MTEQ-MPLS: A Micro-Mobility Architecture for dynamic QoS-aware Traffic Engineering in Next Generation Wireless Access Networks

O. Horvath, T. Drick, O. Drobnik
Institute of Telematics / Department of Computer Science
Johann Wolfgang Goethe-University
60054 Frankfurt am Main, Germany
[horvath,drick,drobnik]@tm.cs.uni-frankfurt.de

Abstract

The Multi Protocol Label Switching (MPLS) is a core technology providing means to ensure quality of service requirements and to control data flows in IP networks. An increasing number of telecommunication core networks are deploying MPLS. In the future MPLS will be integrated into the wireless access networks. But MPLS has been developed with respect to wired networks. In this paper we propose a micro-mobility architecture based on the integration of MPLS for wireless networks. The proposed scheme provides strict support for quality of services guarantees, traffic aggregation and offers traffic engineering capabilities. Furthermore, in this architecture an additional low latency handover scheme is provided to minimise the interruption time for handovers.

1. Introduction

In the last years, there was an increasing growth of mobile devices based on an enormously interest of Internet-based services. This requires new telecommunication networks that can efficiently deal with the available resources and offer an intelligent mobility management. On the other hand, data-based services are getting more and more important. For this reason, next generation telecommunication networks will be based on the Internet Protocol (IP) to harmonise the architecture for voice and data traffic. To meet the demands of different types of services, e.g. real time services, it is necessary to enhance the IP-based infrastructure with additional mechanisms.

Considering the general architecture of telecommunication networks, it consists of three domains: Core Network Domain (CND), Wireless Access Network Domain (WAND) and User Equipment Domain (UED). The CND is responsible for functions like billing, user paging and provides access to other telecommunication providers. The UED consists of mobile devices, e.g. cell phones or laptops with wireless network card, which are called Mobile Nodes (MN). The WAND is placed between the CND and the UED and is responsible for the mobility management and provides MN access to the CND.

A multiplicity of telecommunication CNDs turned to Multi Protocol Label Switching (MPLS) [11] based networks. MPLS is a technology combining the simplicity of IP routing with high speed switching. Furthermore, MPLS provides mechanisms for Quality of Service (QoS) and traffic engineering (TE) by using the extended Resource Reservation Protocol (RSVP-TE) [1] over MPLS. But to benefit of MPLS mechanisms for the complete network it is necessary to deploy MPLS to the WAND as well. Unfortunately, MPLS is not designed to support mobility.

The mobility of a MN can be distinguished into two different levels: macro-mobility (inter-domain mobility) and micro-mobility (intra-domain mobility). Mobile IP (MIP) [10] is the current standard for macro-mobility in IP-based networks. But Mobile IP has several drawbacks, e.g. high signalling load for frequent registration updates and long handover times, depending on the distance between the Home Agent (HA) and Foreign Agent (FA).

To overcome these limitations, the presented architecture combines mechanisms for micro-mobility management with support for different QoS requirements and TE management by using Diffserv-aware MPLS Traffic Engineering (DS-TE) [3]. DS-TE is an extension to RSVP-TE that offers enhanced QoS functionalities. Additionally, the presented architecture reduces the bottleneck between the WAND and CND by offering an enhanced micro-mobility management. Furthermore, a low latency handover mechanism is applied, which minimises the time for intra-domain handover.

The paper is organised as follows. Section 2 introduces the related work on micro-mobility solutions. Section 3 presents the new proposed MPLS micro-mobility solution for WANDs. Finally, section 4 summarises the main results and outline future work.

2. Micro-Mobility in IP-based Networks

In a telecommunication network the MNs move frequently between subnets of one domain. To reduce load and delay of signalling to the home network during movements within one domain, several micro-mobility solutions have been proposed. The goals for micro-mobility solutions are to reduce the number of registration messages and update messages as well as packet loss and latency of packets.

Well-known solutions for IP-based micro-mobility are Cellular IP, Hawaii, hierarchical Mobile IP (HMIP) and Mobile IP Regional Registration (MIP-RR) [4]. Although those proposals addresses micro-mobility management and fast handover mechanisms, they have deficiencies in scalability, flexibility and capabilities for QoS [8]. Therefore those solutions are not appropriate to telecommunication networks.

To overcome these limitations MPLS, will apply in IP-based WANDs to support QoS and TE in the whole network by using the MPLS protocol family only. Therefore it is necessary to develop micro-mobility solutions that can deal with QoS and TE requirements. Classical IP-based protocols, e.g. Integrated Services (IntServ) or Differentiated Services (DiffServ), can deal with the QoS requirements. However IntServ is a host-based reservation scheme which results in poor scalability. On the other hand DiffServ drops the data packets if there isn't any more bandwidth available because it offers no TE support. All these proposals don't provide mobility management.

Several approaches integrate micro-mobility protocols with MPLS. One of the first micro-mobility schemes with QoS support has been proposed in [12]. This scheme integrates HMIP, consisting of a two level hierarchy, with MPLS and uses DiffServ for QoS support. The work of [6] introduces Micro Mobile MPLS which is a combination of MIP-RR and MPLS. Simulation results show, that Micro Mobile MPLS outperforms other (only) IP-based micro-mobility solutions. In [7] Micro Mobile MPLS is enhanced with a fast handover scheme by introduces active and passive Label Switched Paths (LSP). Due to the completion of handover, the passive LSP has only to change to active mode. These pre-established LSPs save setup time after a handover. But they don't offer sufficient QoS mechanisms.

Our approach presented in [4] offers an MPLS-based WAND with HMIP as mico-mobility solution and RSVP-TE for QoS support. But this architecture can only handle one QoS class. Therefore an enhanced architecture M-QoS MPLS was presented in [5]. M-QoS MPLS uses the Differentiated Services Code Point (DSCP) field of DiffServ to assign data packets to the correspondent QoS class. But RSVP-TE only offers a simple TE mechanism: if a LSP cannot handle more MNs, then RSVP-TE tries to establish a second LSP on the same path through the network. In case of exceeding the capacity of the physical line the TE mechanism tries to establish a new LSP on a separate physical path. A better solution would be to assign a fixed capacity to a QoS class and if the capacity is utilised then TE should build an alternative LSP path.

All micro-mobility solutions reduces the registration signalling delay, but fail to address the problem of movement detection. Therefore, cross-layer solutions (layer 2 and layer 3) are proposed to improve the handover mechanism. Cross-layer solutions use link layer information, such as signal strength, to trigger layer 3 mobility management protocols to perform a handover before the signal is lost. With preregistration and postregistration methods, according to [9], packet loss and the handover latency can be reduced [2]. The work of [5] integrates a low latency handover to an MPLS-based WAND based on preregistration.

3. Mobile Traffic Engineering with QoS-aware MPLS

To deal with the limitations of existing approaches, the new architecture Mobile Traffic Engineering with QoS-aware MPLS (MTEQ-MPLS) is proposed. It provides a QoS management that eliminate the drawbacks of RSVP-TE by integrating DS-TE. MTEQ-MPLS reduces the bottleneck between the borders of CND and WAND by introducing a distributed gateway architecture. Furthermore, the different traffic flows are aggregated to reduce administration complexity. MTEQ-MPLS only modifies routers at the borders of the WAND, that means, inside the WAND it is possible to use standard Label Switched Routers (LSR), which reduces the costs of a WAND. A important requirement is also to keep the MIP implementation transparent to the MNs and the HAs.

Our mobility management framework combines MPLS with Mobile IP as macro-mobility solution and applies principles similar to those in the MIP-RR scheme to the network as micro-mobility solution. Figure 1 illustrates the architecture of MTEQ-MPLS.

The architecture is based on a two level hierarchy by introducing two types of FAs: Gateway FA (GFA) and Mobile FA (MFA). The GFAs should reside on the gateway routers between CND and WAND. The MFAs are placed at the border of the WAND between WAND and UED, and they are connected to Access Points (AP) that offer link-layer functionality of the air interface, and the IP-layer mobility (layer 3 handover) when a MN moves between subnets served by different MFAs. The MFAs send agent advertisements to the MNs, handle registration requests and forward

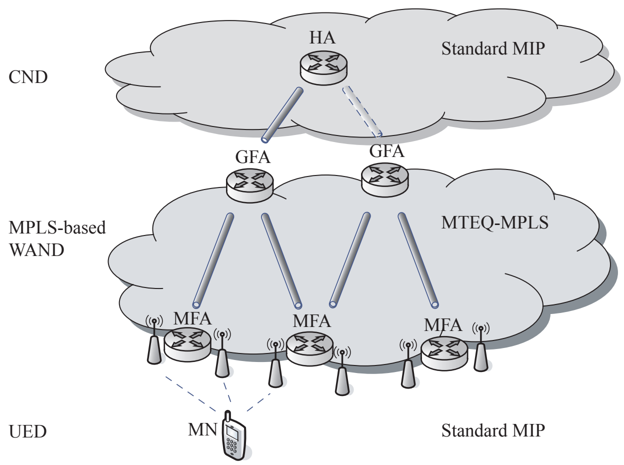

Figure 1. MTEQ-MPLS architecture

data packets from / to the MNs. Between the GFAs and the MFAs are no intermediate FAs. Therefore it is possible to use standard LSRs within the WAND. The LSRs are responsible to forward the traffic from the GFAs to the MFAs and vice versa.

Adding LSRs offers a way to archive scalable granularity in the network. Furthermore, by adding GFAs to the network, it is possible to make the bottleneck less critical to the CND. MTEQ-MPLS uses pre-established LSPs. Moreover, MTEQ-MPLS uses different LSPs to fulfil the different requirements for QoS classes. MTEQ-MPLS uses RSVP-TE with the DS-TE extension to ensure the requirements. RSVP-TE offers support for QoS. But DS-TE allows the possibility to distinguish between different types of traffic. Furthermore, DS-TE is able to make separate bandwidth reservations for different classes of traffic. This implies keeping track of how much bandwidth is available for each type of traffic at any given time on all routers throughout the network. DS-TE provides a very granular support of QoS. It is possible to define different classes that could be optimized for voice applications, video telephony or interactive web applications. The use of MPLS traffic engineering mechanisms distribute the traffic load to the entire network. To reduce the administration effort the data flows belonging to the same QoS class are aggregated.

The GFAs and MFAs are connected via DS-TE tunnels. Every GFA is connected to several MFAs and MFAs are connected to multiple GFAs (see figure 1). One main benefit of the MTEQ-MPLS architecture is the possibility, that more than one GFA can be added to the WAND. This enables MTEQ-MPLS to share traffic from / to the CND to several gateways. Furthermore, in case of a GFA failure, the MFAs will detect that the failuring GFA is not longer available and will route the traffic to another GFA. That improves the scalability and eliminates the single point of failure in comparison to HMIP.

3.1. Network initialisation

MTEQ-MPLS uses pre-established LSPs between GFAs and MFAs. That means, that a registration of a MN doesn't initialise a LSP setup. Each MFA contains two lists at startup: known GFAs and supported QoS classes. Both lists can be dynamically extended during operation. The LSP establishment is triggered by the MFAs. For each GFA a MFA performs the following algorithm:

First, the MFA creates a DS-TE tunnel with a configured bandwidth for each QoS class to the GFA. Then the MFA sends a *GFA registration request* messages to the GFA. The GFA registration request is a MIP registration request message with a modified *type field*. Furthermore, the GFA registration request messages contains an extension that specifics the bandwidth reservations of the different QoS classes for the downstream tunnel. When the GFA receives the message, it creates a DS-TE tunnel to the MFA, also with the configured bandwidth for each QoS class. If there is already an established tunnel, then only a reservation update will be done. Finally a *GFA registration reply* is sent back to the MFA. The reply is a MIP registration reply with a modified type field and the care-of address of the GFA as an extension. The care-of address is the address of the interface that is reachable from the HA.

The GFA registration request message is a modified MIP registration request message. The type field is set to a value that identifies the message as GFA registration request. Because the signalling messages need an appropriate target address, the HA address is set to the address of the GFA. To integrate the QoS requirements for the tunnel establishment, the corresponding information is added to the MIP *Type-Length-Value* extension. The MIP Type-Length-Value extension consists of a type, length and data field (see figure 2). The type field contains the fixed value for the DS-TE extension and the length field consists of the size of the data field. The data field comprises the information about all used QoS classes supported by the WAND. For each QoS class it is necessary to insert a DSCP value that represents a defined QoS class with the responsible bandwidth constraints.

Each data packet that requires a special QoS class needs to set a corresponding DSCP value. On the base of the DSCP value and the target IP address it is possible to map the data packet to the right DS-TE tunnel.

The GFA registration reply is also a modified MIP registration reply message. The type field is set to corresponding value, that identifies the message as GFA registration reply. The extension of the GFA registration reply message includes the care-of address of the GFA. Note, that all extensions are standard compliant and transparent to the HA

Type = (GFA RQ ID)	Length = $5 \cdot n$ Byte
DSCP QoS 1	Bandwidth QoS 1
DSCP QoS 2	Bandwidth QoS 2
...	
DSCP QoS n	Bandwidth QoS n

Figure 2. The structure of the MIP extension

and MNs.

After the establishment of all DS-TE tunnels, the MFAs starts sending MIP agent advertisement messages. These messages include a list with the care-of addresses of all connected GFAs.

3.2. Registration procedure

All MFAs send agent advertisement messages containing a list with the care-of addreses of reachable GFAs. When a MN moves for the first time into a MTEQ-MPLS WAND, it will send a registration request (RQ) to the nearest MFA. This message uses the address of the first GFA of the received list as care-of address. The MFA records the MN home address at the MIP visitor list and relays the RQ to the GFA. When the GFA receives the RQ it updates the MIP visitor list and sends the RQ to the HA. If the RQ is successful at the HA, the HA will send a registration reply (RP) back. When the GFA receives the RP, then the GFA adds an entry for the MN at the Forwarding Equivalence Class (FEC) tables and links the entry to the corresponding DS-TE tunnel. Following, the GFA relays the RP along the established DS-TE tunnel to the MFA. Finally, the MFA adds a FEC entry that links to the corresponding DS-TE tunnel and forwards the RP to the MN. Note, that MNs that are connected to the same MFA will use the same established DS-TE tunnel for traffic forwarding that belongs to the same QoS class (traffic aggregation).

3.3. Handover support

In general, there are two different micro-mobility handovers: intra-MFA and inter-MFA handover. An intra-MFA handover occurs when the MN moves between two AP managed by the same MFA and an inter-MFA handover occurs when a new AP and the old AP are connected to different MFAs.

INTRA-MFA HANDOVER Once the association to the current AP is lost, the MN scans for new APs. If it finds one, the MN will register at layer 2 with the AP and either wait for a agent advertisement message sent from the MFA or it will issue a mobile *solicitation message*. In any case, the MN registers at the MFA. The MFA needs only to update the Address Resolution Protocol tables. The intra-MFA handover is complete and no additional signalling to GFA / HA is necessary.

INTER-MFA HANDOVER In case the new AP is connected to a different MFA, the MN compares the care-of addresses and decides to perform a micro-mobility registration to the GFA or macro-mobility registration to the HA. The new MFA does not have any entry for this MN. Therefore, the MFA adds the MN to the MIP visitor list and forwards the RQ to the GFA. In case of a micro-mobility registration the GFA already contains an entry for the MN. For this reason the GFA updates the FEC tables, maps the FEC to the corresponding DS-TE tunnel and sends the RP back to the MFA. The MFA attaches an entry to the FEC class and reply the RP to the MN and the handover is complete. Note, that no message is sent to the HA of the MN. The HA is not notified unless the MN moves to another WAND or refreshes the binding if the global registration is about to expire.

3.4. Enhanced low latency handover

The aforementioned MTEQ-MPLS architecture works above layer 2 of the IP protocol stack. The disadvantage of this layer separation can be a lower performance caused by missing information about the state of the physical link, that is the reason for a delay in detection the disconnection to an AP. The MN has to wait for the last valid agent advertisement message to expire before it can start a registration procedure with the new MFA. A second reason for additional interruption time is the registration process itself. The RQ has to be forwarded through the WAND to the HA. The HA processes the RQ and the RP has to be sent back to the MN before the communication continues.

MTEQ-MPLS reduces the delay of the registration process. To enhance the MIP handover, MTEQ-MPLS integrates a low latency handover by using the preregistration mechanism. The preregistration mechanism performs the registration process while the MN is still connected to the old MFA, thus there is no latency caused by the registration process anymore. Therefore the MN will be enhanced by a additional component that monitors the radio link between MN and AP. If the link quality falls below a given threshold a signal is sent to the Mobile IP layer to trigger the preregistration process. The MN performs a scan for available APs and sends a list with available MFAs to the old MFA via proxy solicitation message [5]. The old MFA selects a new MFA and sends the reply back via a agent advertisement. The reason, that the MFA makes the decision and intiates the handover process, is that network providers would like to keep control over their network. After the MN received the reply, it starts a RQ MIP registration over the old MFA to the new MFA. If the new MFA received the message, it adds the appropriate information to its tables and forward

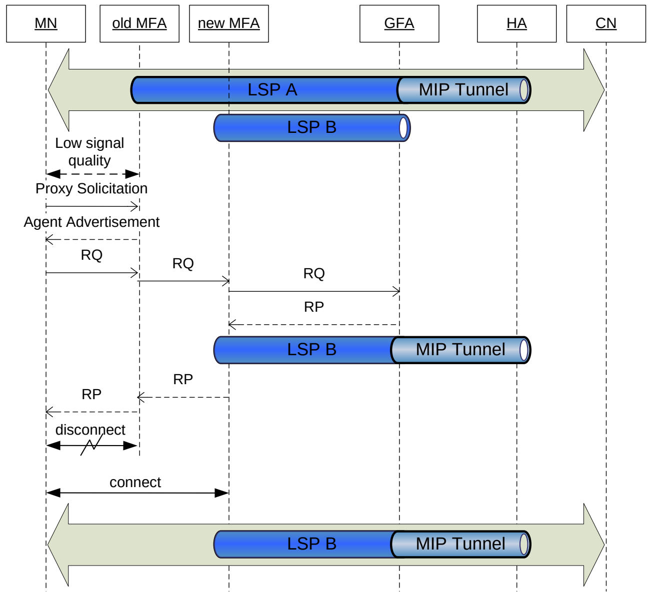

Figure 3. Preregistration procedure

the RQ to the GFA, and if necessary the GFA relays the RQ to the HA. After all involved agents received a positive RP, the MN disconnects immediately from the old MFA and then connects to the new MFA. No additional signalling is required. Technical details can be found in the work of [5]. Figure 3 gives an overview about the preregistration handover.

4. Conclusions

In this paper a new MPLS-based micro-mobility approach called Mobile Traffic Engineering with QoS-aware MPLS for wireless access networks has been proposed that fulfils strict quality of service requirements, traffic engineering and traffic aggregation. In this propose a two level MPLS-based hierarchical architecture is presented. The architecture reduces the signalling traffic to the home agent. Furthermore, the signalling overhead within the wireless access network is reduced too. Further on, the latency for registration messages and update messages is reduced as well. Different quality of service classes can be guaranteed by using RSVP-TE with the DS-TE extension. The differentiation for the different quality of service classes is done by using the Differentiated Services Code Point field of DiffServ. By introducing several GFAs the bottleneck between WAND and CND is made less critical. In general, a MN can receive data packets during the update process after a handover. Therefore, a low latency handover scheme for MN, based on a layer 2 trigger and preregistration mechanism, is proposed. The architecture has been implemented as a prototype. Further research will concentrate on detailed performance evaluation and the integration of a vertical handover mechanism between different radio access technologies and domains.

References

[1] D. O. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, and G. Swallow. RSVP-TE: Extensions to RSVP for LSP Tunnels. RFC 3209, Internet Engineering Task Force, Dec. 2001.

[2] A. Diab, A. Mitschele-Thiel, and R. Böringer. Extension of Mobile IP for Fast Authentication and Low Latency Handoff. In *Proc of 11th European Wireless 2005*, pages 805–811, Cyprus, April 2005. VDE Verlag.

[3] F. L. Faucheur. Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering. RFC 4124, Internet Engineering Task Force, 2005.

[4] O. Horvath, A. Janz, and O. Drobnik. Integration of Hierarchical Mobile IP into an MPLS-based UTRAN. In *Proc. of International Workshop on Mobile Computing and Media Communication in the Internet(MCMI'04)*, pages 296–300, Ulm, Germany, Sept. 2004.

[5] O. Horvath, G. March, and O. Drobnik. M-QoS MPLS: A New Scheme for Micro-Mobility Management with QoS support in Next Generation Wireless Access Networks. In *Proc. of International Workshop of Wireless Networks (IWWN'06)*, pages 920–925, Hongkong, China, June 2006.

[6] R. Langar, G. L. Grand, and S. Tohmé. Mobile MPLS Micro-Mobility protocol in Next Generation Access Networks. In *Proc. of 12th IEEE International Symposium on Modelling, Analysis and Simulations of Computer and Telecommunication Systeams (MASCOTS)*, Volendam, The Netherlands, October 2004.

[7] R. Langar, G. L. Grand, and S. Tohmé. Fast Handoff process in Micro Mobile MPLS protocol for Micro-Mobility. In *Proc. of the Second Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, St. Moritz, Switzerland, January 2005.

[8] R. Langar, G. L. Grand, and S. Tohmé. Micro Mobile MPLS: A New Scheme for Micro-Mobility Management in 3G All-IP Networks. In *Proc. of 10th IEEE Symposium on Computers and Communications (ISCC 2005)*, Cartagena, Spain, June 2005.

[9] K. E. Malki. Low Latency Handoffs in Mobile IPv4. Internet draft, rfc ed queue, IETF, October 2005.

[10] C. E. Perkins. IP mobility support for IPv4. RFC 3344, Internet Engineering Task Force, Aug. 2002.

[11] E. Rosen, A. Viswanathan, and R. Callon. Multiprotocol Label Switching Architecture. RFC 3031, Internet Engineering Task Force, January 2001.

[12] T. Yang, Y. Dong, Y. Zhang, and D. Makrakis. Mobility Support for Differentiated Services in Next Generation MPLS-based Wireless Networks. In *Proc. of International Conference of Communications, Circuits and Systems 2002 (ICCCAS 2002)*, Chengdu, China, June 2002.