TURKISH NAVAL ACADEMY
NAVAL SCIENCE AND ENGINEERING INSTITUTE
DEPARTMENT OF COMPUTER ENGINEERING

USER MOBILITY PATTERN BASED RSVP SCHEME FOR WIRELESS MPLS NETWORKS

MASTER THESIS

İSMAİL BUDANUR

Advisor: Dr. Erdal Çayırcı

İstanbul, 2003

USER MOBILITY PATTERN BASED RSVP SCHEME FOR WIRELESS MPLS NETWORKS

Submitted in partial fulfillment of the requirements for degree of

MASTER OF SCIENCE IN COMPUTER ENGINEERING

from the

TURKISH NAVAL ACADEMY

Author :

İsmail Budanur

Defense Date : 30/07/2003
Approved by:

Dr. Erdal Çayırcı (Advisor)

Yrd.Doç.Dr.Elif Karslıgil(Defense Committee Member)

Yrd.Doç.Dr.Gökhan Yavuz (Defense Committee Member)

ABSTRACT (TURKISH)

TELSİZ MPLS HÜCRESEL AĞLAR İÇİN KULLANICI GEZGİNLİK KALIBI TABANLI RSVP

Anahtar Kelimeler : Telsiz Ağlar, Telli Ağlar, RSVP, El Değiştirme, Kullanıcı Gezginlik Kalıbı, U-RSVP, Yer Yönetimi, İletim Katmanı, Gezginlik, MPLS

İnternet kullanıcılarının isteği özellikle gerçek zamanlı uygulamalar için büyük bir hızla artmaktadır. "Best effort service" modeli çoklu ortam uygulamaları için yetersiz kalmaktadır. Uçtan uca servis kalitesini sağlamak maksadıyla IETF tarafından değişik yaklaşımlar geliştirilmiştir. Bütünleşik service mimarisi gerçek zamanlı uygulamalar için geliştirilmiştir ancak, omurgada ilave sinyal oluşmasına sebep olmaktadır. Ayrık servis mimarisi internette ölçeklenebilir bir yapı ortaya koyarken aynı zamanda sinyal trafiğini de azaltmak için önerilmiştir. Çoklu protokol etiket anahtarlama (MPLS) hem trafik mühendisliğini ve servis kalitesini hem de ayrık ve bütünleşik servislerin sunduğu avantajları sağlamaktadır.

Bu tezde, yer yönetimi olarak tasarlanan UMP tekniği kullanarak hücresel telsiz MPLS ağlar için UMP tabanlı bir RSVP tekniği önerilmiştir. UMP tekniğinde gezgin kullanıcı, ziyaret edilecek hücreler ve hücreye giriş zamanından oluşan gezginlik kalıbını kaydeder. UMP kullanarak sadece bir sonraki hücre değil hücre giriş zamanı da tahmin edilebildiğinden yer değiştirme gerçekleşmeden önce yeni bir LSP yaratabiliriz. Bu husus MPLS'i gezgin kullanıcıların servis kalitesi bakımından daha güçlü ve güvenilir yapar. Simülasyon sonuçları çağrı engelleme oranında hafif bir yükselişle birlikte çağrı düşme oranının %20 oranında azaldığını göstermektedir.

ABSTRACT (ENGLISH)

USER MOBILITY PATTERN BASED RSVP SCHEME FOR WIRELESS MPLS NETWORKS

Keywords : Wireless Networks, User Mobility Pattern, RSVP, U-RSVP, MPLS, QoS, Handoff, Location Management, Transport Layer, Middleware, Mobility,

The Internet users' demand is explosively growing especially for the real-time multimedia applications. The best-effort service model of the Internet is inadequate to support multimedia service requests. Various approaches have been introduced by IETF to support end-to-end quality of service (QoS. The integrated services architecture was developed to provide QoS for real time applications but it creates additional signaling traffic in the backbone. Differentiated services was proposed to reduce signaling cost as well as to enable scalable services in the Internet [7]. multiprotocol label switching (MPLS) [11] provides traffic engineering (TE), QoS, differentiated services, and supports integrated services. Thus it leverages both the advantages of the integrated and differentiated services architectures.

In this thesis We propose a new user mobility pattern based RSVP scheme for wireless cellular systems where the user mobility pattern based registration and paging (UMP) technique is used for location management. In the UMP technique a mobile registers a mobility pattern made up of the cells to be visited and the expected cell entry times. Since not only the next cell but also the cell entry time can be predicted by using a UMP, we can create a new label switched path (LSP) for the new cell before a handoff occurs. This makes MPLS more powerful and more reliable for QoS requirements of mobile terminals. The simulation results show that our scheme reduces the call termination rate 80 % in the average in the expense of slightly increased call blocking rates.

DISCLAIMER STATEMENT

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Turkish Navy, Naval Academy, and Naval Science and Engineering Institute

DEDICATION

To my wife Şaziye for her love, inspiration, and support

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr.Erdal Çayırcı for his intelligent supervision and constructive guidance, inspiration and friendship. I would also like to thank Vice Admiral Lütfü Sancar for his efforts for obtaining the required materials and support. I would also like to thank to Prof. Dr Captain Süleyman Özkaynak and Dr. Vedat Coşkun for their support, helpful discussions and comments.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS, ACRONYMS, AND SYMBOLS

ID	Identification
ARIS	Aggregate Route-based IP Switching
ATM	Asynchronous Transfer Mode
BS	Base Station
BU	Bandwidth Unit
COA	Care of Address
CR-LDP	Constraint-Based LDP
Δ	A period of time that an advance LSP lasts on the average
Diffserv	Differentiated Services
dRSVP	Dynamic resource Reservation Protocol
DS	Differentiated Services
EETI	Expected Entry Time Interval
ERO	Explicit Route Object
FA	Foreign Agent
FEC	Forwarding Equivalence Class
FH	Fixed Host
FR	Frame Relay
HA	Home Agent
H-MPLS	Hierarchical Mobile MPLS
ID	Identification
IETF	The Internet Engineering Task Force
IGP	Interior Gateway Protocol
ILM	Incoming Label Map
IMT	International Mobile Telecommunications
IntServ	Integrated Services
IP	Internet Protocol
IS	Integrated Services
ISP	Internet Service Provider
LA	Location Area
LDP	Label Distribution Protocol
LSN	Label Switching Nodes
LSP	Label Switched Path
LSR	Label Switched Router
MH	Mobile Host
MM-MPLS	Micro Cell Mobile MPLS
MPLS	Multiprotocol Label Switching
MSPEC	Mobility Specification

MT	Mobile Terminal
NHLFE	The Next Hop Label Forwarding Entry
NNS	Next Node Structures
NV	Number of Visits
PCN	Personal Communications Network
PCS	Personal Communications Service
θ	A variable to make t_{max} flexible according to administrative purposes.
QoS	Quality Of Service
RSVP	Resource Reservation Protocol
RSVP-TE	Extensions to RSVP for LSP Tunnels
SA	Synchronization Agent
TE	Traffic engineering
t_{max}	A function of D and UMP accuracy (Ψ)
TOS	Type of Service
UMH	User Mobility History
UMP	User Mobility Pattern
U-RSVP	UMP based RSVP
VPN	Virtual Private Networks
WMPLS	Wireless MPLS
Ψ	UMP Accuracy

I. INTRODUCTION

The Internet users' profile is changing and growing rapidly. A worker who is trying to be on time for his or her job will want to read his/her e-mails on a bus. An other one will want to watch a remote video or a videoconference with someone else during a tiring train trip. The users demand is explosively increasing especially for the real-time multimedia applications already now via GPRS technology. The best-effort service model of the Internet is inadequate to support multimedia service requests since it causes variable queuing delays and congestion delays [50]. Various approaches have been introduced by IETF to support end-to-end quality of service (QoS). The integrated services architecture [50] and Resource Reservation Protocol (RSVP) [2,12] was developed to provide QoS for real time applications. In integrated service RSVP model every end-to-end connection needs additional signaling traffic in the backbone. Differentiated services [13] is proposed to reduce signaling cost as well as to enable scalable services in the Internet [7]. Multiprotocol label switching (MPLS) [11] provides traffic engineering (TE), QoS, differentiated services, and supports integrated services. Thus it leverages both the advantages of the integrated and differentiated services architectures.

While the demand or different Internet service for the mobile users increases, the problems that have to be solved increase as well. The most important ones are the QoS needs and using wireless bandwidth most efficiently. Terminal mobility imposes additional difficulties in the realization of reservation-based schemes designed to guarantee QoS. If the next cell that an MT will handoff and the cell entry time can be predicted during the lifetime of a data flow, resources can be reserved just in the right cell at the right time. User mobility pattern (UMP) based location update and paging (UMP) scheme [1] helps us to realize a mobility prediction in this detail.

In this thesis a new UMP based resource reservation protocol (RSVP) scheme that reduces the cost of bandwidth reservation in MPLS is introduced to support end-to-end QoS and using wireless frequency spectrum most efficiently. MPLS does not support end-to-end QoS. It solves the problems in the backbone. Our work makes MPLS more powerful since it supports both mobility and end-to-end QoS for MPLS networks. Furthermore we do not propose any change in original MPLS and original RSVP

The power of our work is the UMP based location management scheme [1]. Since not only the next cell but also the cell entry time can be predicted by using UMP, we can create a new LSP in the next cell before a handoff occurs. This makes MPLS more powerful and more reliable for QoS requirements of MTs. This work shows that a new LSP for which the source node is the new cell can be effectively created for an active connection before an MT handoffs, and this can be achieved transparent to the MT when the UMP based location management scheme is used.

In Section II, QoS and traffic engineering are explained. The UMP based location update and paging is presented in Section III. We introduce our new UMP based RSVP (U-RSVP) scheme in Section IV. The performance of U-RSVP is evaluated in Section V. We conclude our paper in Section VI.

II. QUALITY OF SERVICE

A. OVERWIEV

Quality of Service (QoS) is defined as a set of service requirements to be met by the network while transporting a flow in [51]. There are generally three phases to support QoS as shown in Fig. 1. First one is the procedure between an end user and an Internet Service Provider (ISP), the second one is the procedure in the backbone and the third one is the procedure between the other end user and his/her (ISP).

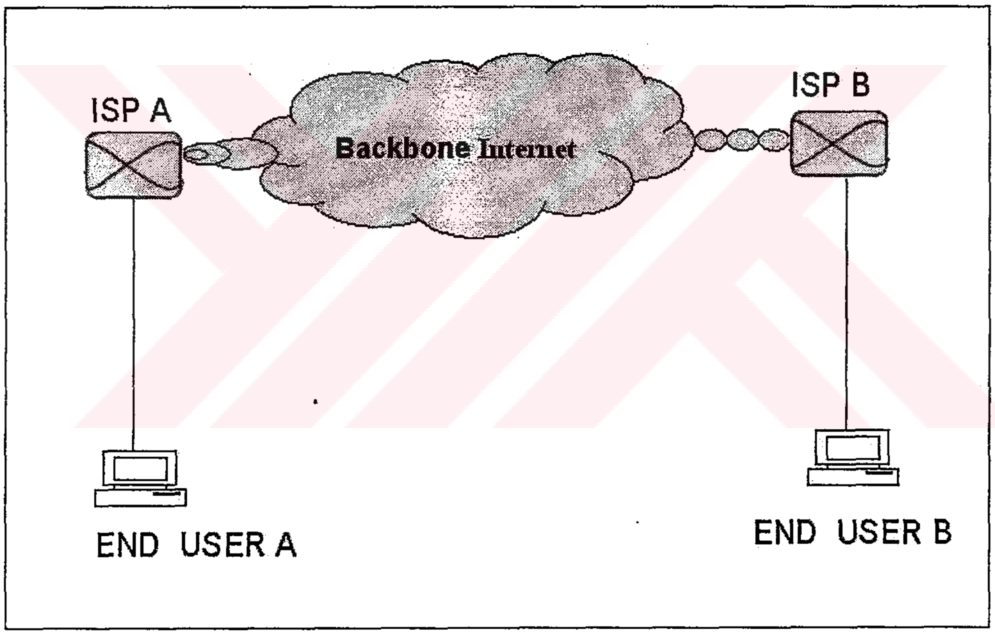


Fig. 1 QoS Phases

The study for QoS [16,17,35,41,43,57-75] aims to solve all the problems according to users requests now and for future. While the number of users and the demand for different services increase, the best effort service becomes not enough. Because best effort service just tries to send packets when the lines permits. There may be delays in the transmission of packets based on the network load. This may cause problems on voice, remote video or videoconference. The Internet Engineering Task Force (IETF) has developed different service models to meet the demand for QoS request such as Integrated Services/RSVP [12] model, the Differentiated Services (DS) model [13], MPLS [11]. We will explain these service models below.

B. TRAFFIC ENGINEERING

Traffic engineering deals with the controlling a network's data flow traffic to use the Internet resource most effectively with the best performance [28].

"Traffic engineering can be viewed as assistance to the routing infrastructure that provides additional information in routing traffic along specific paths, with the end goal of more efficient utilization of networking resources.

Traffic engineering is performed by directing trunks along explicit paths within the ISP's topology. This diverts the traffic away from the shortest path computed by the IGP and presumably onto uncongested links, eventually arriving at the same destination. The Specification of the explicit route is done by enumerating an explicit list of the routers in the path. Given this list, traffic engineering trunks can be constructed in a variety of ways. For example, a trunk could be manually configured along the explicit path. This would involve configuring each router along the path with state information for forwarding the particular label. Such techniques are currently used for traffic engineering in some ISPs today. Alternately, a protocol such as RSVP can be used with an Explicit Route Object (ERO) so that the first router in the path can establish the trunk. The computation of the explicit route is beyond the scope of this document but may include considerations of policy, static and dynamic bandwidth allocation, congestion in the topology and manually configured alternatives" [52].

While the demands increases and greater the Internet by means of technology and the infrastructure greater the study to scale the Internet about Traffic Engineering [20,21,23-26,28-33,54-57]

C. BEST EFFORT SERVICE

The first use of Internet is to send e-mails and to transfer files. When a request arrives at a router, it waits until the network resources are available to send. There are delays and congestion losses [50] but it is not important when you send e-mail or a file. We call this type of service as best-effort service. Best effort service is quite well suited to applications with lax demands for timely and predictable packet delivery. Such applications tend to generate their data in bursts, or at least can tolerate data being transferred in bursts [115]. This service is useful for e-mails and files transfers. But the demand for Internet forces ISP's to use new system models to answer for voice, vide etc. We cannot use best effort service for voice or video streaming.

"Over the last decade the Internet community came up with Integrated Services and Differentiated Services, two new architectures for resource allocation in the Internet. The two architectures introduced a number of new concepts and primitives that are important to QoS support in the Internet:

Frameworks for resource allocation that support resource assurance and service differentiation

New Service models for the Internet in addition to the existing best-effort service

- Language for describing resource assurance and resource requirements
- Mechanisms for enforcing resource allocation" [114].

Integrated Services and Differentiated Services represent two different solutions. Integrated Services provide assurance through resource reservation for individual application flows, whereas Differentiated Services use a combination of edge policing, provisioning, and traffic prioritization. [114]

D. INTEGRATED SERVICES (IS)

1. OVERVIEW

Today the digital voice and video hardware technology is inexpensive and highly-sophisticated software applications have been developed but it is seen that the real-time applications often do not work well across the Internet because of variable queuing delays and congestion losses. The Internet, as originally conceives, offers only a very simple quality of service (QoS), point-to-point best-effort data delivery. Before real-time applications such as remote video, multimedia conferencing, visualization, and virtual reality can be broadly used, the Internet infrastructure needs to be modified to support real-time QoS, which can be done by providing some control over end-to-end packet delays. This extension has to be designed from the beginning for multicasting; simply generalizing from the unicast (point-to-point) case does not work. [50] This is the reason why the Intserv model is developed. The study for integrated service continues [54,55,56]

2. INTEGRATED SERVICES MODEL

The IS model proposes two types of service targeted towards real-time traffic: guaranteed and predictive service on best effort service. It integrates these services with controlled link sharing [50]

In this model resources must be definite to guarantee a real time service. So a reservation setup protocol is necessary to work with. This is Resource Reservation Protocol (RSVP) [12]

3. RSVP

RSVP [2] is developed to provide QoS for real time applications. It is a receiver-oriented signaling protocol where a resource reservation starts from the receiver. To establish a connection, a source node first sends a PATH message to the receiver. If the receiver accepts the connection request and has enough resources for the connection, it sends a RESV message to the source. The nodes that can provide resources required for the connection repeat the RESV message to the next node on the reverse path to the source node. Fig. 2 shows RSVP signaling mechanism. If all the nodes can provide the resources needed, then the RESV message arrives at the source and the connection is established. If a node along the route does not have enough resource, the connection cannot be established. RSVP is defined for wired networks but there are studies to use it for wireless mobile networks.

In [77] a dynamic resource allocation scheme is proposed to run Intserv applications within a DiffServ network by adding RSVP extensions for real time services in wireless mobile networks.

In [78] a reservation protocol, Mobile RSVP, to provide real time services to mobile users in an integrated services packet network is defined to make advance reservation for the future location of mobile nodes. In this model, a mobile host can make advance reservations from a set of locations, called *Mobility Specification* (MSPEC). MSPEC should be the set of locations the mobile host will visit while it participates in the flow. It assumes that the mobile host acquires its MSPEC, either from the network or from its mobility profile when it initiates a reservation. It supports two types of reservations: active and passive. The MT makes active reservation from its current location and it makes passive reservation from the other locations of its MSPEC.

In [80] Dynamic resource Reservation Protocol (dRSVP) is proposed for dynamic bandwidth management for a variable bandwidth wireless environment.

It describes an approach for providing dynamic QoS support in a variable bandwidth network, which may include wireless links and mobile nodes. The dynamic QoS approach centers on the notion of providing QoS support at some point within a range requested by applications. To utilize dynamic QoS, applications must be capable of adapting to the level of QoS provided by the network, which may vary during the course of a connection. To demonstrate and evaluate the dynamic QoS concept [80] propose dRSVP

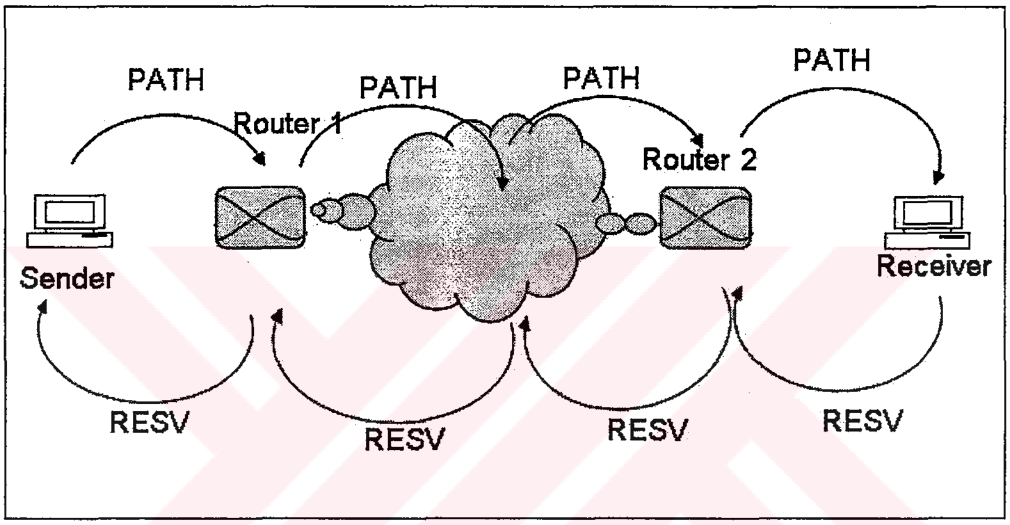


Fig. 2 RSVP Mechanism

E. DIFFERENTIATED SERVICES

Integrated services provide QoS but the cost is very high. Because RSVP has to handle each flow separately. It means every router must be aware of this model and has to handle per-flow information. In Diffserv model networks can be divided into two parts: The first is the access networks and the second is the backbone. The purpose is to send Internet packets as early as possible and classify the packets according to their QoS demand. There are two types of routers in this model; core routers and edge routers. The packets are classified in the edge routers. This procedure may take time but this is not important because the link between the user and edge routers make the bottleneck. When the packets arrive the core router they are just forwarded simply. The packets are classified according to the TOS (Type of Service) [15] byte in IP header. Applications set TOS byte (DS field) according to the need for *low-delay, high throughput* or *low-loss-rate service.* Using classification various types of services can be provided by ISPs such as:

Premium service for applications that require low-delay and low-jitter service

Assured service for applications that require better reliability than best-effort service

Olympic service, which provides three tiers of services: *gold, silver* and *bronze,* with decreasing quality [14].

Diffserv model was designed for wired networks. But there are some approaches that use Diffserv model in wireless networks. In [76] bandwidth reservation for a Diffserv wireless mobile networks is defined. [16] describes a model combining RSVP and Differentiated services to realize the benefits of each.

IntServ, RSVP and DiffServ can be seen as complementary technologies in the pursuit of end to-end QoS. Together these mechanisms can facilitate deployment of multimedia applications. IntServ enables host to request per-flow, quantifiable resources, along end to end data paths and to obtain feedback regarding admissibility of these requests. DiffServ enables scalability across large networks. The primary benefit of combining IntServ and DiffServ is the increased scalability, provided through the aggregate traffic control of DiffServ [17].

F. MPLS

1. OVERVIEW

The routing mechanism used today is IP forwarding. IP forwarding depends on datagram model In this model each packets is sent to its destination address independently. The path is calculated according to existing routing protocols. This is called destination based forwarding mechanism. Each router in the backbone looks up the header of each packet to decide where to send it. The IP header has more information than it is needed to forward the packets. This makes the Internet to slow down. There is another mechanism for data networking, called *virtual circuit.* Asynchronous transfer mode (ATM) and, Frame Relay (FR) use this mechanism. Before data transfer a virtual connection is setup by signaling protocols. This approach is called *connection oriented.* Today lots of the Internet backbones use ATM. There are mechanisms to run IP over ATM but, there are some scalability problems with ATM. The need for more seamless IP/ATM integration led to the development of label switching. The IP switch uses a property signaling protocol to set up an ATM connection on the fly for long lasting IP flows. Several other approaches were soon proposed, including Cisco's Tag Switching and IBM's Aggregate Route-based IP switching (ARIS). Although these approaches differ in detail, they are all based on the same paradigm of label switching [114].

MPLS has been developed for QoS, gigabit forwarding, network scaling, and traffic engineering to support both the integrated and differentiated services [3]. The Internet packets that have similar characteristics are assigned to a forwarding equivalence class (FEC) and labeled according to this FEC in the MPLS networks. These packets are conveyed through a virtual path called a label switched path (LSP). A label switched router (LSR) that also assigns the traffic engineering (TE) parameters to the packets does labeling. LSPs are created by a label distribution protocols (LDPs) such as LDP [10] or RSVP-TE [9]. Once the packets are assigned to a FEC, no other analysis of packets is needed in the backbone LSRs. Instead, packet forwarding is used in the MPLS network. Therefore MPLS is faster than the other systems as well as it provides service differentiation.

2. IP FORWARDING VERSUS MPLS

When a packet travels from one router to another according a network layer protocol (i.e., IP), on the backbone, each router decides how it is going to forward it. This means for example every packet for a TCP connection may not tracks the same routers. When the next router gets a packet it has to control the network layer header of that packet and has to make a forwarding decision for that. If there some address prefix in that router's routing table matches these packets are assigned to the same FEC for next hope. As the packets arrives the other router each one of them reexamines the packets.

In MPLS when a packet is assigned to a FEC no other analysis is done at the other routers. Next router just changes it label and forwards it to the next router simply.

3. LABEL SWITCHING

Label switching is the forwarding mechanism and it is performed by *Label Switched Routers* (LSRs) by adding a fix length *label* to conventional IP packet's header to forward packets. When an LSR gets such a packet it learns the next hope of the packet by looking up the information in the label of it. Then swap the label according to the next hope and forward the packet. A Labeled packet is sent through the network within a path according to the labels. This path is called an LSP.

The label switching mechanism is used in both ATM and FR. But supporting the label-switching paradigm within IP networks, enhances the services provided in the IP networks by offering scope for traffic engineering, guaranteed QoS, and virtual private networks (VPNs) [114]. When the power of label switching is noticed IETF has decided to form a working group to develop label switching standards and label distribution protocols. As a result MPLS has been held up as the solution to IP QoS, gigabit forwarding, network scaling, and traffic engineering.[22]

a. Label

A label is a short fix length physically contiguous identifier, which is used to identify a Forwarding Equivalence Class (FEC), usually of local significance [11]. It does not encode any network layer header. It means it is independent from network layer construction in encoding or decoding a packet. When a packet is assigned to a FEC no other analysis is done in the backbone routers.

b. Forwarding Equivalence Class

The IP packets that are treated in the same manner are assigned to a FEC. For example the multicast packets that have the same source and destination address may be assigned to a FEC according to their network layer header. Another example is transfer layer port numbers for a specific application within a computer.

c. Label Switched Router (LSR)

A Label Switched Router is the router that can encode and decode the label information. An LSR encapsulate a network layer packet in a network layer 2 packet.

(1) MPLS domain:

A contiguous set of LSRs, which runs MPLS routing and forwarding, and which are also in one Routing or Administrative Domain [11].

(2) MPLS edge node:

An MPLS router that connects the MPLS domain with a non MPLS domain or a router that con not run MPLS since it is in out of an MPLS domain.

(3) Ingress Label Switched Router

It is an LSR that handles the traffic entering an MPLS domain.

(4) Egress Label Switched Router

It is an LSR that handles the traffic leaving an MPLS domain.

d. Label Switching Path (LSP)

A Label Switching Path is a virtual path, which consists of the routers in which the packets in a FEC travel. The path information is placed in a label at the ingress Label Switched Router (LSR).

e. Label Stack

MPLS permits more than one label to be encoded in a packet organized by first in last out stack. This is called label stack. It is useful on a path, used by lots of LSPs. In this situation the backbone LSRs are not related to the LSPs being creation or released between them. Actually this mechanism is called LSP tunnel.

4. LABEL SWITCHING TABLE

The *label-switching table* resembles packet-forwarding table in IP routers. It is also called an *incoming label map* (ILM). It contains information to map an incoming label to the outgoing label for the next hop.

5. THE NEXT HOP LABEL FORWARDING ENTRY (NHLFE)

The entry that the incoming label points to is called the next hop label forwarding entry (NHLFE). Each incoming label typically points to one NHLFE. However, in the case of load sharing, there may be multiple NHLFEs for an incoming label. Typically the NHLFE contains the next hop and the outgoing label for that next hop. If an LSR is the ingress or egress of an LSP, the NHLFE also specifies the actions for manipulating the label stack. NHLFEs may also contain additional state information related to the LSP; for example, hop count and data link encapsulation to use when transmitting the packet. LSRs use the label-switching table for forwarding labeled packets. When a packet arrives, an LSR finds the corresponding NHLFE for the incoming label by performing a look up in the label-switching table The LSR then replaces the incoming label with the outgoing label and forwards the packet to the interface specified in the corresponding NHLFE [114].

6. WHY MULTIPROTOCOL?

In MPLS Forwarding is performed at Layer 2. Network layer header's of the packets are encapsulated without any encoding. This means label encapsulation can be done with any network layer protocol. For example label switching can be used with IP as well as IPX. This feature makes MPLS multiprotocol according to the network layer protocols.

Furthermore MPLS can operate with any virtually link layer protocol. This makes MPLS multiprotocol according to the ink layer protocols.

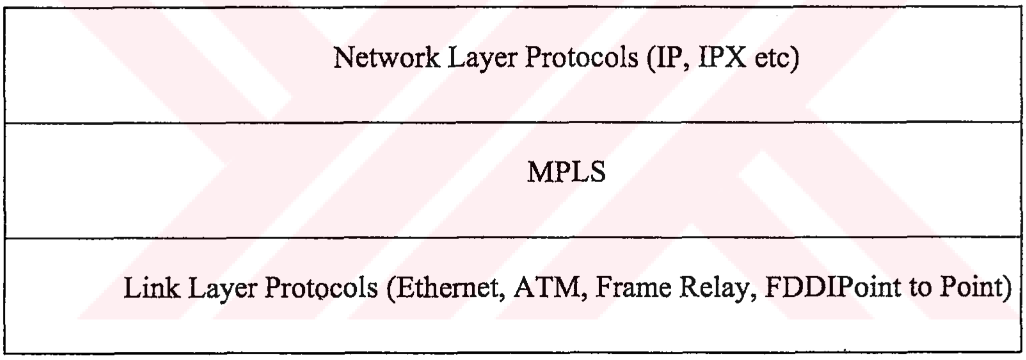


Fig. 3: MPLS in OSI Network Layers

7. LABEL DISTRIBUTION MANAGEMENT

a. Label Assignment and Distribution

In the MPLS architecture, the decision to bind a particular label L to a particular FEC F is made by the LSR, which is DOWNSTREAM with respect to that binding. The downstream LSR then informs the upstream LSR of the binding. Thus labels are "downstream-assigned", and label bindings are distributed in the "downstream to upstream" direction. If an LSR has been designed so that it can only look up labels that fall into a certain numeric range, then it merely needs to ensure that it only binds labels that are in that range [11].

b. Attributes of a Label Binding

A particular binding of label L to FEC F, distributed by Rd to Ru, may have associated "attributes". If Ru, acting as a downstream LSR, also distributes a binding of a label to FEC F, then under certain conditions, it may be required to also distribute the corresponding attribute that it received from Rd [11].

8. LABEL DISTRIBUTION PROTOCOLS

A label distribution protocol is a set of procedures by which one LSR informs another of the label/FEC bindings it has made. Two LSRs, which use a label distribution protocol to exchange label/FEC-binding information, are known as "label distribution peers" with respect to the binding information they exchange. If two LSRs are label distribution peers, we will speak of there being a "label distribution adjacency" between them. (N.B.: two LSRs may be label distribution peers with respect to some set of bindings, but not with respect to some other set of bindings.) The label distribution protocol also encompasses any negotiations in which two label distribution peers need to engage in order to learn of each other's MPLS capabilities The architecture does not assume that there is only a single label distribution protocol. In fact, a number of different label distribution protocols are being standardized. Existing protocols have been extended so that label distribution can be piggybacked on them. New protocols have also been defined for the explicit purpose of distributing labels [11].

a. Label Distribution Protocol (LDP)

LDP is the first Label Distribution protocol for distribution of labels to support hop-by-hop routing Two LSRs that use LDP to exchange label/FEC mapping information are known as LDP peers [114].

b. Constraint-Based LDP (CR-LDP)

CR-LDP is developed for traffic engineering by adding explicit route and resource reservation extensions to LDP. An explicit route is determined in the label request message. CR-LDP both supports loose and strict explicit routes. In strict explicit route, the nodes along the LSP cannot be changed. All the routers on the way of the LSP is determined with an IP address. On the other hand loosely explicit route allows some routes that become unavailable during an LSP connection, to change. Some nodes (routers) can be defined as abstract. An abstract node consists of routers. The nodes except the nodes between two known nodes may change according to a network failure.

The CR-LDP LSP setup is shown in Fig. 4. The ingress node LSR A initiates a LSP between LSR A and LSR C. LSR A sends a label request message to LSR B, containing an explicit route request to LSR C. Then LSR B sends the label request to LSR C. When the request message arrives LSR C it understand that it is the egress node and sends the label-mapping message to LSR A over LSR B. resource reservation is allowed in CR-LDP.

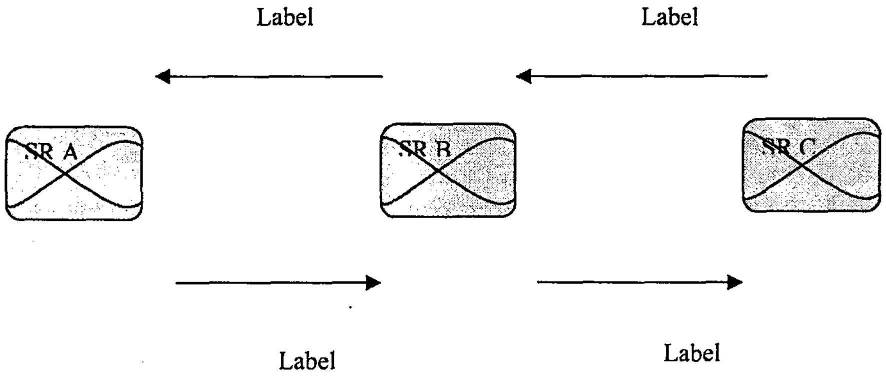


Fig. 4:CR-LDP initiation

(1) Path Preemption and Priorities

If the requested resources cannot be reserved over an LSP, the LSP may reroute and modify the existing LSP. This is called preemption of the existing LSP. There are two parameters for this purpose: setup priority and holding priority. The setup priority and holding priority reflect the preference for adding a new LSP and holding an existing LSP. A new LSP can preempt an existing LSP if the setup priority of the new LSP is higher than the holding priority of the existing LSP. The setup and holding priority values range from 0 to 7, where 0 is the priority assigned to the most important path, or the highest priority. [114]

(2) LSP Modification Using CR-LDP

Modification of the bandwidth and other parameters of an established Constraint-based Routed Label Switched Paths (CR-LSP) using CR-LDP without service interruption is possible. After a CR-LSP is set up, its bandwidth reservation may need to be changed due to the new requirements for the traffic carried on that CR-LSP. LSP modification also supports reroute of the existing LSP. For modification the LSP must be already set up and active. That is, modification is not defined nor allowed during the LSP establishment or label release/withdraw phases [18].

c. Extensions to RSVP for LSP Tunnels (RSVP-TE)

RSVP is originally designed for resource reservation for Integrated services especially for multimedia applications such as real time applications (video conference, voice conversations), remote video. RSVP-TE is developed by adding label distribution and explicit route extensions.

(1) LSP Tunnel

Original RSVP protocol was designed to setup reserved paths across IP networks like LSP. But it is not enough to setup LSP with RSVP Because In original RSVP a reserved path is always associated with particular destination and transport-layer protocol, and forwarding packets over the intermediate nodes is performed due to the IP header. In contrast, with an LSP set up by RSVP-TE, the ingress node of the LSP encapsulates the IP packets according to heir FEC. The intermediate nodes over the LSP just looks up the label and just forward the packet according to the label and they do are not related to the IP packet encapsulated within the label. To determine this feature, an LSP in the RSVP-TE specification is referred to as an LSP tunnel [114]. RSVP-TE LSP creation is shown in Fig. 5.

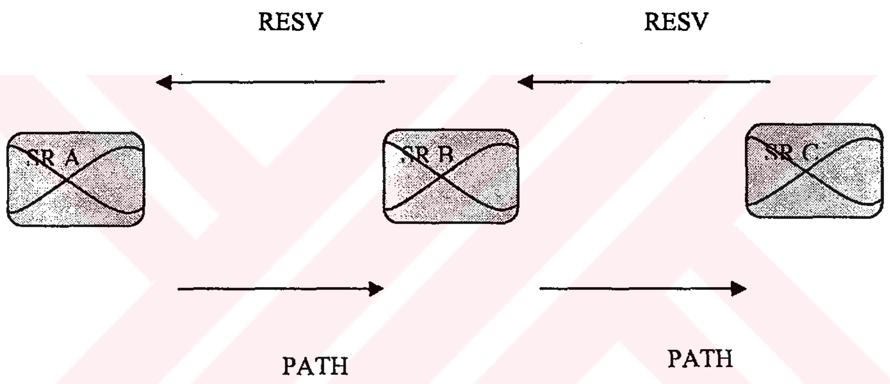


Fig. 5: LSP Tunnel setup

G. MOBILITY SUPPORT FOR MPLS

Mobility issues in original MPLS are not supported. There are different proposals to support user mobility.

In [4] Wireless MPLS (WMPLS) is introduced. There are two main factors that influence the design of WMPLS: wireless medium and terminal mobility. WMPLS, which tackles the issues introduced by these factors, resembles the original MPLS protocol. It introduces loosely explicit routing concept where the wireless mobile part of the network is defined as an abstract node. An abstract node is a group of nodes in an LSP [9] and the packets of a data flow can be forwarded through any set of the nodes in an abstract node. WMPLS uses the abstract node approach to solve the handoff problem. Mobile terminals can handoff among the nodes of an abstract node during an LSP connection [4].

In [23] A hierarchical architecture using label switched packet forwarding to enhanced traffic control and accommodate service differentiation supporting wireless users is defined. In this architecture the network consists of base stations, connected to label switching nodes (LSN), along with the components of the standard MPLS network, such as label switching routers. LSN provides services to base stations and supports fast handoff and location management mechanisms. The hierarchical architecture extends label-switched paths up to the base station, which in turn supports wireless users. The architecture proposed here is independent of the underlying radio technology, and the base station terminating the label switched paths requires functions for forwarding of packets between wireless terminals and LSN interfaces Location management and routing is realized by adapting the current interior protocols for mobile users. The path is calculated according to the link state routing protocols and LSPs are created with signaling protocols [23].

In [19] Hierarchical Mobile MPLS (H-MPLS) supporting delay sensitive applications over wireless MPLS is introduced. In a hierarchical mobile MPLS architecture, the mobile host determines whether it is at home or in a foreign domain when it receives agent advertisement messages, broadcasted by the Foreign Agent (FA). If the mobile host determines that it is in a foreign domain, the mobile host acquires a temporary Care of Address (COA) from the FA and sends a registration request to the FA. The FA will forward this Registration Message to the FDA instead of the HA of the mobile host by using normal IP forwarding. Then the FDA will forward this Registration Message to the HA of the mobile host. When the HA gets the Registration Message and knows the IP address of the FDA, it will send a label request using LDP protocol to the FDA with the IP address of the FDA as the FEC. The FDA replies with an LDP label mapping message back to the HA and sends a label request to the FA of the subnetwork in which the mobile host currently is located. When the label mapping message arrives at Home Agent (HA), the LSP from HA to the FDA is established. Similarly, the FA replies with an LDP label-mapping message back to the FDA and when this message arrives at FDA, the LSP from FDA to the FA is established. Then, the HA will search the label table and find the row with MH's home address as the FEC and change the outgoing port and out label to the same values as the ones for the LSP from HA to the FDA. Finally, the HA sends a registration reply to the FDA along the LSP from HA to FDA and the FDA will forward this registration reply to the FA along the LSP from the FDA to the FA. When the FA receives the registration reply, it adds a new row in its label table and puts the incoming registration reply's label value and port number into the "In label field" and "incoming port field" of the table. Fig. 6 shows the registration procedure for mobile hosts in Hierarchical Mobile MPLS.

In [49] Micro Cell Mobile MPLS (MM-MPLS), which is an incorporation of MPLS forwarding, localized signaling, and soft-state location management scheme is introduced and a comparison of the two approaches, Hierarchical Mobile MPLS and Micro Cell Mobile MPLS is given management. Similar to H-MPLS, MM-MPLS also hides the movements of a mobile host, moving within a domain, from its remote home agent. Two levels of foreign agents, Foreign Domain Agent and Foreign Agent, are comprised in each MM-MPLS domain. RSVP-TE, is employed as the signaling protocol to establish and manage the LSPs. This means it must refresh the states of each LSP periodically. Otherwise, the states are removed automatically when the timers associated with them expire. This property of RSVP-TE allows us to apply the soft-state location managements on the mobile hosts within the MM-MPLS domain. The home address of the mobile host, instead of the address of the current FA in H-MPLS, is used as Forwarding Equivalence Class (FEC) to establish the LSP from the FDA to the FA. This allows, most likely, an intermediate LSR between the previous FA and FDA, instead of the FDA defined in H-MPLS, to handle the local location update message, and steer the traffic to the current FA when the mobile host handoffs to another network that belongs to the same domain. Thus, the FDA may even not notice the movement of the mobile host in MM-MPLS. This has the benefit of further reducing the delay and jitter during the handoff period of the mobile host, when compared to H-MPLS, since it will take a shorter period of time for the location update message to reach an intermediate LSR than the FDA. A simulation-based study of the two different approaches, H-MPLS and MM-MPLS, supporting micro mobility is presented in the paper.

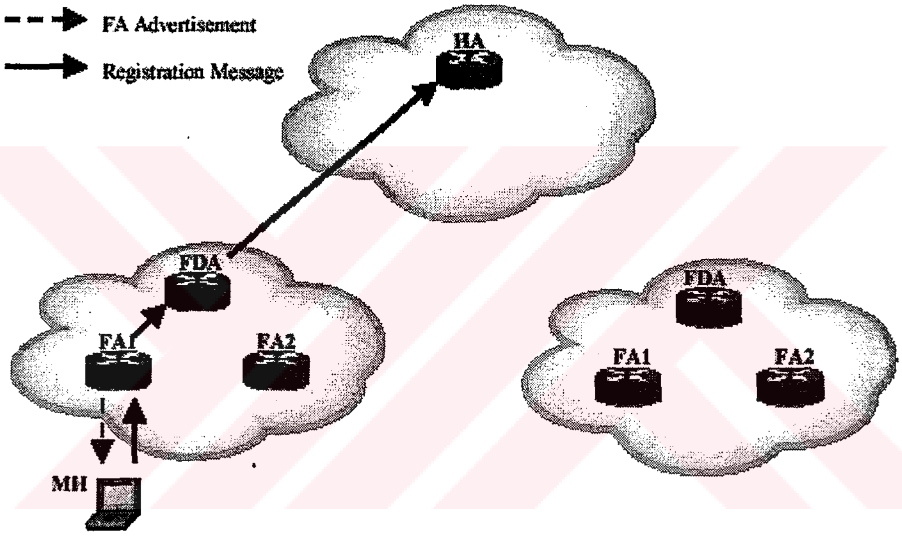


Fig. 6. Hierarchical Mobile MPLS Architecture [19]

III. LOCATION MANAGEMENT AND RESOURCE RESERVATION FOR MOBILE NETWORKS

A. OVERVIEW

Wireless communications is growing incredibly nowadays. personal communications service (PCS) scribers are increasing at an exponential rate and will continue to increase in the near future. The next generation personal communications network (PCN) is being standardized as part of the international mobile telecommunications 2000 (IMT-2000) system, whose goal is to unify many diverse systems existing today (including PCS, two yaw paging, mobile satellite, etc) into a seamless radio infrastructure capable of offering a wide range of service. The current PCNs use a cellular architecture. The geographical coverage area is portioned into cells, each served by a base station. Mobile users and their terminals are connected to the network via the base stations. Cells can have different sizes: picocells are commonly used indoor environment; microcells are used within cities; macrocells are used in rural areas and cover highways. Smaller cells use less power for transmission and allow greater frequency reuse. Several base stations are connected to a base station controller, and a number of base station controllers are then connected to a mobile switching center. The connection of the base stations, base station controllers, and mobile switching centers, along with the radio links between the base stations and mobile terminals, form the access network. [108]

Mobility management enables telecommunications networks to:

- Locate roaming MTs for call delivery
- Maintain connections with MTs that change their point of attachment.

Location management tracks and locates the MT for the delivery of incoming calls, while handoff (or handover) management allows a call in progress to continue as the MT changes channels or moves between cells. In location management, the MT periodically performs location registration (location update) to explicitly notify the network of its new access point and store changes to its user location profile. Then, when incoming calls arrive, the network performs call delivery by querying the user profile to deliver the calls to the current call location of the MT. Location management protocols deal with querying and storing information in location databases and sending paging signals to locate the user within the network [117].

It is well known that there is a trade-off between the cost of location update and paging. If the mobile terminal updates its location whenever it crosses a cell boundary, the network can maintain its precise location, thus obviating the need for paging. However, if the call arrival rate is low, the network wasres its resources by processing frequent update information and the mobile terminal wastes its power transmit the update signal [108].

B. LOCATION UPDATE

The network has to know the exact location of each user to so that it can connect a fix host or another MT that is calling it. If the network doesn't know the location of an MT when a call arrives it searches the MT. This is called paging. An MT has to make location update according to different approaches explained above. Increasing the location update reduces paging. Location update is performed by each MT. This procedure begins with an update message sent by the mobile terminal over the uplink control channels, which is followed by some signaling procedures, which update the database. Location update algorithms can be divided into two groups: static and dynamic. In a static algorithm, location update is triggered based on the topology of the network. In a dynamic algorithm, location update is based on the user's call and mobility patterns. [108]

1. SELECTIVE LOCATION UPDATE

An MT may move among. For example an office worker goes to his/her job in the morning and return in the evening. He is in his office all the day. This means he enters some LAs just for short periods on his/her way to and from work. In this model the residence time and transition probabilities is needed [117] and this can be estimated from the behavior of MTs according to their movements for a period of time from the system databases.

2. MOVEMENT BASED LOCATION UPDATE

In this model location update is performed according to the number of cells (*d*) that the mobile user visits. The value *d* is called location update movement threshold. When an incoming call arrives, the network pages (searches) the MT within a location area including all cells within a distance *d* from the last registered location of the called mobile user. [95]

3. TIME BASED LOCATION UPDATE

Location update is performed in a period of time t in this model. The location area is calculated according to the movement behavior of the mobile user during location update period. This does not mean that the location information is saved.

4. DISTANCE BASED LOCATION UPDATE

Location update is realized when a mobile user moves a distance (in terms of number of cells) since the last location update process.

5. PREDICTIVE DISTANCE BASED LOCATION UPDATE

In Predictive Distance Based Location Update scheme [119], the network predicts the MT's future location according to the information that the MT sends. This is performed by location update process. The MT sends its location and speed while sending the location update signal. The network makes the prediction information available for the MT. If the MT goes somewhere else that the network has calculated, the MT updates its new location.

6. STATE BASED LOCATION UPDATE

In this model location update is performed when the state of the MT changes. The state can be different according to the scheme; the number of cells that the MT moves among, the time elapsed since the last location update etc.

C. PAGING

We can define paging as the process of being searched and designated the exact location of the MT by the network. The purpose of both paging and location update schemes is similar, reducing the bandwidth consumption. All the MTs listen this message and just the MT called answers. The network needs paging when a call arrives to an MT that the network does not know its exact location. Paging costs bandwidth consumption. This means that if the network knows the future location of an MT, both the location update and paging costs reduce. Generally there is a trade-off between location update and paging.

1. BLANKET PAGING

In blanket paging all cells in an LA are simultaneously paged upon a call arrival. The least paging delay is guaranteed by the blanket polling technique. This scheme has high paging cost.

2. SELECTIVE PAGING

The selective paging is an alternative to the blanket polling and reduces the paging cost, but increases the paging delay. In the selective paging, the location of the MT is predicted based on its location probability, and cells are paged sequentially starting from the cells where the MT is most likely to be present.

3. VELOCITY PAGING

In the velocity paging, the system calculates the maximum distance that a mobile can travel based on its average velocity and the last registration time, and pages the cells, which are within this distance from the last registered cell.

4. SEQUENTIAL PAGING

In this scheme the location areas are partitioned into clusters. The number of clusters in a location area guarantees that the system can page them sequentially without exceeding the given delay bound.

D. MOBILTY PREDICTION

1. OVERVIEW

The next generation wireless systems will consist of smaller cells for the advantage of efficient use of frequency and bandwidth. Thus location management and handoff will be paramount and handover process have to be done fast [8]. To accomplish this, location update and resource reservation must be done as fast as possible because users' mobility implies that reservation has to be realized at all places that an MT may visit, during the life time of a data flow. These are location management issues. There are time based [92,93,107], movement based [92,94,95,106], distance based [92,96,97,109,110], direction based [98], selective [99,100], profile based [1,38,74,104,105,112,113,116], hierarchical partioning [101], load balancing [102], intersection-oriented [111]. In [118] a location-tracking mechanism that consists of intersystem location update and intersystem paging is introduced. If the future location of the MT and expected cell entry time can be predicted during the lifetime of a connection, resources can be reserved just in the right cell where the MT will visit next. The problem is the mobility of the MTs. If the future location is known resource is used most effectively. But this is not easy to predict someone's movement.

The goal advantage of the profile based location update scheme is to reduce the update cost by taking advantage of the user mobility pattern. The network maintains a profile for each user, which includes a sequential list of the LAs the user is most likely to be located at in different time periods. This list is sorted from the most to least likely LA where a user can be found. When a call arrives, the LAs on the list are paged sequentially. As long as the mobile terminal moves between LAs within the list, no location update is necessary. Location update is performed only when the mobile terminals moves to anew LA not on the list. The list may be derived from the user's movement history [108].

2. USER MOBILITY PATTERN (UMP)

a. Overview

The user mobility pattern (UMP) scheme [1] is introduced for location update and paging in wireless systems where mobile terminals (MTs) maintain their history data in a database called user mobility history (UMH). During a location update, an MT derives a UMP from its UMH and registers it to the network. Unless the MT detects that it has moved out of the registered UMP, it does not perform any other location update. On the other hand, cells are paged selectively according to the cell entry times in the registered UMP upon a call arrival for the MT. The related data structures and the protocols for the UMP scheme are presented in [1].

Mobile subscribers usually follow a limited number of mobility patterns in their daily lives. For example, people generally take almost the same path and same time to go to work every day. In the UMP scheme, an MT collects the data related to these patterns in a UMH, and predicts its current UMP based on the collected data. During location updates, the expected UMP is registered to the network by the MT.

A UMP is a list of cells expected to be visited starting from a given time according to the mobility history of a mobile, and is made up of a number of nodes that have two fields, namely, cell identification (cell id) and expected cell entry time. A cell id and expected cell entry time pair identifies a node, which is unique in a UMP. However, a UMP may have two different nodes with the same cell id and different expected cell entry time values. Similarly, two different UMPs may have nodes with the same cell id because the same cell may be visited in multiple UMPs or multiple times in the same UMP. For example, one may go to work passing through a cell in the morning. The user may come back home passing through the same cell in the evening. Moreover, the same person may go somewhere else through the same cell at noon in the weekends. In these cases, an MT visits the same cell in different UMPs or in the different phases of the same UMP, i.e., the subscriber may have a single UMP that represents his way both to and from his work where some cells are visited twice at different times within the same UMP.

In Fig. 7, we show an example UMP with four nodes. The cell in the first node is expected to be entered starting from 1080 according to the UMH of the MT that registered this UMP. Expected cell entry time is the representation of daytime in minutes with a 24-hour scale starting at midnight, i.e., 06:00 p.m. is represented as 1080. Cell 12 is expected to be visited twice in Nodes 2 and 4 according to the UMP shown in Fig. 1. The difference between these two nodes is the expected cell entry times. Assuming that this UMP represents the pattern that a subscriber follows to go to work and to come back home, that user passes through Cell 12 once in the morning and once in the evening. Note that a UMP is valid after its registration as long as another UMP is not registered, which means that a UMP can span many days. For instance, a subscriber may go to work and nowhere else during some days where he does not need another location update when the UMP representing the path to work is registered before the beginning of this period.

Node 1	cell id	8
	expected cell entry time	1080
Node 2	cell id	12
	expected cell entry time	480
Node 3	cell id	21
	expected cell entry time	510
Node 4	cell id	12
	expected cell entry time	1050

Fig. 7. An example UMP. [1]

The UMP scheme outperforms the time-based and movement-based location update schemes as well as the blanket, selective, and velocity paging schemes according to experimental results. It is an effective location management scheme. The algorithms, the protocols and the results from the performance evaluation for the UMP scheme are presented in [1].

b. User Mobility History (UMH)

UMPs are derived from UMH, which is a data structure where the mobility history of an MT is stored. Each MT manages its own UMH, which is composed of a limited number of records with the following fields:

- flag,
- UMP identification (UMP Id),
- cell Id,
  - expected entry time interval (EETI),
  - the earliest entry time t_{ee},

- the latest entry time t_{le},

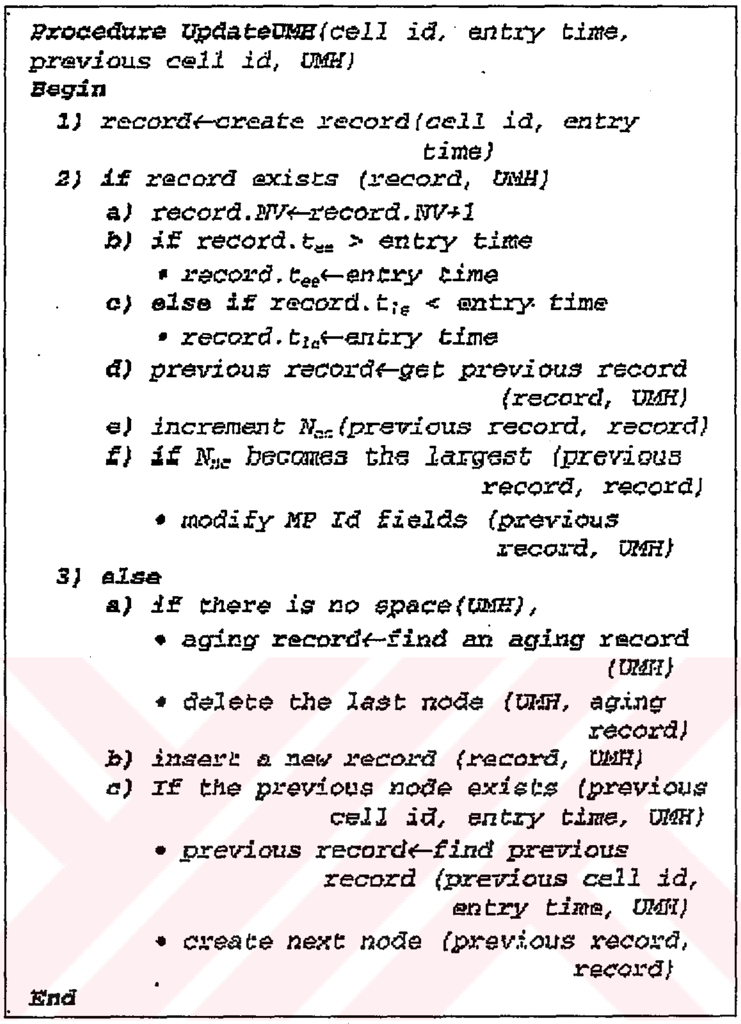
- . number of visits (NV), and
- . next node structures (NN)
  - - index of the next node I_{nc},
  - - number of times that the next node is visited N_{nc}.

The records are the parameters, which are related to nodes of a UMP.

When an MT enters a new cell, it starts the procedure shown in Fig. 8 to update its UMH. This procedure is composed of "if" statements, and a UMH has generally less than 100 records as explained in [1]. Therefore, this procedure generally requires a limited number of comparisons (i.e., less than 100), and one or two update operations, which makes it simple enough to run on an MT hardware.

If an MT enters a new cell, its UMH probably holds a record for that new cell. Therefore, the UMH is searched for the record of the new cell first. If a record with the same cell id is found and the expected entry time is not larger than the maximum entry interval I_{max}, the N_{nc} field in the current node record is incremented. The previous node is the node from which the MT crosses to the current node.

The expected entry time interval, which is the difference between tee and tle, must be lower than I_{max}. The minimum t_{ee} and maximum t_{le} values can be determined as lower and upper control limits by using techniques like statistical quality/process control where the difference between these two values gives I_{max}. If the cell entry time is not between t_{ee} and t_{le} when a new cell is entered, the expected cell entry time interval for the node is also modified (i.e., t_{ee} is replaced if the cell entry time is lower than tee or t_{le} is replaced if the cell entry time is higher than t_{le}).



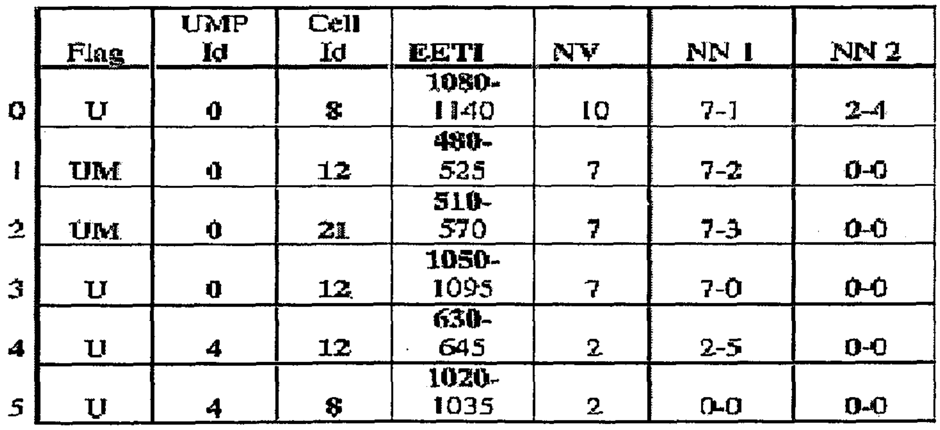
```
Procedure UpdateUMH(cell id, entry time,
previous cell id, UMH)
Begin
  1) record←create record(cell id, entry
                                   time)
  2) if record exists (record, UMH)
     a) record.NV←record.NV+1
     b) if record.$t_{ee}$ > entry time
          • record.$t_{ee}$←entry time
     c) else if record.$t_{le}$ < entry time
          • record.$t_{le}$←entry time
     d) previous record←get previous record
                                   (record, UMH)
     e) increment $N_{nc}$(previous record, record)
     f) if $N_{nc}$ becomes the largest (previous
                                  record, record)
          • modify MP Id fields (previous
                                   record, UMH)
  3) else
     a) if there is no space(UMH),
          • aging record←find an aging record
                                       (UMH)
          • delete the last node (UMH, aging
                                      record)
     b) insert a new record (record, UMH)
     c) If the previous node exists (previous
                    cell id, entry time, UMH)
          • previous record←find previous
                  record (previous cell id,
                          entry time, UMH)
          • create next node (previous record,
                                      record)
End
```

Fig. 8. The UMH update procedure. [1]

When the maximum value for an NV field is reached, all NV fields are decremented with the exception of the records where the NV value has already become 0. The NV field is used to find the aging UMPs. When the UMH is full and a new record is needed to be inserted, one of the records with the lowest NV is replaced with the new record.

When the Nnc field pointing to the new cell in the previous record becomes the largest Nnc of its record after update, it indicates that the UMP changed starting from the previous node. For example, if we modify next node 2 structure of Node 0 as 8 ¡ 4 (i.e., Nnc à 8 and Inc à 4) in Fig. 9, the nodes in the UMP starting from Node 0 become 4 and 5. In this case, we must modify the UMP Id fields of Nodes 4 and 5 as 0, and the UMP Id fields of Nodes 1, 2, and 3 as 1. Because Node 0 is the member of both UMP 0 and 4, and after this modification, the probability that the MT crosses from Node 0 to Node 4 becomes higher than the probability that it crosses from Node 0 to Node 1.

If there is no record for the node entered in the UMH (i.e., there is no record that has the same cell id or the expected entry time interval becomes larger than I_{max} after its modification of the cell entry time), the data about the new node is inserted into the first unused record, unless UMH data structure is full. Otherwise, the record with the minimum NV is found, and the next node links are followed until the next node with an NV greater than the minimum NV or a record with no next node data is reached. This ensures that we select the record related to the last node of one of the least used UMPs. The next node field in the previous node record is also modified such that it does not point to the node to be replaced anymore. Then, this record is used to store the data related to the new node. If the previous node is known and there is a record for it in the UMH, the links between the previous and the new node are also created after the data about the new node is inserted into the UMH.



	Flag	UMP Id	Cell Id	EETI	NV	NN 1	NN 2
0	U	0	8	1080-1140	10	7-1	2-4
1	UM	0	12	480-525	7	7-2	0-0
2	UM	0	21	510-570	7	7-3	0-0
3	U	0	12	1050-1095	7	7-0	0-0
4	U	4	12	630-645	2	2-5	0-0
5	U	4	8	1020-1035	2	0-0	0-0

Fig. 9. An example UMH. [1]

c. Location Update By Using The UMP Scheme

An MT starts a location update process when one of the following happens: when the MT is turned on, or returns back to the coverage area of the network, and . if a new node is entered, which is not a member of the UMP registered in the last location update. During a registration process, an MT first finds the UMP that it follows. Knowing the current cell and the current time is enough to determine the current UMP. By searching through the UMH, the MT finds the node according to its current cell and time. For example, if the current time is 540 (i.e., 9:00 a.m.) and the current cell id is 12 when an MT is turned on, its current UMP will be UMP 0 for the UMH given in Fig. 9 because 540 is within the expected cell dwell time interval for Cell 12 in UMP 0. At the end of this search, if the MT cannot find a UMP for its current situation, it starts to create a new UMP until it enters an existing UMP. During the time when it cannot register a UMP, a static or a dynamic location update technique may be used. In other words, when an MT does not have a UMP in its UMH for its current location, then any other location update scheme can be used. After finding its current UMP, the MT creates a UMP registration message, which has the structure illustrated in Fig. 10. The first field of this message is a header and consists of two subfields, namely the flag and number of nodes fields. The flag field is a single bit, and the network uses it to interpret the content of the message. If this bit is 0, it indicates that this message is related to a UMP, which has been previously registered to the network. In this case, the number of nodes field gives the number of nodes in the registration message. The network first modifies that many nodes of the previously registered UMP and uses the modified UMP as the current UMP of the MT. If the number of nodes field is 0, it indicates that there is no need to modify the UMP before using it. When the flag bit is 1, the meaning of the message depends on the content of the number of nodes field. When the flag bit is 1 and the number of nodes field is greater than 0, either an existing UMP is replaced by a new one or UMP, which was not registered before, is sent to the network. When the flag bit is 1 and the number of nodes field is 0, it indicates that this registration message is for another location update technique. The second field in a registration message is the UMP Identification (UMP Id) field. It may also be the location area identification when the system is in the static location area management mode. The fields for the nodes in the UMP follow the UMP Id field. Each of these fields is composed of two subfields, namely the cell identification and the expected cell entry time. When this structure is created, it is sent to the network in the payload of a registration message. Note that the network keeps the registered UMPs and modifies them by the registration messages coming from MTs. Therefore, an MT reports only the modifications that are made in the nodes of a UMP since its last registration. It is also possible that the network discards the previously registered UMP of an MT as soon as the MT registers a new UMP. In this case, the MT must send all of the nodes in the UMP as registered to the network for location updates. Hence, when the network keeps only the current UMP and forgets the previously registered UMPs, the average size of registration messages increases. Also, note that our solution is adaptive. If an MT enters a region where it has not been before, it cannot find a UMP in its UMH for the current location. In this case, we can use another location update technique.

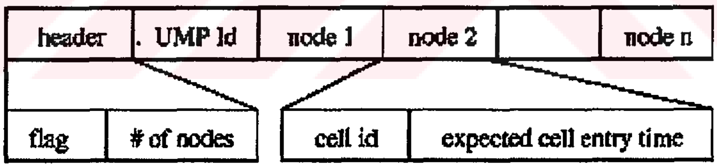


Fig. 10. The structure of a location update message.

d. Paging By Using The UMP Scheme

When a call arrives for an MT, the cells registered by the MT in the UMP are paged sequentially starting from the cell where the called MT is most likely to be at the call arrival time. This cell is determined by searching the nodes in the UMP registered by the called MT in order to find the last node which has an expected entry time lower than or equal to the current time, i.e., there is either no node after the last node or the next node has an expected entry time greater than the current time. For example, if a call arrives for the MT which has registered the UMP in Fig. 9, and the current time is 10:30 a.m., Cell 21 is paged first. If the MT is not found in the first paged cell, the cell with the closest cell residency interval to the current time is paged next. For example, Cell 12 is paged second, and Cell 8 is paged third. One cell at a time search is carried until either a response from the MT is received or a paging delay bound is reached. If the MT is not found in one of the paged cells until the paging delay bound is reached, all remaining cells in the UMP are paged simultaneously. If the MT is still not found, it indicates that the MT either left the coverage area of the network or turned off without deregistration (i.e., sudden death of power supply) because MTs are supposed to register a new UMP whenever they move out of the registered UMP. In such a case, the called MT is deregistered and recorded as turned off until it registers a UMP. As explained in the location update process in the previous section, whenever the MT returns to the coverage area or is turned on, it starts a location update process.

E. RESOURCE RESERVATION FOR MOBILE NETWORKS

Resource reservation is a promising to provide QoS for real time applications. There are two kinds of reservations First Immediate and advance reservation [34]. Resource reservation is very important for cellular networks since the resource is limited. And it is very important to support bandwidth to moving mobile users when they have open connection. Lots of studies have been made about resource management for mobile networks [35,36,38,75-91]. There are different approaches for reserving the resources. [35] compares five different schemes (named CHOI, NAG, AG, BHARG, NCBF respectively) for bandwidth reservation and admission control in QoS sensitive cellular networks. Probabilistic resource estimation and semi-reservation for flow oriented multimedia wireless networks is defined in [36], an optimization for adaptive bandwidth reservation in wireless multimedia networks using full information on near future traffic arrivals and an event-sensitive rescheduling method [36]. In [38] Reservation is realized according to user mobility prediction.

F. PREDICTED RESOURCE MENAGEMENT

Efficient resource use in mobile networks is very important for QoS requirements for MTs since limited bandwidth in wireless media. There are various resource management schemes and studies. Minimizing the amount of system resources must fulfill QoS. Generally you must select one of both in traditional approaches, either minimizing system resources or QoS requirements. If the user movements can be predicted both efficient use of system resources and QoS requirements of MTs can be maintained.

The *shadow cluster* concept can be used to estimate future resource requirements and to perform call admission decision in wireless networks. In [81], it is shown how BSs determine the probabilities that a mobile terminal will be active in other cells at future times, define and maintain shadow clusters by using probabilistic information on the future position of their MTs with active calls and predict resource demands based on shadow cluster information.

Consider a microcell wireless network system that can support MTs running applications which demand a wide range of resources. Users can freely roam within the networks coverage area and experience a large number of handoffs during a typical connection. The wireless network users expect good QoS from the system, e.g. low delays, small call dropping and packet loss probabilities. The wireless network must provide the requested level of service even if an active MT moves to a congested cell. In this case, the corresponding BS must provide the expected service even if this implies denying network access to new connection requests. Ideally, BSs should deny network access to certain connection requests only when it is strictly necessary. This constitutes a problem that could only be optimized if knowledge is available regarding the future movement and cell holding times of the active MTs in the wireless network, as well as the future movements and call holding times of the MTs with connection requests. As in related problems, solutions close to optimal can often be obtained by using knowledge of past events to predict future behavior.

The fundamental idea of the shadow cluster concept is that, every MT with an active wireless connection exerts an influence upon the cells (and their BSs) in the vicinity of its current location and along its direction of travel. As an active MT travels to other cells, the region of influence also moves, following the active MT to its new location. The BSs (and their cells) currently being influenced are said to form a shadow cluster, because the region of influence follows the movements of the active MT like a shadow. The shadow (and therefore the level of influence) is strongest near the active MT and fades away depending on factors such as the distance to the MT, current call holding time and priority, bandwidth resources being used and the MT's trajectory and velocity. Because of these factors, the shape of a shadow cluster is usually not circular and can change over time.

We say that the center of a shadow cluster is not the geometric center of the area described by the shadow, but the cell where MT is currently located. We also refer to this cell as the MT's current home cell. A bordering neighbor is a cell that shares a common border with shadow cluster's center cell. In contrast, a nonbordering neighbor cell, although being a part of the shadow cluster, does not share a border with the shadow cluster's center cell. In a manner consistent with the above definitions, we also use the terms current home BS, bordering BS and nonbordering BS. Conceptually, the number and the darkness of the shadows covering a cell reflect the amount of resources that the cell's BS needs to reserve in order to support the active MTs currently in its own and neighboring cells. With the information provided by shadow clusters, BSs can determine, for each new call request, whether the request can be supported by the wireless network. In practice, a shadow cluster is a virtual message system where BSs share probabilistic information with their neighbors on the likelyhood that their active MTs will move to neighbor cells (while remaining active) in the near future. Within the information provided by the shadow clusters, BSs project future demands and reserve resources accordingly. BSs reserve resources by denying network access to new call requests and by waiting for active users to end their calls.

The decision process for the acceptance of a new call request also involves a shadow cluster. Every new call request results in the implementation of a tentative shadow cluster. BS exchange information on their new call requests and decide, based on this and other information, which requests should be accepted and which requests should be denied. When implementing shadow clusters, it is important to consider that the amount of information shared among neighbor BSs should be limited, so that the wireline network is not overburdened with control messages. In practice, after a mobile call has been admitted, only a small amount of information needs to be shared.

When an active MT handoffs to another cell, the shadow cluster moves. After a handoff, BSs within the old shadow cluster are notified about this movement and the MT's new current BS has to assume the responsibility of supplying the appropriate information to the BSs within the new shadow cluster. BSs which were in an old shadow cluster that has just moved away must delete any entries corresponding to the active MT that established the shadow clustered free reserved resources if appropriate. BSs which become part of the influence region of a shadow cluster must be given appropriate information on the shadow cluster's active MT, such as respective QoS requirements, e.g., bandwidth demands, call dropping probabilities and any other useful information such as the wireless connection's elapsed time, for the establishment of the shadow cluster. In return for the communication and processing overheads involved, the shadow cluster concept improves some QoS requirements of the network, providing the ability to control the call dropping probability by establishing resource reservation requirements and a call admission policy. [81]

In [38] resource reservation is done according to user mobility that is predicted according to MTs' movements saved in a database. In this scheme both the next cell and cell entry time is predicted. The main difference between [1] [38] is that in [38] the user mobility pattern is saved in the network.

In [120] the BS of a cell calculates the required bandwidth to be reserved for anticipated handoffs from adjacent cells upon arrival of a new connection request. The mobility i.e. handoff behavior of each user is estimated using history of handoffs observed in each cell. Using this estimation, one can compute the bandwidth required to handle the handoffs that are predicted to occur within a specific time window. It also adaptively controls the window size depending on the observed handoff dropping events.

In [121], the BS considers not only incoming handoffs from adjacent cells, but also outgoing handoffs to adjacent cells. The BS then calculates the total required bandwidth in its cell for both handed-off and existing connections. Originally, this scheme is evaluated based on: 1) an exponentially-distributed time each mobile stays in a cell; and (2) the perfect knowledge about the mobility and lifetime of each user connection, i.e. known handoff and connection-termination rates. [35]

IV. SYSTEM DESCRIPTION

A. OVERVIEW

U-RSVP is a middleware between application and transport layers, and designed for the cellular networks where the UMP scheme [1] is used for the location management. An MT registers the cells expected to be visited, and the predicted cell entry times (*t*) during a UMP based location update [1]. For the active flows of the MT, an additional LSP is established for the cell expected to be visited Δ secs before the predicted cell entry time *t*. We call this LSP, i.e., the LSP other than the *actual* one, *advance LSP*. Δ is a system parameter, which is a short period of time in seconds. Δ can be determined by network administrators according to the QoS requirements and efficient resource using considerations. Thus an *advance* LSP lasts Δ seconds on the average. After a handoff occurs, the advance LSP for the new cell becomes the actual LSP and the actual LSP is terminated. LSPs are created according to the used signaling protocols.

There may be multiple advance LSPs for a single actual LSP based on the mobility of two endpoints of a flow. Only one of the endpoints or both endpoints may be mobile. In the case where both endpoints are mobile, mobiles may change their current cell almost at the same time, or one of the mobiles may change its cell long after the other one changes.

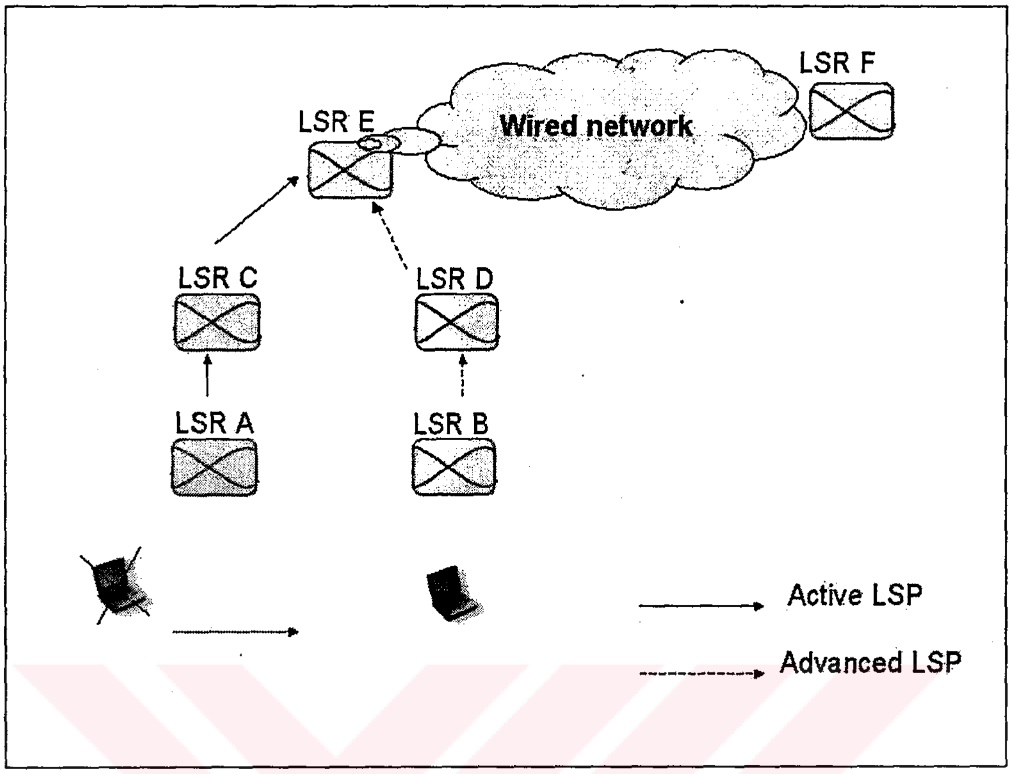


Fig. 11: Advanced LSP Creation.

B. ESTABLISHING AN ADVANCE LSP

For establishing an LSP, there are three-label distribution protocols as explained in Section 2.F.8. which are LDP, CR-LDP and RSVP-TE. LDP supports hop by hop label switching. It is not useful for us because it does not support end-to-end QoS. CR-LDP and RSVP-TE supports explicit routing. We can use both CR-LDP and RSVP-TE for creating LSPs.

An LSP is setup from ingress LSR. A modification on an LSP is needed when the mobile user handoffs to a new access network in an MPLS domain. The LSP setup is shown in Fig. 11. At the beginning the mobile user enters the access network by registering itself to LSR A. When a call arrives or the mobile user requests a call, the MT sends this request to the LSR A. Then the LSP is setup among LSR A, LSR C AND LSR E, LSR F. When the mobile user handoffs,

1. If the access network does not change the current LSP does not change either (A,C,E,F).

2.If the access network changes and will be LSR B, then there must be a modification in the current LSP (B,D,E,F). The SA informs the network before handoff and the change in the current LSP is calculated. The advanced LSP is performed with LSR (B,D,E,F).

1.SYNCHRONIZATION AGENT (SA) PROCEDURES

An SA is software, which the network collaborates the location management and call management systems of the network with, is used for the management of active and advance LSPs.

The next BS in the UMP of an MT should start the RSVP procedure on behalf of the MT before the handoff. The network sends an advance LSP request to the SA in the next BS along the UMP of the MT. This message includes also the identification of the source and destination nodes, the flow number and other QoS related parameters. The SA establishes an LSP by using these parameters. When the MT handoffs, it starts using the advance LSP, established by the SA in the new cell. The SA in the former cell terminates the actual LSP.

As soon as the network finds out that an advance LSP is not required anymore, e.g., when an MT registers a new UMP, the related SA is notified, and the SA terminates the LSP. Advance LSPs do not last longer than the maximum advance LSP duration t_{max}, which is a system parameter and a function of Δ. t_{max} is a function of Δ and UMP accuracy (Ψ). There must be a relation between t_{max} and Ψ. Because when Ψ is higher, it indicates that Δ is pretty much accurate. We can show this relation by $(1-\Psi) \times \Delta$. If we add this time to Δ then the formula becomes

$$t_{max} = \Delta + (1-\Psi) \times \Delta \qquad (1)$$

$$t_{max} = [1+(1-\Psi)] \times \Delta \qquad (2)$$

System administrators can use a variable θ to make t_{max} flexible according to their administrative purposes. Applying θ to Equation 2, we find Equation 3. Then the formula becomes:

$$t_{max} = [(1-\Psi) \times \theta + 1] \times \Delta \qquad (3)$$

The advance LSPs are terminated without waiting a termination request from the network unless they become the actual LSP before the end of t_{max}.

2.LSP SETUP WITH RSVP-TE

In conventional RSVP a PATH message should be sent by the source node to the destination. Therefore, we need a solution from the MT to the ingress node. An SA solves this problem. The SA procedures are explained above. The LSP Creation with RSVP-TE is shown in Fig. 12. The LSP setup is performed with LABEL_REQUEST message that is sent with RSVP PATH message. in RSVP LSP is created after the resource is reserved.

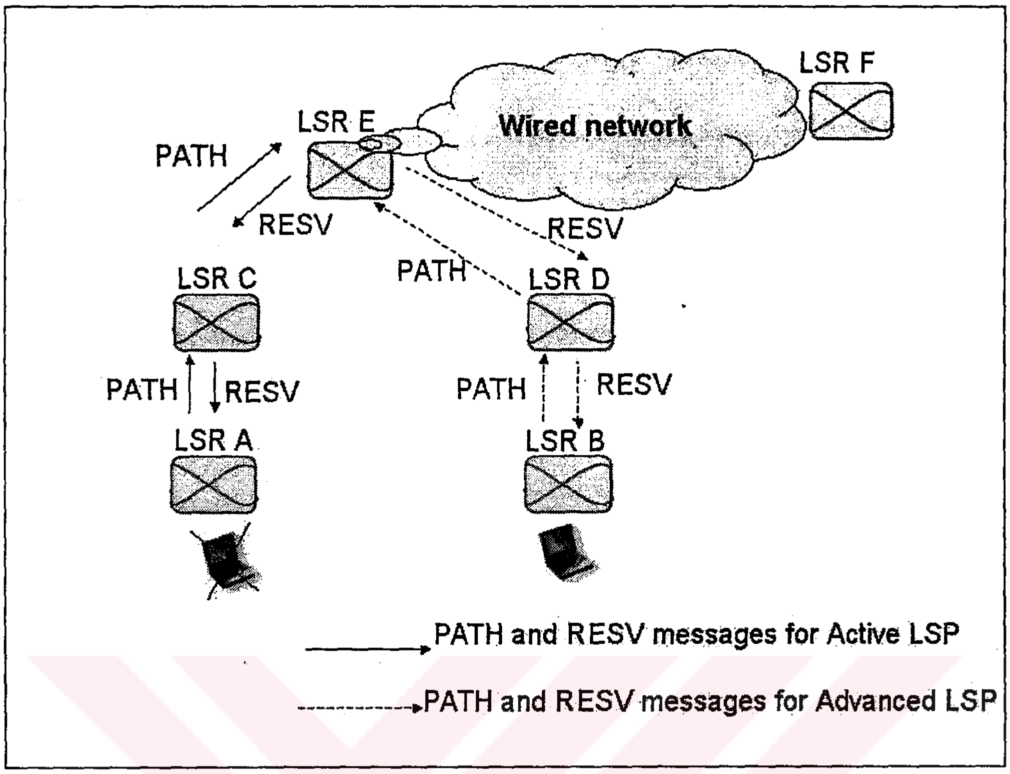


Fig. 12: LSP Creation with RSVP-TE

3.LSP SETUP WITH CR-LDP

The difference between RSVP-TE and CR-LDP during LSP setup is that in RSVP-TE first the LSP is created before resource reservation. On the other hand in CR-LDP reservation is performed after LSP is setup. The LSP setup for CR-LDP is shown in Fig. 13.

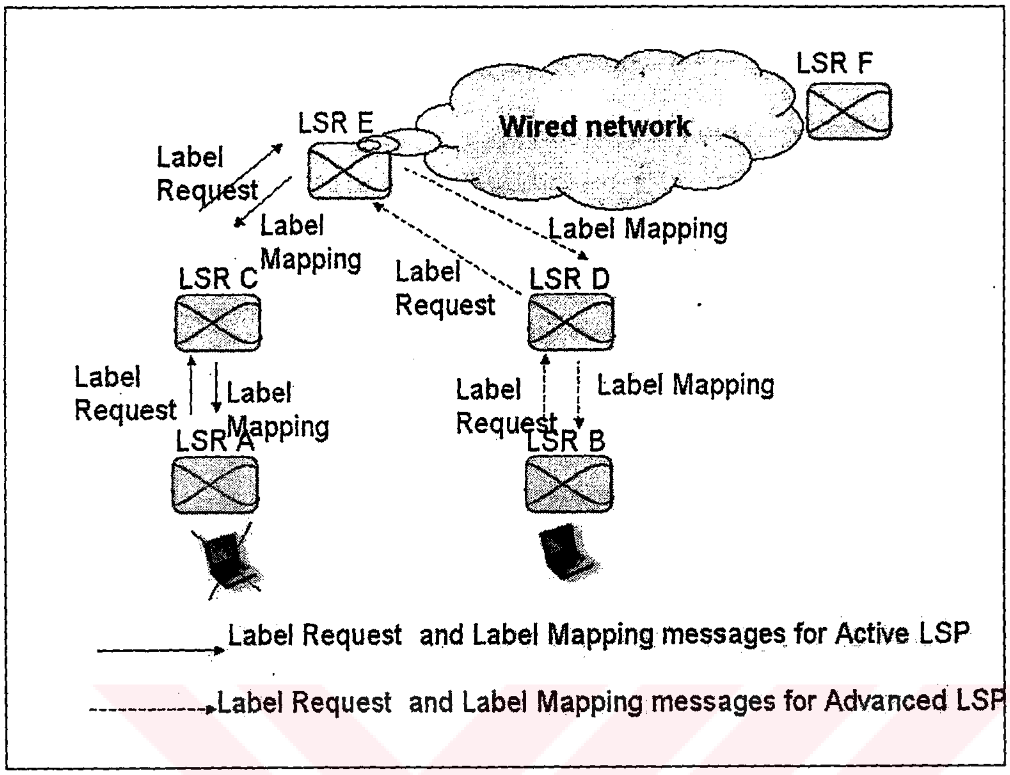


Fig. 13: LSP Creation with CR-LDP

C. CASE 1: ONLY ONE OF THE ENDPOINTS IS MOBILE

An example scenario for Case 1 where only one of the endpoints is mobile is shown in Fig. 10. In this scenario an MT establishes a connection with a fixed host (FH) when it is in Cell 1, i.e., the cell served by Base Station (BS) 1. The network knows the current UMP that the MT follows because the MT registers it. According to this UMP, the network predicts that the MT will make a handoff to Cell 2 at time *t*. At time *t*-Δ, RSVP is used to establish an advance LSP between Cell 2 and the FH for the MT. When the MT enters the Cell 2, it starts forwarding its flow through the advance LSP, and terminates the actual LSP. After this, the advance LSP becomes the actual LSP. Hence the resource reservation process is completed before the handoff, and thus the effects of mobility, such as slowdown in transmission rate, jitter and high connection preemption rate are alleviated.

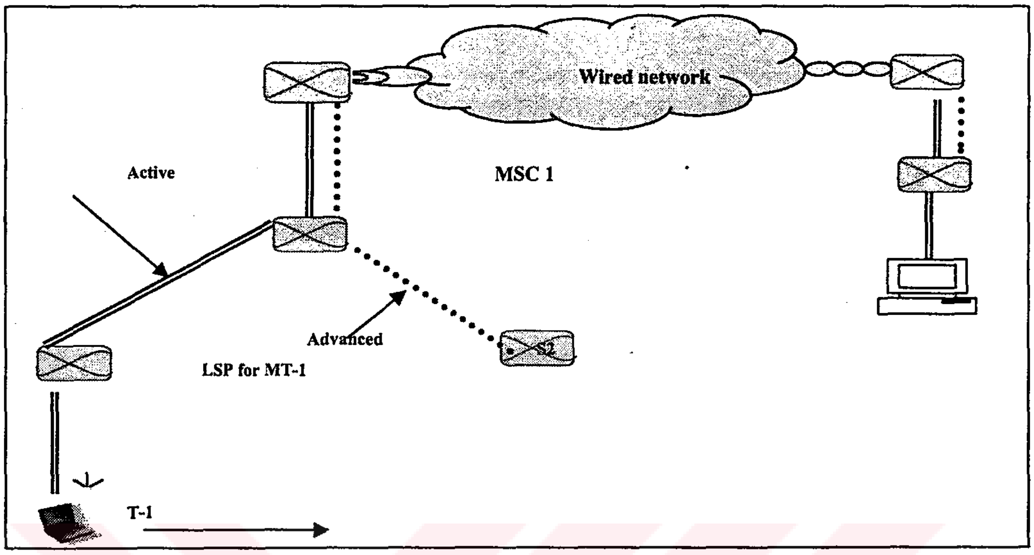


Fig. 14. Example scenario for the case where one of the end points is mobile.

D. CASE 2: BOTH OF THE ENDPOINTS ARE MOBILE

Fig. 11 shows an example scenario for Case 2 where a flow is established between two mobile hosts. When the flow is established, MT1 is in Cell 1, and MT 2 is in Cell 4. As in Case 1 the UMPs of MTs are known by the network because the UMP based location management is used. According to their UMPs, MT1 is expected to handoff to Cell 2 at time t_1, and MT2 is expected to handoff Cell 3 at time t_2. In this case, there may be up to three advance LSPs for the actual LSP at a time. The number of advance LSPs depends on the difference between t_1 and t_2.

If $|t_1-t_2| \geq \Delta$, there can be a single advance LSP at a time although both of the end points are mobile. In this case, the scenario can differ based on which MT makes handoff earlier as explained below:

If $t_1<t_2$, an advance LSP is established between BS2 and BS4 at time t_1-Δ. When MT1 handoffs, the actual LSP is terminated, and the advance LSP becomes the actual LSP. Then another advance LSP is established between BS3 and BS2 at time t_2-Δ. This advance LSP becomes the actual LSP when MT2 handoffs to Cell 3. If $t_2<t_1$, the first advance LSP is established between BS3 and BS1 at time t_2-Δ, and the second advance LSP is established between BS2 and BS3 at t_2-Δ.

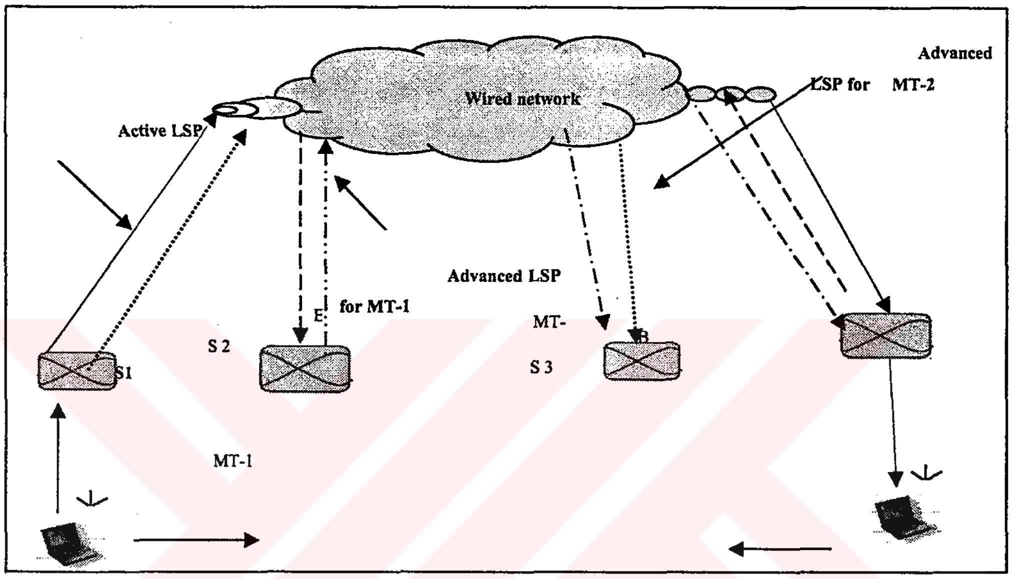


Fig. 15. Example scenario for the case where both of the endpoints are mobile.

When $|t_1-t_2| < \Delta$, three advance LSPs are present at time $\max(t_1-\Delta, t_2-\Delta)$. If $t_1<t_2$, the first advance LSP is established between BS2 and BS4 at time t_1- Δ, and the second and third advance LSPs are created between BS3 and BS1, and between BS2 and BS3 at time t_2- Δ. As the MTs handoff, the actual LSP is terminated, and an appropriate advance LSP, e.g., if MT2 handoffs first, the advance LSP between BS1 and BS3 is the appropriate advance LSP, becomes the actual LSP.

V. PERFORMANCE ANALYSIS

For performance evaluation, we simulated a cellular network with 40 cells. We assume that each cell has 40 bandwidth units (BUs) capacity. In our experiments we use the User Mobility Pattern scheme [1] as the location management technique where users follow their UMP according to a parameter called UMP accuracy which is the probability that an MT follows its current UMP when it handoffs. *Palicade Decision Tools Risk 4.5 for Excel* is used as the simulation tool. The simulation parameters are as follows: Number of users is distributed according to Poisson distribution with λ=20 and with a maximum value of 40. Call arrival rate is distributed according to Poisson distribution (λ changes according to the desired call arrival rate and the maximum value it can take is the number of users in the related). Average delay (unavoidable jitter) is 600 ms. Number of handoffs is distributed according to Exponential distribution with λ=5, the maximum value it can take is the number of the unblocked calls. Calls require either 1 BU or 4 BUs (30 % 4 BUs, 70% 1BU). Call duration is exponentially distributed with β=120. Velocity is distributed according to exponential distribution with β=25. Each MT derives a UMP from its UMH, and registers it during the location updates. We measure the call dropping and blocking rates for the varying UMP accuracy and node density, i.e., number of nodes in the simulation area, in our simulations. We also evaluate the overhead of U-RSVP for varying advance LSP duration Δ. On the other hand U-RSVP's jitter performance is measured for varying call duration, velocity and advance LSP duration Δ during handoff. We also measure our system sensitivity versus velocity.

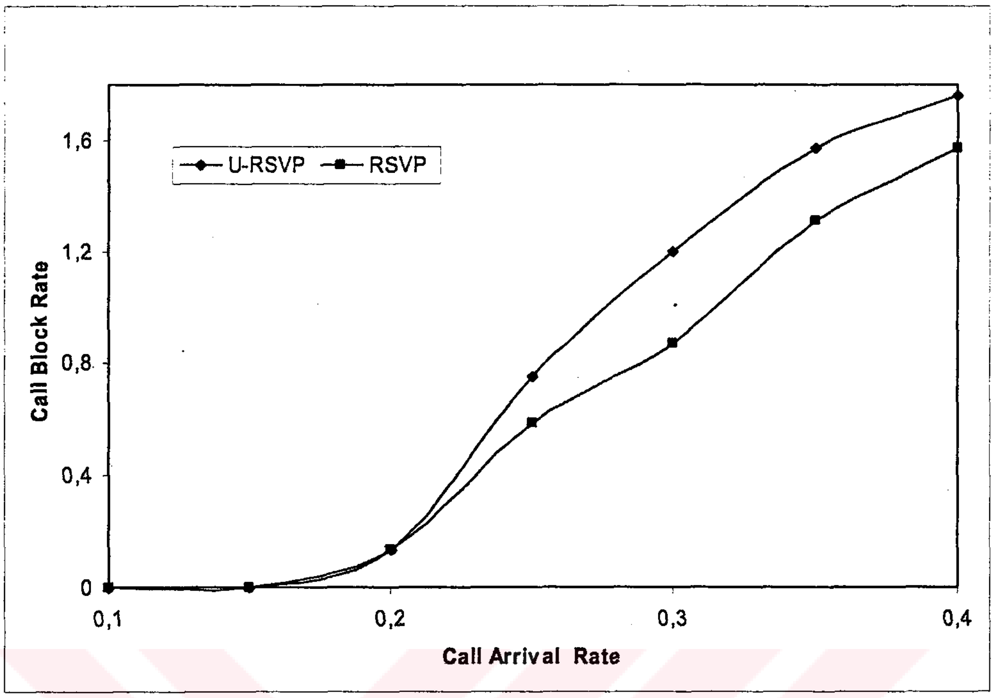


Fig. 16. The call blocking rate versus call arrival rate for UMP accuracy $\Psi = 1.0$.

Fig.16 shows call blocking rate versus call arrival rate for UMP accuracy $\Psi = 1.0$. When compared with RSVP there is a slight increase in U-RSVP since system resources are kept for open connections.

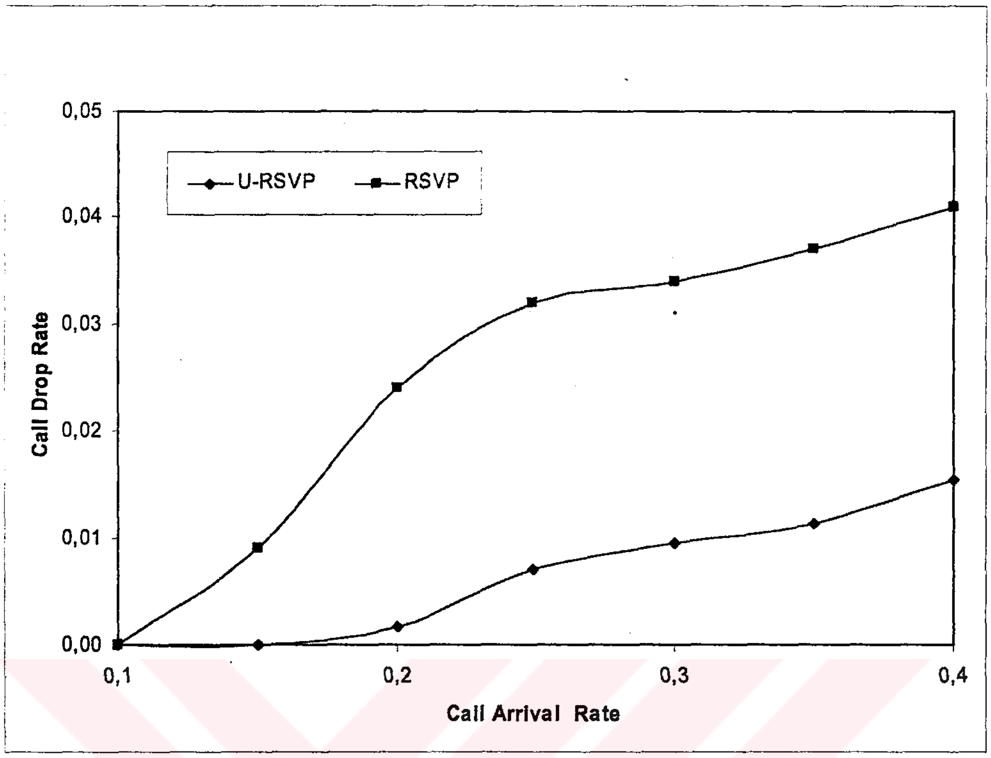


Fig. 17. The call dropping rate versus call arrival rate for UMP accuracy $\psi= 1.0$.

As seen in Fig.17 U-RSVP outperforms RSVP for call drop rate. The reason is the power of UMP. In UMP MTs' future location can be predicted. When the future location is known UMP helps us to reserve resources when an MT needs at that time.

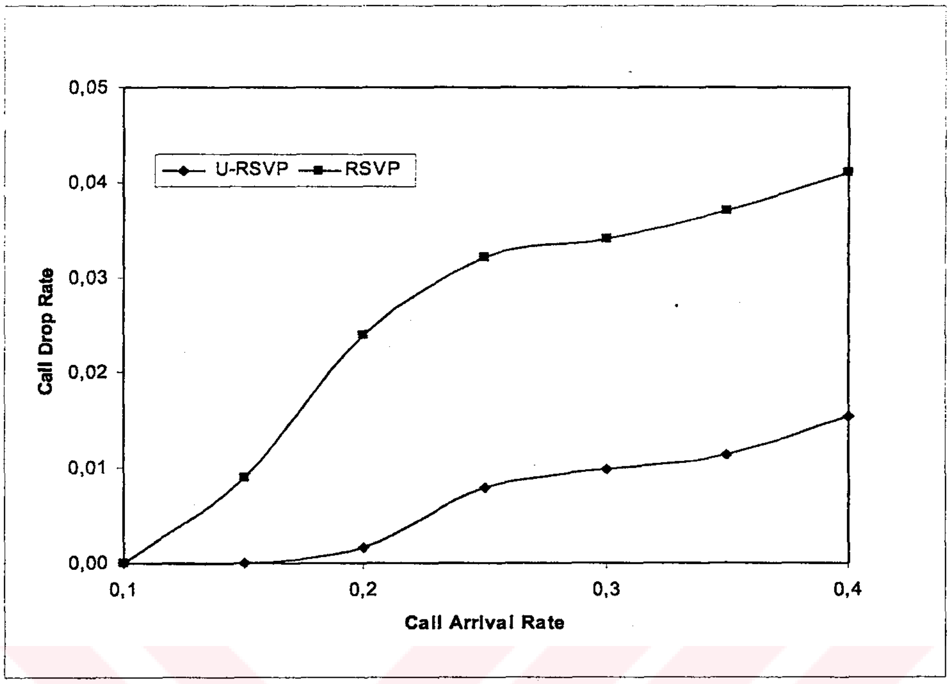


Fig. 18. The call dropping rate versus call arrival rate for UMP accuracy ψ= 0.9

Fig.18 shows that when the UMP accuracy ψ= 0.9, U-RSVP is again very powerful when it is compared with RSVP. When U-RSVP is used the call drop rate decreases %93 for call arrival rate is 0,2.

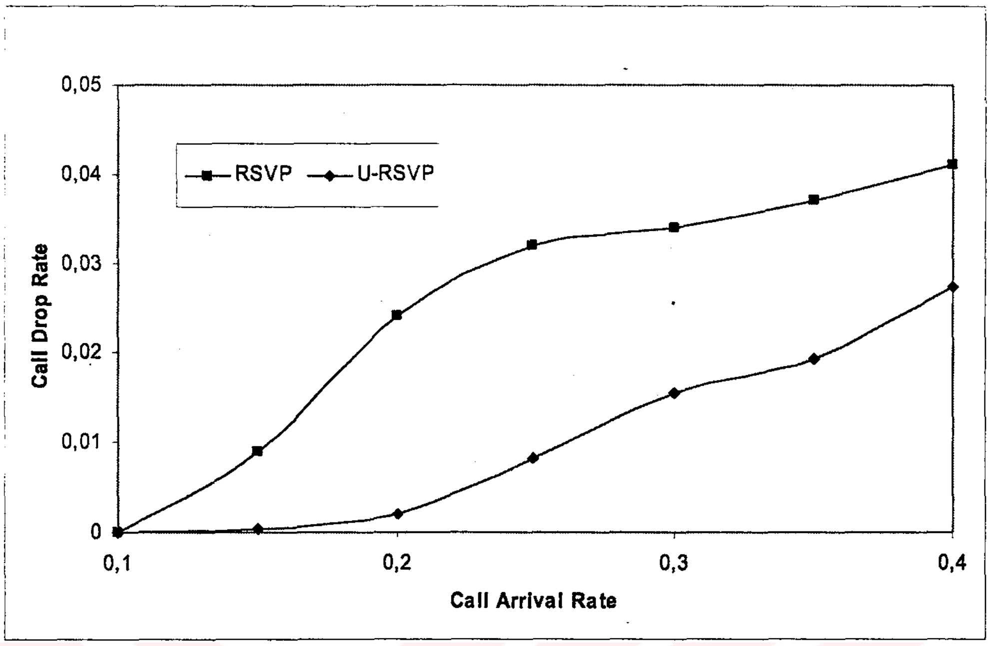


Fig. 19. The call dropping rate versus call arrival rate for UMP accuracy ψ= 0.8 .

There is an increase in call drop rate when the UMP accuracy ψ= 0.8 But the call drop rate still too low for U-RSVP when it is compared with RSVP.

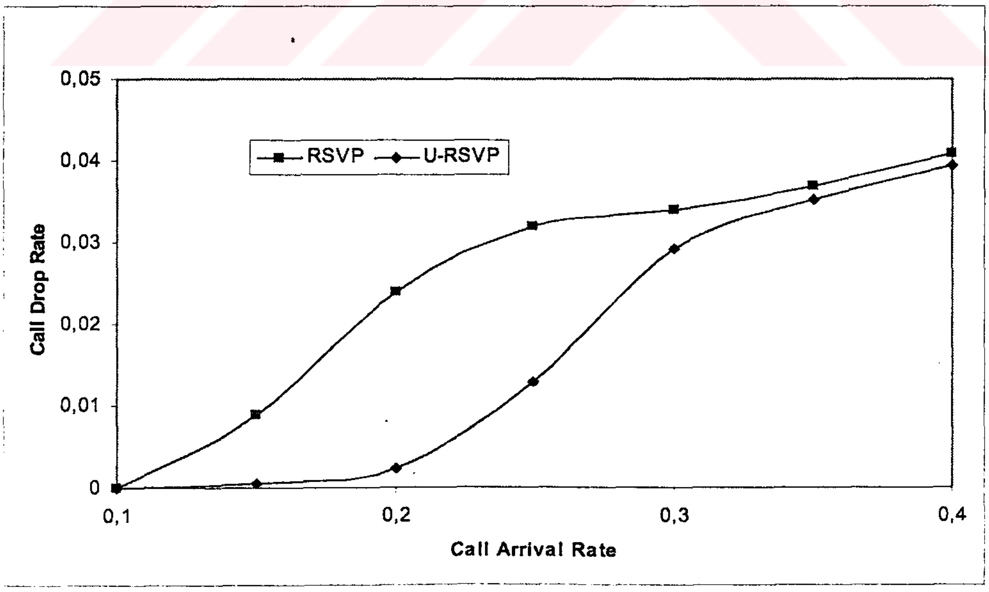


Fig. 20. The call dropping rate versus call arrival rate for UMP accuracy ψ= 0.7 .

Fig. 16-20 show the call blocking and dropping rates for the varying UMP accuracy ψ= 0.7-1.0 and call arrival rates. These figures show that our scheme reduces the call-dropping rate 79 % on the average in the expense of slightly increased call blocking rates.

Figures 17-20 prove that U-RSVP outperforms RSVP in call dropping rates. We repeat the experiments for varying UMP accuracies, and observed that U-RSVP performs better than RSVP even for UMP accuracy ψ=0.7. It is shown by a statistical study that the UMP accuracy is higher than 0.8 in the average [1].

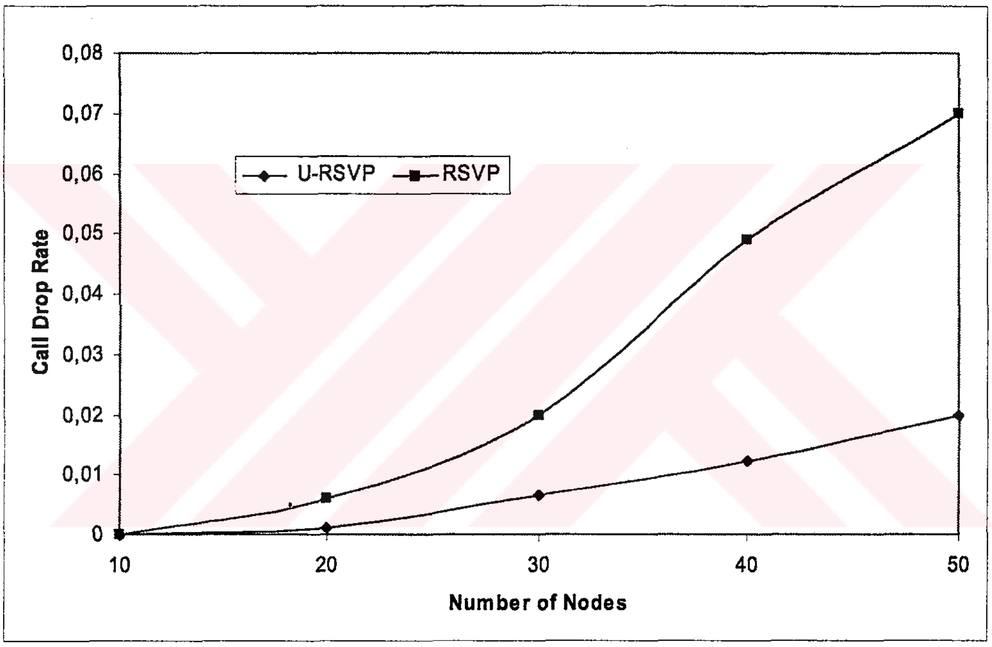


Fig. 21. The call dropping rate versus number of the nodes

While the number of nodes increases the call request also increase. Fig.21 shows the call dropping rate versus number of the nodes for traffic load is 0.2. It is seen that especially when the number of users is 30 call drop rate for RSVP increases exponentially while call drop rate for U-RSVP is under 0,01.

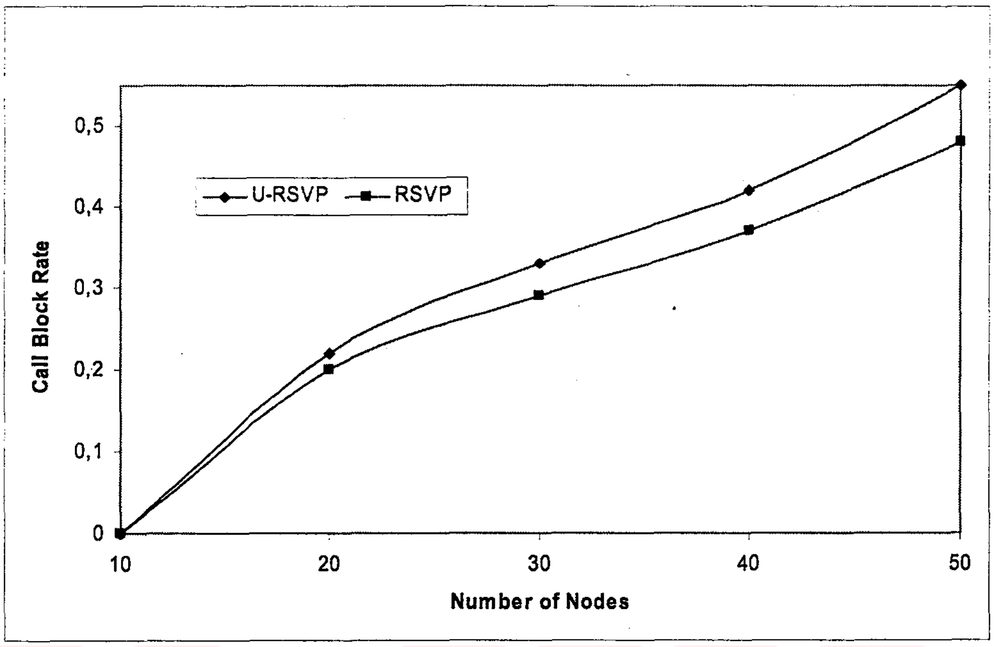


Fig. 22. The call blocking rate versus number of the nodes

Fig. 21 and 22 show the relationship between the call blocking rates and the node densities. Since higher node density indicates higher traffic load, call blocking rates increases in the higher node densities.

Fig. 23 shows the overhead of advance LSPs, i.e., the capacity used to establish the advanced LSPs and the resources reserved for them, for varying advance LSP duration Δ between 1 and 10 minutes. The number of nodes is 50, the call arrival rate is 0.2 call/h, and the average speed of the nodes is 10 km/h in these experiments. The overhead of U-RSVP is less then 1% for the advance LSP duration Δ is 1 minute.

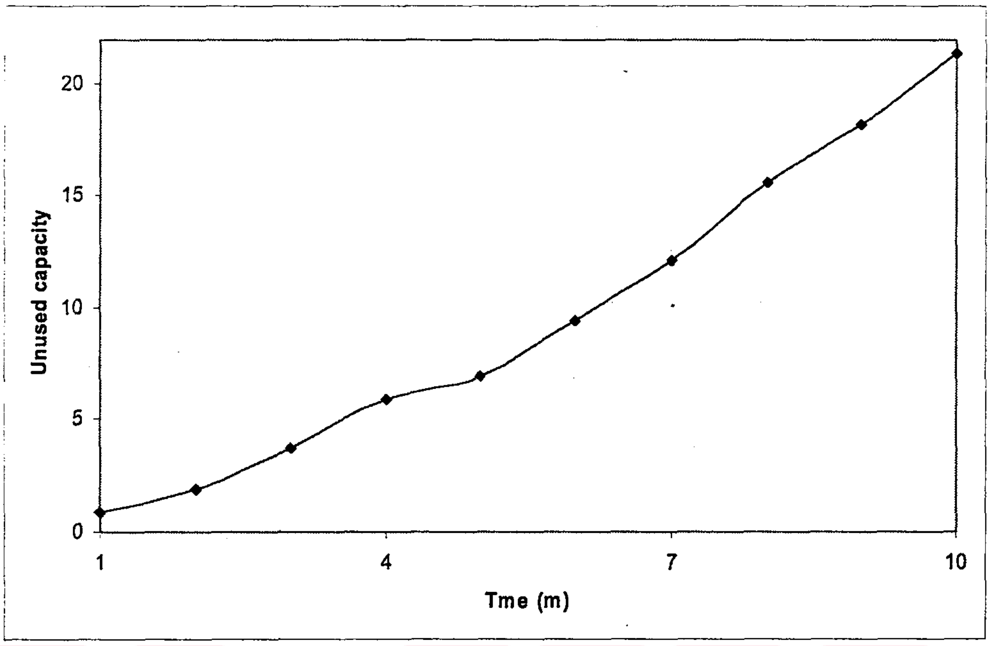


Fig. 23. The overhead of advanced LSPs for varying advance LSP duration Δ.

Fig.23 shows the overhead of advanced LSPs for varying advance LSP duration Δ. When Δ is 1 minute the overhead is under 1%. U-RSVP decreases call drop rate %78 while it causes just %1 overhead.

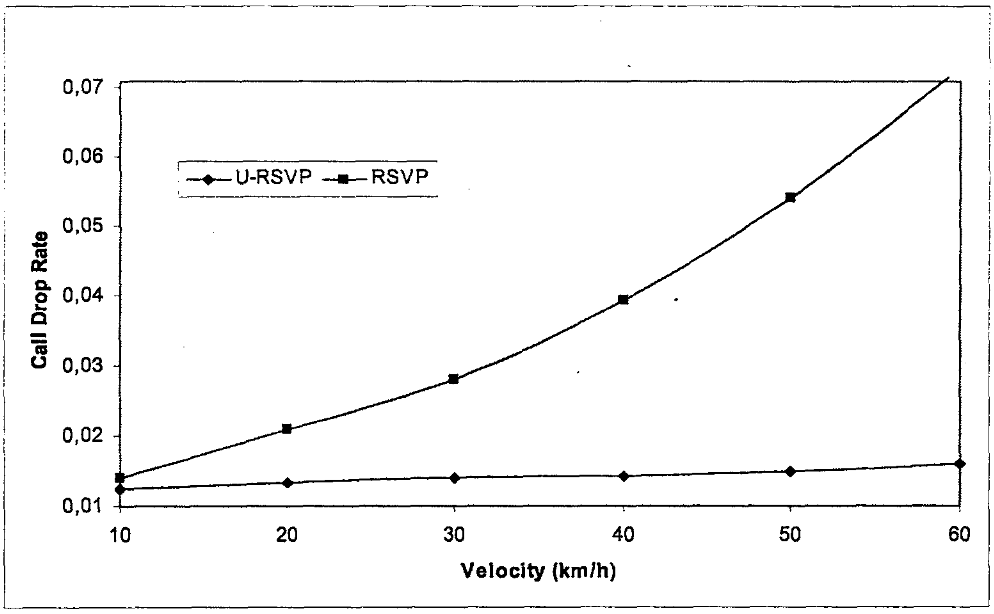


Fig. 24. The call dropping rate versus velocity

Higher speed of MTs indicates more handoff. Fig.24 shows the call dropping rate versus velocity. While the call dropping rate increases very fast for RSVP, it tracks nearly a horizontal pattern with U-RSVP. This shows that higher handoff number causes huge call drop rates with higher speeds for RSVP. But U-RSVP supports seamless connection to MTs.

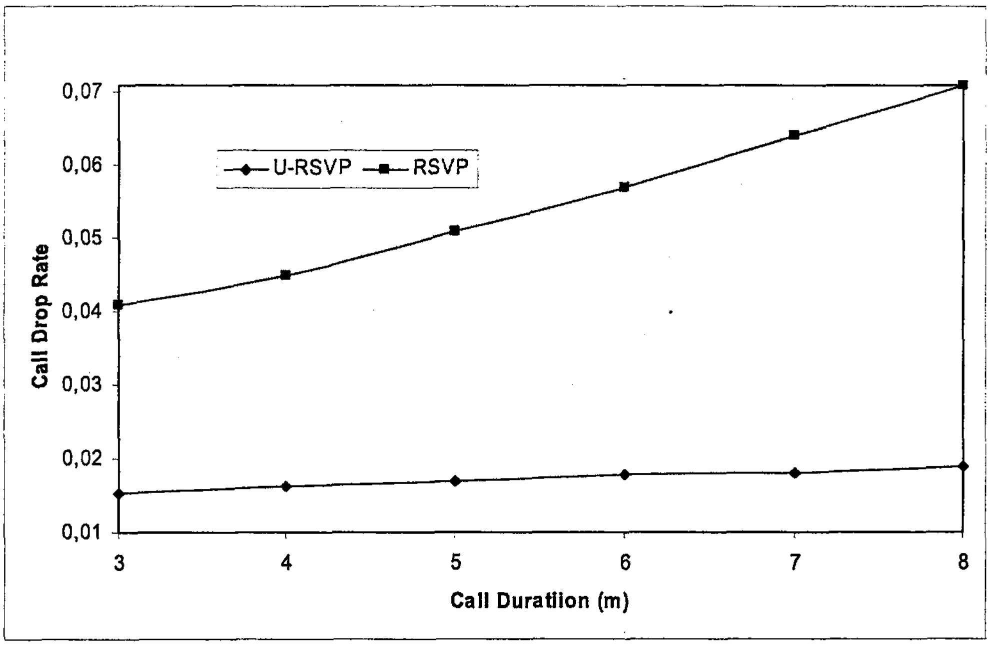


Fig. 25 Call drop rate versus call duration time

Fig. 25 Call drop rate versus call duration time. Since longer call duration means more handoff RSVP causes higher Call drop rates while it is under 0,02 for U-RSVP.

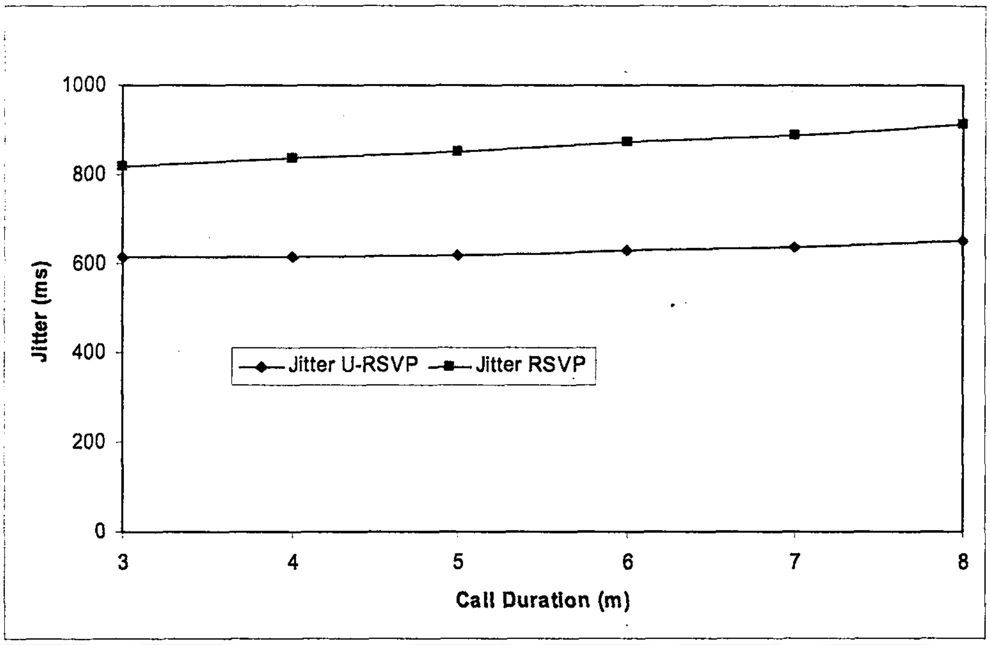


Fig. 26. U-RSVP's jitter sensitivity to call duration during handoff.

The average RTT values for GSM networks are around 600 ms as it is discussed in [122]. Since transmission speed decreases after handoff, an additional jitter occurs. In . Fig. 25. U-RSVP's jitter sensitivity to varrying call duration during handoff is shown. In U-RSVP when handoff occurs the new LSP connection is ready while in RSVP the connection must be done after handoff.

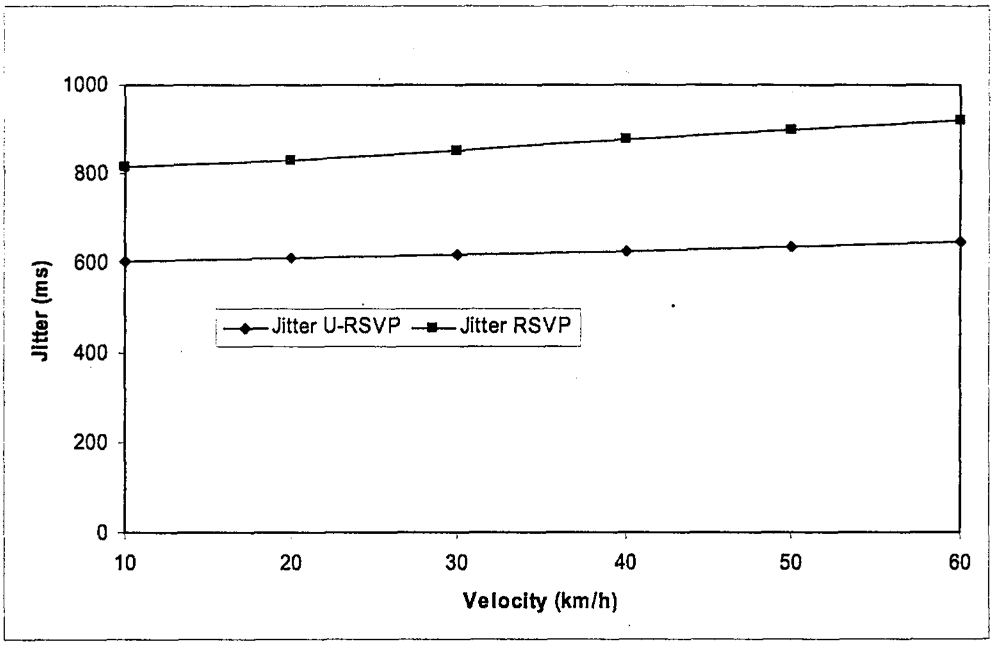


Fig. 27. U-RSVP's jitter sensitivity to velocity during handoff

Higher speed of MTs indicates more handoff. Fig.27 shows U-RSVP's jitter sensitivity to velocity during handoff. Since velocity affects handoff, the delay with RSVP is higher for higher velocities.

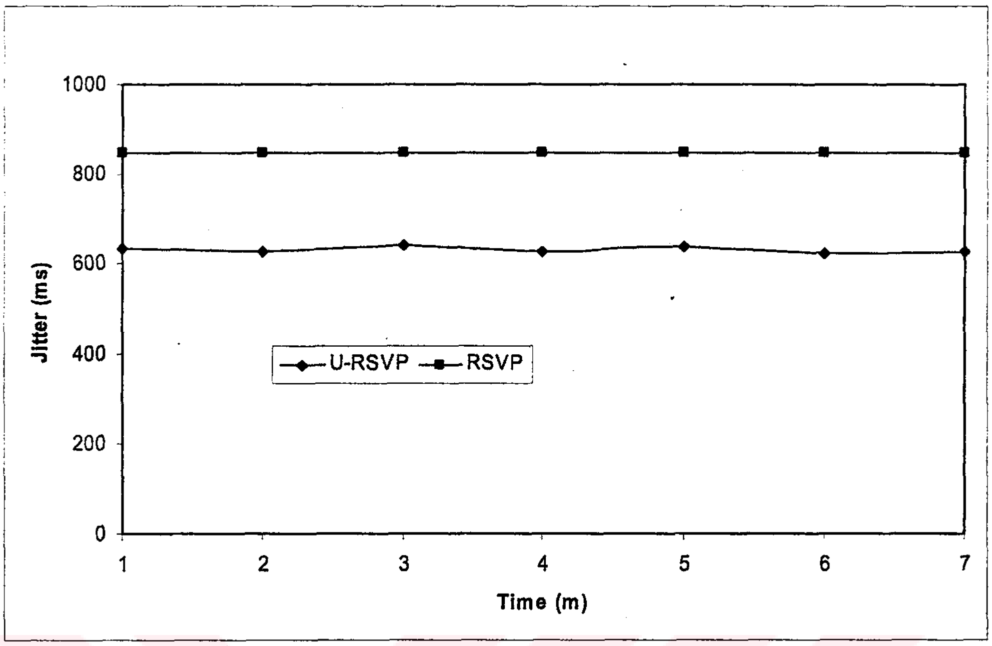


Fig. 28. U-RSVP's jitter sensitivity to Δ during handoff

Fig. 28. shows U-RSVP's jitter sensitivity to Δ during handoff where Δ indicates advance LSP time. Since higher Δ values cause more system resource use, higher Δ values causes higher drop rates. On the other hand as seen in the figure this does not affect delay for open. connections. It is because the advance LSP is ready for the coming connection.

VI.CONCLUSION

The demand or different Internet service increases and number of mobile users will grow in the future decade and the problems that have to be solved increase. Efficient use of resources and QoS needs are paramount topics for wireless mobile networks. Terminal mobility imposes additional difficulties in the realization of reservation. Anybody who watches a remote video or a videoconference with someone else does not want to connection dropped. But predicting the future location of MTs is not an easy task. The best solution is reserve resource all the neighbor cells that the MT is located. But resources is wireless media is not infinitive. UMP knows the next cell that an MT will handoff and the cell entry time can be predicted during the lifetime of a data flow, so resources can be reserved just in the right cell at the right time.

MPLS has been developed for QoS, gigabit forwarding, network scaling, and traffic engineering to support both the integrated and differentiated services [3]. Mobility issues in original MPLS are not supported. In this thesis a new UMP based resource reservation protocol (RSVP) scheme that reduces the cost of bandwidth reservation in MPLS is introduced to support end-t-end QoS and using wireless frequency spectrum most efficiently. Since not only the next cell but also the cell entry time can be predicted by using UMP based location management scheme [1], we can create a new LSP in the next cell before a handoff occurs. This makes MPLS more powerful and more reliable for QoS requirements of MTs. In this thesis it is shown that a new LSP for which the source node is the new cell can be effectively created for an active connection before an MT handoffs, and this can be achieved transparent to the MT when the UMP based location management scheme is used. Our scheme complies with the current RSVP protocol, and does not impose any need to modify it. We just add a new extension in application layer for advance reservation. The network predicts the next cell that an MT will handoff and the expected cell entry time by using the UMP registered by the MT. Then the BS in the next cell notified. When our scheme is used, the call dropping rate decreases 78% in the average, in the expense of a slightly increased call-blocking rate. Our system also utilizes the system resources more effectively.

A. FUTURE WORK

MPLS supports QoS between ingress and egress LSPs. It means MPLS does not support end-to-end QoS. In our thesis we propose an end-to-end QoS by no changes in original MPLS. Our work makes MPLS more powerful since it supports both mobility and end-to-end QoS. There must be new studies to make MPLS available for end-to-end QoS both for fix users and mobile users.

REFERENCES

[1] E. Cayirci and Ian F.AKYILDIZ, "User Mobility Pattern Scheme For Location Update and Paging in Wireless Systems", IEEE Transaction on Mobile Computing, Vol. 1, No. 3, July-September 2002.

[2] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D.Zappala "RSVP :A new Resource Reservation Protocol", IEEE Network, Sptember 1993.

[3] G.Armitage, "MPLS: The Magic Behind the Myths." IEEE Communications Magazine, January 2000 pp:124-131.

[4] J.M.Chung, K.Srinivasan, M.A.S..Benito, "Wireless Multiprotocol Label Switching (WMPLS)", IETF Internet Draft, draft-chung-mpls-wmpls-00.txt, March 2002.

[5] S.Gai, D.G.Dutt, N.Elfassy, Y.Bernet, "RSVP Proxy", IETF Internet Draft, draft-ietf-rsvp-proxy-03.txt.

[6] Y. Bernet, P. Ford, R. Yavatkar, F. Baker, L. Zhang, M. Speer, R. Braden, B. Davie, J. Wroclawski, E. Felstaine, "A Framework for Integrated Services Operation over Diffserv Networks", RFC 2998, November 2000.

[7] K. Nichols, S. Blake, F. Baker, D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.

[8] Tolga Onel, Cem Ersoy, and Erdal Cayirci, 'A Fuzzy Inference System for the Handoff Decision Algorithms in the Virtual Cell Layout Based Tactical Communications Systems', IEEE MILCOM 2002, 7-10 October 2002, Anaheim-California-USA.

[9] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[10] L. Andersson, P. Doolan, N. Feldman, A. Fredette, B. Thomas, ”LDP Specification”, RFC 3036, January 2001.

[11] E. Rosen, A. Viswanathan, R. Callon, “Multiprotocol Label Switching Architecture”, RFC 3031, January 2001.

[12] R. Braden, L. Zhang, S. Berson, S. Herzog and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, Sept. 1997

[13] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, Dec. 1998.

[14] X. Xiao and L. M.Ni ”Internet QoS: A Big Picture”, IEEE Network,March/April 1999

[15] J. Postel, "Service Mappings," RFC 795, Sept. 1981.

[16] Y.Bernet, “The Complementary Roles of RSVP and Differentiated Services in the Full-Service QoS Networks”, IEEE Communications Magazine, February 2000 pp:154-162

[17] J. Manner, A.L. Toledo, A.Mihailovic, H.L V Mu-noz E.Hepworth, Y. Khouaja, “Evaluation of mobility and quality of service interaction”, Elsevier Compter Networks 38 (2002) 137–163

[18] J. Ash, Y. Lee, P. Ashwood-Smith, B. Jamoussi, D. Fedyk, D. Skalecki, L. Li, “LSP Modification Using CR-LDP”, RFC 3214, January 2002

[19] T. Yang, and D. Makrakis, “Hierarchical Mobile MPLS: Supporting Delay Sensitive Applications over Wireless Internet”, International Conference on Info-Tech & Info-Net (ICII 2001), Beijing, China, October 2001.

[20] A.Banerjee, J.Drake, P.Lang, B.Turner, K.Kompella, Y.Rekhter, “Generalized Multiprotocol Label Switching: An Overview of Routing and Management Enhancements.”, IEEE Communications Magazine, January 2001 pp:144-150

[21] Sergio González-Valenzuela, Victor C. M. Leung, "QoS routing for MPLS networks employing mobile agents", IEEE Network, no. 3, May 2002 pp. 16-21

[22] G.Armitage, “MPLS: The Magic Behind the Myths.” IEEE Communications Magazine, January 2000 pp:124-131

[23] B. Jabbari, R.Papneja, E.Dinan, “Label Switched Packet Transfer for Wireless Cellular Network”, IEEE Wireless Communications and Networking Conference, 2000

[24] J.Lawrence, “Designing Multiprotocol Label Switching”, IEEE Communications Magazine, July 2001 pp:134-142

[25] C.Scoglio, “Design and Management of Diffserv MPLS-Based Networks”

[26] T. Anjali, C. Scoglio, J.C. de Oliveira, I.F. Akyildiz, G. Uhl, “Optimal Policy For Label Switched Path Setup In Mpls Networks”, Elsevier Computer Networks 39 (2002) 165–183

[27] D.O.Awduche, “MPLS and Traffic Engineering in IP Networks”, IEEE Communications Magazine, December 1999 pp:42-47

[28] X.Xiao,A.Hannan, B.Bailey, L.M.Ni.,”Traffic Engineering with MPLS in the Internet” IEEE Network, March/April 2000

[29] C. Scoglio, T. Anjali, J. C. de Oliveira, I. F. Akyildiz, and G. Uhl, "A New Threshold-Based Policy for Label Switched Path Setup in MPLS Networks," in proceedings of the 17th International Teletraffic Congress (ITC 2001),

Salvador da Bahia, Brazil, Elsevier Science Publisher, pp.1-11, December 2001

[30] C. Bruni, C. Scoglio, S. Vergari, "Optimal Capacity Provisioning for Label Switched Paths in MPLS Networks," in proceedings of IFIP-TC6 Networking 2002, Pisa, Italy, May 2002

[31] T. Anjali, C. Scoglio, I. F. Akyildiz, J. A. Smith, and A. Sciuto,"MABE: A New Method for Available Bandwidth Estimation in an MPLS Network," to appear in proceedings of NETWORKS 2002 (Joint Conference: ICWLHN 2002 and ICN 2002), Atlanta, Georgia, August 26-29, 2002

[32] J. C. de Oliveira, C. Scoglio, I. F. Akyildiz, G. Uhl, and J. Smith, "A New Topology-Aware LSP Preemption Policy for DiffServ-MPLS Networks," to appear in proceedings of NETWORKS 2002 (Joint Conference: ICWLHN 2002 and ICN 2002), Atlanta, Georgia, August 26-29, 2002

[33] J. C. de Oliveira, C. Scoglio, T. Anjali, L. Chen, I. F. Akyildiz, J. A. Smith, G. Uhl, and A. Sciuto. "Design and Management Tools for an MPLS Domain QoS Manager," proceedings of SPIE ITCOM 2002, July 29 - August 2, 2002, Boston, Massachusetts

[34] Y.T.Chen, K.K.Lee, “A Flexible Service Model for Advance Reservation”, Elsevier Computer Networks 37 (2001) 251-262

[35] S.Choi, K.G.Shin, “A Comparative Study of Bandwidth reservation and Admission Control Scheme in QoS-Sensitive Networks”, ACM Wireless Networks 6 (2000) 289-305

[36] G.S.kuo, P.C.Ko, M.L.Kuo, “A Probabilistic Resource Estimation And Semi-Reservation For Flow Oriented Multimedia Wireless Networks”, IEEE Communications Magazine, February 2001 pp:136-141

[37] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D.Zappala "RSVP :A new Resource Reservation Protocol", IEEE Network, Sptember 1993

[38] F. Yu, V. Leung, "Mobility based predictive call admission control and bandwidth reservation in wireless cellular networks", Elsevier Computer Networks 38 (2002) pp. 577-589

[39] F. Yu, V. C :M. Leung, "A Framework of Combining Mobility Management and Connection Admission Control in Wireless Cellular Networks"

[40] N.Antunes, A.Pacheco and R.Rocha," An Integrated Traffic Model For Multimedia Wireless Networks", Elsevier Computer Networks, 38 (2002).

[41] S.Singh "Quality of Service Guarantees in Mobile Computing" Computer Communications

[42] R. Ghai , S.Singh "An Architecture and Communication Protocol for Picocellular Networks", IEEE Personal Communications Magazine Vol. 1(3), 1994, pp. 36-46.

[43] K.Lida, T. Takine, H. Sunahara, Y.Oie, K.Kawahara, "Performance Evaluation of the Architecture for End-to-End Quality-of-Service Provisioning", IEEE Communications Magazine, April 2000 pp 76-81

[44] S. Casner, V. Jacobson "Compressing IP/UDP/RTP Headers for Low-Speed Serial Links",IETF RFC 2508, February 1999

[45] S. Shenker, C. Partridge and R. Guerin, "Specification of Guaranteed Quality of Service", RFC 2212, Sept. 1997

[46] R. Braden, L. Zhang, S. Berson, S. Herzog and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, Sept. 1997

[47] D.O.Awduche, “MPLS and Traffic Engineering in IP Networks”, IEEE Communications Magazine, December 1999 pp:42-47

[48] C.Lee, Y.Chen, J.Chen, ”Providing Service Differentiation in Wireless Networks”, IEICE Trans. Communications Vol.E84-B, No.4 April 2001

[49] T. Yang, Y. Dong, Y. Zhang, and D. Makrakis, “Practical Approaches for Supporting Micro Mobility with MPLS”, International Conference on Telecommunications 2002 (ICT 2002), Beijing, China, June 23, 2002.

[50] R. Braden, D. Clark, S. Shenker « «Integrated Services in the Internet Architecture: an Overview RFC 1633 June 1994

[51] E. Crawley, R. Nair, B. Rajagopalan, H. Sandick, “A Framework for QoS-based Routing in the Internet”, RFC 2386, August 1998

[52] T. Li, Y. Rekhter, Provider Architecture for Differentiated Services and Traffic Engineering, RFC 2430, October 1998

[53] S.Shenker, C. Partridge and R.Guerin, “Specification of Guaranteed Quality of Service“, RFC 2212, Sept.1997

[54] N.Antunes, A.Pacheco, R.Rocha, “An Integrated Traffic Model for Multimeida Wireless Networks”

[55] P.M.L.Chan, R.E.Sheriff, Y.F. Hu, P.Conforto, C.Tocci, “”Design and Evaluation of Signaling Protocols for Mobility Management in an Integrated IP Environment”, Elsevier Computer Networks, 38 (2002), 517-530

[56] T.Tjelta, A.Nordbotten, M.Annoni, E.Scarrone, S.Bizzarri, L.Tokarchuk, J.Bigham, C.Adams, K.Craig, M.Dinis, “Future Broadband Radio Access systems for Integrated Services with Flexible resource Management”, IEEE Communications Magazine, August 2001

[57] P.V.Mieghem, H.D.Neve, F.Kuipers, "Hop-by-Hop Quality of Service Routing", Elsevier Computer Networks, 37 (2001), 407-423

[58] S.K.Das, R.Jayaram, N.K.Kakani, S.K.Sen, "A Call Admission and Control Scheme for Quality of Service (QoS) Provisioning in Next Generation Wireless Networks, J.C. Baltzer Wireless Networks 6 (2000) 17-30"

[59] I.D.D Curcio, Ville Lappalainen, M E.Mostafa, Elsevier Computer Networks 37 (2001), 425-445

[60] G. xylomenos, G.Polyzos, "Link Layer Support for Quality of Service on Wireless Internet Links", IEEE Personal Communications, October 1999

[61] N.R.Figueira, J.Pasquale, Providing Quality of Service for Wireless Links: Wireless/Wired Networks, IEEE Personal Communications, October 1999

[62] S.Chen, K.Nahrstedt, "An Overview of Quality of Service Routing for Next Generation High-Speed Networks: Problems and Solutions", IEEE Network, November/December 1998

[63] L.Zhang, L.Zheng, "Ipv6 Traffic with Multi-Class QoS in VPN", Elsevier Computer Networks 37 (2001) 263-275

[64] A.Orda, "Routing with End-to-End QoS Guarantees in Broadband Networks, "IEEE/ACM Transactions on Networking, Vol 7, No.3, June 1999, pp 365-374

[65] R.Guerin, A.Orda, "QoS Routing in Networks with Inaccurate Information: Theory and Algorithms", "IEEE/ACM Transactions on Networking, Vol 7, No.3, June 1999, pp 350-364

[67] G. Xylomenos, G.Polyzos, "Quality of Service Support Over Multi-Service Internet Links", Elsevier Computer Networks 37 (2001), pp. 601-615

[68] G.Bianchi, N.B.Melazzi, P.M.L.Chan, M.Holzbock, Y.F.Hu, A.Jhan, R.E.Sheriff, "Design and Validation of QoS Aware Mobile Internet Access Procedures for Heterogeneous Networks", Kluwer Academic Publishers Mobile Networks and Applications 8, 11-25, 2003

[69] C.Casetti, M.Gerla, S.Mascolo, M.Y.Sanadidi, R.Wang, "TCP Westwood:End-to-End Congestion Control for Wired/Wireless Networks", Kluwer Academic Publishers Mobile Networks and Applications 8, 467-479, 2002

[70] C.B.Ahmed, N.Boudriga, M.S.Obaidat, "Adaptive QoS Schemes in DWDM Networks", IEEE International Conference on Communications 2002. ICC 2002 Volume:5, 2002, pp 2881-2885

[71] I. Mahadevan, K.M.sivalingam, "Architecture and Experimental Results for Quality of Service in Mobile Networks using RSVP and CBQ", J.C.Baltzar Wireless Networks 6 (2000) pp 221-234

[72] S.W. Suthon, G.Meng, H.K.Pung, "An Adaptive End-to-End QoS Management with Dynamic Protocol Configurations", ICON 2002, 10th IEEE International Conference on Networks, 2002, pp 106-111

[73] E.W.Knightly, N.B.Shroff, "Admission Control for Statistical QoS: Theory and Practice", IEEE Network, March/April 1999

[74] A.Aljadhai, T.F.Znati, "Predictive Mobility Support for QoS Provisioning Mobile Wireless Environments", IEEE Journal on Selected Areas in Communications, Vol.19. No.10, October 2001,

[75] S.Dixit, Y.Guo, Z.Antoniou, "Resource Management and Quality of Service in Third-Generation Wireless Networks", IEEE Communications Magazine, February 2001 ,

[76] Y.Cheng, W.Zhuang, "DiffServ Resource Allocation For Fast Handoff in Wireless Mobile Internet", IEEE Communications Magazine, May 2002

[77] B.Moon, H.Aghvami, "RSVP Extensions for Real Time Services in Wireless Mobile Networks", IEEE Communications Magazine, December 2001

[78] A.Talukdar, A.Acharya, "MRSVP: A Resource Reservation Protocol for an Integrated Service Network", Kluwer Academic Publisher Wireless Networks 7, 5-19, 2001

[79] G.C.Lee, T.P.Wang, C.C.Tseng, "Resource Reservation with Pointer Forwarding Schemes for the Mobile RSVP"

[80] M.Mihakkak, N.Schult, D.Thomson "Dynamic Bandwidth Management and Adaptive Applications for a Variable Bandwidth Wireless Environment", IEEE Journal on Selected Areas on Communications, Vol.19, No.10 October 2001 pp 1984-1997

[81] D.A.Levine, I.F.Akyildiz, M.Naghshineh, "A resource Estimation and Call Admission Algorithm for Wireless Multimedia Networks Using Shadow Cluster Concept", IEEE/ACM Transactions on Networking 5 (1) 1997, pp1-12

[82] S.Ganguly B.Nath, N.Goyal, "Optmization Based Bandwidth Allocation in Mobile Cellular Networks", INFOCOM 2002

[83] B.Li, L.Yin, K.Y.M.Wong, S.Wu, "An Efficient and Adaptive Bandwidth Allocation Scheme for Mobile Wireless Networks Using an Online Local Estimation", Kluwer Wireless Networks, 7, 107-116, 2001

[84] P.Ramanathan, K.M.Sivalingam, P.Agrawal, S.Kishore, "Dynamic Resource Allocation Schemes During Handoff for Mobile Multimedia Wireless

NEtworks", IEEE Journal on Selected Areas in Communicationa, Vol.17 No.7, July 1999, pp 1270-1283

[85] Ö.Gürbüz, H.Owen, "Dynamic Resource Scheduling Schemes for W-CDMA Systems", IEEE Communications Magazine, October 2000

[86] M.Moustafa, I.Habib, M.Naghshineh, "Wireless Resource Management Using Genetic Algorithm for Mobiles Equilibrium", Elsevier Computer Networks, 37 (2001) pp 631-643

[87] J.H.Lee, T.H.Jung, S.U.Yoon, S.K.Youm, C.H.Kang, "An Adaptive Resource Allocation Mechanism Including Fast and Reliable Handoff in IP-Based 3G Wireless Networks", IEEE Personal Communications, December 2000

[88] X.Liu, E.K.P.Chong, N.B.Shroff, "Opportunistic Transmission Scheduling with Resource-Sharing Constraints in Wireless Networks ", IEEE Journal on Selected Areas in Communicationa, Vol.19 No.10, October 2001, pp 2053-2064

[89] M.Xiao, N.B.Shroff, E.K.P.Chong, "Resource Reservation in power Contrdolled cellular wireless Systems", Wireless Communications and Mobile Computing 2001;1;183-199

[90] J.Li, N.B.Shroff, E.K.P.Chong, "Channel Carrying: A Novel Handoff Scheme for Mobile Cellular Networks", IEEE/ACM Transactions on Networking Vol.7, No.1, February 1999, pp.38-50

[91] K.D.Lee, S.Kim, "Optimization for Adaptive Bandwidth Reservation in Wireless Multimedia Networks", Elsevier Computer Networks 38 (2002) pp.631-643,

[92] A. Bar-Noy, I. Kessler, and M. Sidi, "Mobile Users: To Update or not to Update," ACM/Baltzer J. Wireless Networks, vol. 1, no. 2,pp. 175-186, July 1995.

[93] C. Rose, "Minimizing the Average Cost of Paging and Registration: A Timer-Based Method," ACM/Baltzer J. Wireless Networks, vol. 2, no. 2, pp. 109-116, June 1996.

[94] I.F. Akyildiz, J.S.M. Ho, and Y-B. Lin, "Movement-Based Location Update and Selective Paging for PCS Network," IEEE/ACM Trans. Networking, vol. 4, pp. 629-638, Aug. 1996.

[95] J. Li, H. Kameda, and K. Li, "Optimal Dynamic Mobility Management for PCS Networks," IEEE/ACM Trans. Networking, vol. 8, pp. 319-327, June 2000.

[96] J.S.M. Ho and I.F. Akyildiz, "Mobile User Location Update and Paging under Delay Constraints," ACM/Baltzer J. Wireless Networks, vol. 1, no. 4, pp. 413-425, Dec. 1995.

[97] U. Madhow, M. Honig, and K. Steiglitz, "Optimization of Wireless Resources for Personal Communications Mobility Tracking," IEEE/ACM Trans. Networking, vol. 3, pp. 698-707, Dec. 1995.

[98] D. Gu and S.S. Rappaport, "A Dynamic Location Tracking Strategy for Mobile Communication Systems," Proc. IEEE VTC '98, pp. 259-263, May 1998.

[99] S.K. Sen, A. Bhattacharya, and S.K. Das, "A Selective Location Update Strategy for PCS Users," ACM/Baltzer J. Wireless Networks, vol. 5, no. 5, pp. 313-326, Sept. 1999.

[100] S.K.sen, A.Bhattacharya, S.K.Das, “A Selective Location Update Strategy for PCS Users“, ACM/Baltzar J.Wireless Networks, Vol.5. No.5, September 1999, pp 313-326

[101] Y.Bejerano, I.Cidan, “An Efficient Mobility Management strategy for Personal Communication Systems”, ACM/IEEE Mobicom 98, Dallas, TX, October 1998, pp 215-222

[102] G.Krishnamurthi, M.Azizoğlu, A.Somani, “Optimal Location Management Algorithms for Mobile Networks”, ACM/IEEE Mobicom 98, Dallas, TX, October 1998, pp 223-232

[103] J.Jannink, “Efficient and Flexible Location Management Techniques for Wireless Communication Systems”, ACM/Baltzar J.Wireless Networks, Vol.3, October 1997, pp 361-374

[104] G.P.Loppini, C.L.İ, “A Profile Based Location Strategy and Its Preformance”, IEEE JSAC, Vol. 15, No.8, October 1997, pp1415-24

[105] S.Tabbane, “An Alternative Strategy for Location Tracking”, IEEE JSAC, Vol 13, no.5, 1995, pp 880-92

[106] Z.Naor, H.Levy, “Cell Identification Codes for Tracking Mobile Nodes”, IEEE INFOCOM 99, New York, NY, March 1999

[107] Z.Noar, H.Levy, “Minimizing the Wireless Cost of Tracking Mobile Users: An Adaptive Threshold Scheme”, IEEE INFOCOM 98, San Francisco CA,March/April 1998, pp 720-27

[108] V.W.S.Wong, V.C.M.Leung, “Location Management for Next Generation Personal Communications Networks”, IEEE Network, September/October 2000, pp.18-23

[109] V.W.S.Wong, "An Adaptive Distance-Based Location Update algorithm for Next Generation PCS Networks", IEEE JSAC vol 19 no.10 October 2001, pp.1942-52

[110] V.C-Giner, "Global Versus Distance Based Local Mobility Tracking Strategies: A Unified Approach", IEEE JSAC vol.51,no3, May 2002, pp.472-485

[111] K.-T. Chen, S.L.Su, „Design and Analysis of Dynamic Mobility Tracking in Wireless Personal Communications Networks",IEEE JSAC vol.51,no 3, May 2002, pp 486-97

[112] X.Shen, J.W.Mark, J.Ye, "User Mobility Profle prediction: An Adaptive Fuzzy Interference Approach", J.C. Baltzar Wireless Networks 6 (2000), pp. 363-374

[113] K.D.Lee, S.Kim, "Modeling Variable User Mobility with Stochastic Correlation Concept", Elsevier Computer Networks 38 (2002), pp. 603-12

[114] Z.Wang, "Internet QoS", Morgan Kaufmann Publishers, Academic Press, USA, 2001,

[115] G.Armitage, "Quality of Service in IP Networks", Published by:MTP, Indianapolis, IN, USA, April 2000, First Edition

[116] B.P.V.Kumar, P.Venkataram, "Prediction-Based Location Management Using MultilayerNeuralNetworks", J.Indian Inst.Sci., 2002, 82, 7-21

[117] I.F.Akyildiz, J.McNair, J.Ho, H.Uzunalioğlu, W.Wang, "Mobility Management in Current and Future Communications Networks", IEEE Network, July/August 1998

[118] I.F.Akyildiz, W.Wang, "A Dynamic Location Management Scheme for Next-generation Multitier PCS Systems", IEEE Transactions on Wireless communications vol 1.no 1, January 2002

[119] B.Liang, Z.Haas, "Predictive Distance Based Mobility Management for PCS Network", Proc. IEEE INFOCOM 99 NY, March 1999

[120] S. Choi and K.G. Shin, Predictive and adaptive bandwidth reservation for handoffs in QoS-sensitive cellular networks, in: Proc. ACM SIGCOMM '98, Vancouver, British Columbia (September 1998) pp. 155–166

[121] M. Naghshineh and M. Schwartz, Distributed call admission control in mobile/wireless networks, IEEE Journal on Selected Areas in Communications 14(4) (May 1996) 711–717.

[122] G. Xylomenos and G.C. Polyzos, TCP Performance Issues over Wireless Links, IEEE Communications Magazine, pp. 52-58, April 2001.

INDEX