., SIP for Real-time traffic and MIP for TCP).

After I developed the proposed policy-based mobility management scheme back in 2003, few other integrated mobility management schemes were developed that use SIP for personal mobility and
Chapter 7: Optimizations for simultaneous mobility

Summary of Key contribution and Indicative results:

**Problem:** Without any thorough analysis of the simultaneous mobility problem that arises due to non-receipt of binding updates when both the hosts that are in communication move, it is difficult to predict the parameters that affect the simultaneous mobility and propose solutions to mitigate these problems. Currently, there is no comprehensive study that analyzes the simultaneous mobility problem or provides solutions to mitigate these problems in an infrastructure-based mobility environment.

**My proposals:** I analyze the simultaneous mobility problem and develop an analytical framework to study the effect of inter-handoff rate and binding update latency on the probability of occurrence for simultaneous mobility problem. I proposed timer-based retransmission, forwarding and redirecting mechanisms using binding update and location update proxies and use of simultaneous bindings by the mobile to eliminate the vulnerability of binding update. I applied these techniques to several mobility protocols namely, SIP-based mobility, MIPv6 and MIP-LR.

**Key benefits:** My proposed analytical framework for simultaneous mobility can predict the probability of simultaneous mobility based on inter handoff time and binding update latency. Each of my proposed techniques can be applied either at the sender side or receiver side for both network layer and application layer mobility protocols. Each of these techniques reduces the vulnerability interval of simultaneous binding update.

Chapter 8: Handoff optimizations for multicast streaming media

Summary of key contribution and Indicative results

**Problem:** Currently, multicast-based content distribution networks do not provide flexibility in terms of local and global program management, nor do these provide automatic advertisement insertion features. These also lack fast-handoff support when the mobile moves between 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 layer 2 and layer 3 “join” latency during its movement between layer 2 access points or layer 3 subnets. Multicast Join latency is contributed due to IGMP periodic router query interval and IGMP client report.
**My proposal:** I proposed and implemented a hierarchical scope-based multicast streaming architecture that offers local and global program management and real-time advertisement insertion using feedback data such as RTCP. I developed an application layer triggering technique that uses RTCP to join the multicast tree after the mobile hands off to the new network. In order to reduce the handoff latency for multicast join, I proposed both proactive and reactive triggering techniques. As part of the proactive technique, I use an application layer proxy or multicast address announcer that joins the multicast tree on behalf of the mobile as the mobile is impeding to handover to the new network. While the multicast proxy joins the upstream router using IGMP, the mobile triggers the multicast stream using RTCP “Join”. As part of the reactive technique, the mobile joins the upstream multicast tree by sending an “unsolicited” join request during its configuration process.

**Key advantages:** 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, it provides the ability to control the advertisement duration without relying on additional signaling. By using an application layer triggering technique such as RTCP, the mobile does not need to depend upon IGMP router query interval nor does it depend upon the kernel that needs multicast support. Having the ability to trigger multicast streaming during configuration, the mobile optimizes the operations in parallel. Compared to traditional multicast handoff approaches, my proposed proactive optimization techniques can reduce the handover latency by a factor of 100 when the probability of presence of 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)

**Chapter 9: System Evaluation**

**Summary of key contribution and Indicative results**