A Multi-Layer Mobility Management Architecture Using Cross-Layer Signalling Interactions

Qi Wang and Mosa Ali Abu-Rgheff

Mobile Communications Research Group
School of Computing, Communications and Electronics
The University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, UK
<qwang, mosa>@plymouth.ac.uk

Abstract

Every layer of the mobile protocol stack contributes to next-generation advanced mobility management (MM). However, present MM architectures are largely single-layer specific and thus can hardly meet the simultaneous low- and high-level mobility support requirements. A co-operation of multiple layers would lead to a promising solution. Notably, there is an increasing interest to adopt cross-layer design for future wireless systems. Through this approach, the refined systems can obtain extended functionality and/or improved performance, which are hard to gain via a single-layer signalling. In this paper, we identify and abstract each layer's contributions to various mobility support tasks. By introducing cross-layer interactions to the TCP/IP protocol stack, we present an IP-based multi-layer MM architecture taking advantage of contributions from individual and combined layers.

1. Introduction

The complexity and requirements of mobility management (MM) are growing with the evolution of wireless systems. The well-established 2G (second generation) MM procedures were designed only for terminal mobility in a homogeneous system. In the next generation (NG, or 3G and beyond), a mobile user can access to heterogeneous networks for various services and multimedia sessions via a set of personal devices. Consequently, there will be three new types of high-level mobility: service mobility, session mobility and personal mobility [1].

From the perspective of the protocol stack, the network layer is the most appropriate level to converge heterogeneous networks in an all IP vision of the NG. Therefore, a large number of IP-based network-layer MM schemes have been proposed, in place of the traditional link-layer ones. So far only Mobile IP [2] has been standardised but is only suitable for terminal mobility and can hardly support the high-level mobility.

On the other hand, the application-layer Session Initiation Protocol (SIP) [3] has the potential capabilities to support the three high-level mobility types by augmented signalling [1]. However, currently SIP is not so much a mature solution than an initial framework. When extended for terminal mobility, it is no easy work for SIP to outperform Mobile IP in either supporting mid-session TCP mobility, or supporting real-time applications like voice over IP [4]. Therefore, the SIP-alone approach for a complete mobility support seems questionable.

Furthermore, some functions of the traditional link-layer mobility support could be utilised wherever available and appropriate. In addition, the link layer, together with the physical layer, could also help to tackle network-specific problems resulting from mobility, such as adaptation to the NG heterogeneous communication environments. These tasks are beyond the network- and the application-layer MM schemes including both SIP and Mobile IP approaches. However, an advanced MM architecture should consider all the above issues as well as support various mobility types.

In sum, the lessons we have learned are that a single-layer-specific MM architecture can hardly provide the advanced mobility support required by the NG. The intrinsic reason is that mobility brings about significant impacts on each layer, which in turn has its convenience to deal with different level mobility impacts. Thus, introducing a single collocated layer for various MM tasks, if possible, would be too complex and heavy. Therefore, it would be simpler and more flexible to develop a co-ordinated multi-layer architecture that can make full use of each layer's contributions while still keeping the basic structure of the TCP/IP protocol suite.

In the meantime, cross-layer design [5], especially via cross-layer signalling methods, has justified its introduction into wireless systems [6]. In fact, this methodology has been successfully applied in several areas, such as error correction [7], adaptation of wireless protocols [8], and optimisation of ad hoc networks [9]. Obviously, there exists a good case to combine the multi-layer MM architecture and the cross-layer design.

The remainder of this paper is organised as follows: Section 2 addresses related work. Section 3 presents our MM architecture making use of each layer's contributions via cross-layer interactions, based on an efficient cross-layer signalling method. Section 4 concludes and outlines the future work.

2. Related Work

Related work falls into two categories: one is multi-layer MM architectures and the other is cross-layer signalling methods.

2.1 Multi-Layer MM Architectures

Previous MM architectures are mainly located in one specific layer, and there is little work on a multi-layer protocol stack.

Since terminal mobility and the high-level mobility could be handled by the network and the application layer respectively, it is natural to combine Mobile IP and SIP for a complete MM solution. References [10-12] introduced SIP for selected high-level mobility and Mobile IP based network-layer MM protocols for terminal mobility. However, it seems that the two approaches were applied independently, rather than optimised jointly. Furthermore, these frameworks are concerned only with signalling in the network side, without taking any cross-layer design in the mobile host (MH) into account. Thus, no intrinsic integration across the protocol stack has been achieved. Anyway, straightforward as it seems, a simple "Mobile IP + SIP" approach can hardly be equivalent to our desired results. For fast handoffs, link-layer hints [13] and extended information [14] have been utilised respectively in specific contexts. However, these schemes were limited to terminal mobility and thus no high-level mobility or layers were involved.

2.2 Cross-Layer Signalling Methods

1) Method 1- Packet Headers:

The "Interlayer Signalling Pipe" [7] stores the cross-layer information in the Wireless Extension Headers of IPv6 packets. This method makes use of IP data packets as in-band message carriers with no need to use a dedicated message protocol. However, normally an IP data packet can only be processed layer by layer, and this can lead to inefficiency. In other words, this layer-by-layer signalling mechanism constructs a bottom-to-top "pipe", which seems excessive in most cases.

2) Method 2- ICMP Messages:

Method 2 [8] propagates information across layers by using ICMP (Internet Control Message Protocol) [15] messages. Since a message could be generated from any layer and then terminates at a higher layer, cross-layer signalling is carried out through these selected "holes", rather than the "pipe" in Method 1 as shown in Fig. 1. Thus, Method 2 appears more flexible and efficient. However, ICMP messages encapsulated by IP packets have to pass by the network layer even if the signalling is only desired between the link layer and the application layer. Furthermore, only upward ICMP messages were reported.

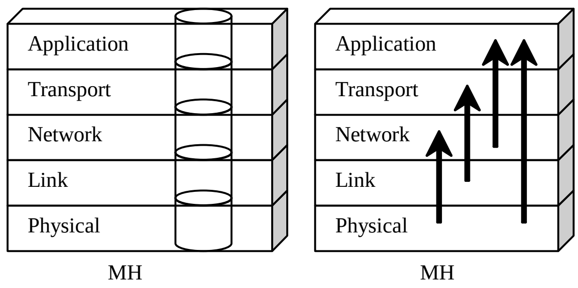

Figure 1: Comparison of Method 1 and Method 2

3) Method 3- Local Profiles:

In Method 3 [9] as shown in Fig. 2, cross-layer information is abstracted from each related layer respectively and stored in separate profiles within a MH. Other interested layer(s) can then select profile(s) to fetch desired information. Method 3 is flexible since profile formats can be tailored to specific layers, which in turn can access to the information directly. However, it is not suitable for time-stringent tasks.

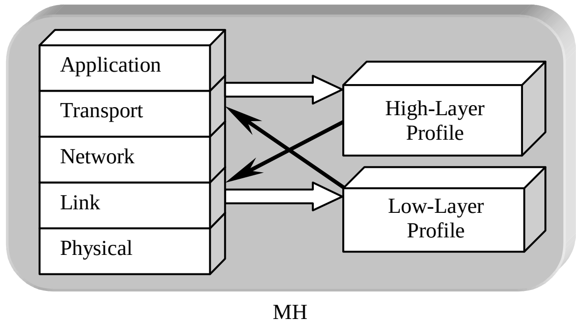

Figure 2: Concept model of Method 3

4) Method 4- Network Service:

In [16], channel and link information from the physical and link layers are gathered, abstracted and managed by third parties, the distributed Wireless Channel Information (WCI) servers. Interested applications access to WCI servers for their desired parameters as shown in Fig. 3. As a network service, it is complementary to the former schemes within a MH although some overheads in the air would be incurred and interfaces have to be defined among the MH, the WCI server and application servers.

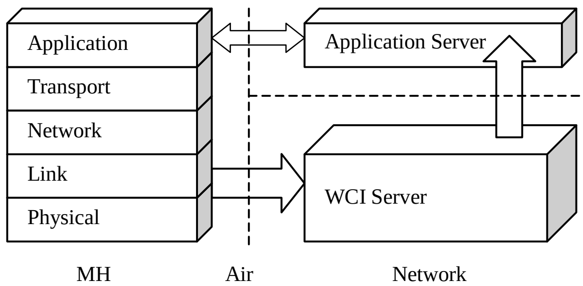

Figure 3: Concept model of Method 4

3. Class-Based Multi-Layer Mobility Management Architecture

We are working on a multi-layer MM architecture, which is aimed at an advanced mobility support for personal mobility, session mobility, service mobility as well as the traditional terminal mobility in a hybrid heterogeneous system [17]. The basic functions of the three high-level MM schemes are located in the application layer while the terminal MM in the network layer. Through proper cross-layer interactions, these schemes could work in a jointly optimised way and benefit from other layers' contributions. Consequently, improved performance and extended functionality could be expected, beyond the single-layer approach.

3.1 Contributions to MM from Each Layer

In the IP-based protocol stack, each layer has some positive or negative effects on mobility management. Here we abstract their possible contributions.

The physical and link layers can report current channel conditions and link properties respectively to upper layers, which can then adapt to mobility-incurred issues, in particular, the network heterogeneity. These reports can also be used to determine an imminent handoff. In an overlay-networking environment, the better/best system could then be chosen by performing an inter-system handoff. Meanwhile, by detecting and reporting the handoff arrival to the network layer in advance, the link layer expedites the IP-based handoff significantly.

Furthermore, in the link layer, different MAC (medium access control) techniques enable different Layer-2 handoff schemes, which are network-specific but could be utilised by the network layer in order to improve handoff performance.

The major job of the network layer is the support for terminal mobility, including IP-based handoff management and location management. With additional application-layer information such as traffic types and QoS (quality of service) requirements, multimedia-aware handoffs could be achieved in the IP level, with the help of IP-based QoS schemes like DiffServ (Differentiated Services) [18]. In addition, the network layer is in charge of generating and processing external cross-layer messages based on ICMP or IP header.

The transport layer, by providing reliable end-to-end data delivery, can decrease the packet loss ratio due to the error-prone wireless channels and the user's mobility. Moreover, for connection-oriented transport, live connection states should be maintained during a temporary disconnection. Collaboration with the layers below would facilitate these functions.

As to the application layer, it should take care of the high-level mobility tasks and their possible interactions with the network-layer terminal MM. In addition, getting aware of the timely information from the lowest two layers, many multimedia applications could become adaptive in resource requirements and transform themselves automatically. Therefore, the live session dropping possibility could be reduced during handoffs from the current system or bearer to another one with fewer resources. In addition to adaptation, location-based services also entail mobility awareness in the application layer.

Therefore, the advanced mobility support calls for a co-ordination of all the layers, which can be enabled by a proper cross-layer signalling method.

3.2 Introduction of CLASS (Cross-LAyer Signalling Shortcuts)

Aiming at the advanced MM, we can identify a couple of desirable features of the cross-layer signalling method. Above all, MM is a time-stringent task. Any delay during handoffs or location updates would lead to deteriorated performance of on-going sessions. Hence, we propose to enable the direct bi-directional interactions between non-neighbouring layers, especially between the application layer and the network layer to facilitate a real-time co-ordination. Otherwise, the transport layer has to be turned to as a middleman. In theory, the signalling can take place between any two arbitrary layers, although only selected paths are needed to fulfil a specific task. Furthermore, taking account of the multiplicity of MM requirements and the heterogeneity of the NG, suitable message formats are needed for rich and active internal and external signalling. Since no existing cross-layer signalling methods can satisfy these requirements, we introduce a new method, named CLASS (Cross-LAyer Signalling Shortcuts) [6], as shown in Fig. 4.

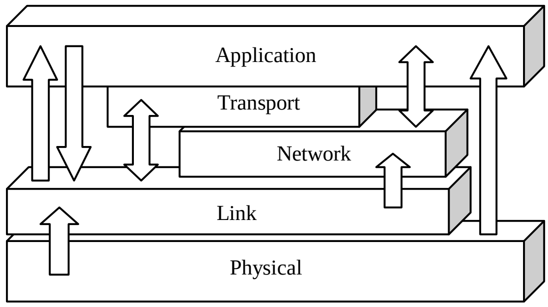

Figure 4: Concept model of CLASS

1) Enabling Direct Interactions between Non-Neighbouring Layers:

The basic idea is to break the layer ordering constrains while keeping the layering structure, i.e., to introduce local out-of-band signalling in the protocol stack of a MH. Previously, this approach only appeared as exceptions in some proprietary protocol stacks, and has not been designed for generic management functionality. For instance, the Layer-3 entity RRM (Radio Resource Management) directly accesses the physical layer in GSM [19]. This feature can be exploited and extended to the TCP/IP protocol stack. Notably, CLASS is also applicable to signalling between neighbouring layers.

2) Decomposing Internal/External Messages:

For external signalling, IP headers can be used for short notifications while ICMP for complex messages.

For internal signalling, it is not cost-effective to use standardised protocols, which are normally heavy-weighted. Reducing additional headers and minimising the fields can thus simplify the internal message format. Essentially, only three fields are required:

Destination Address, including destination layer and destination protocol(s) or application(s).

Event Type, indicating a parameter.

Event Contents, the value of the parameter.

Messages can also be propagated in an aggregate way by introducing an optional field, *Next Event*.

For message generation and reading, the mechanisms in Method 2 [8] can be based on. In general, a message with a layer-specific parameter is generated from the layer when a significant value change of the parameter has occurred. Function calls are used to set and fetch the parameter, and system calls read the message.

Notably, the actual interactions between layers are task-dependent and protocol-specific. In the design of CLASS, we focus on the mechanisms of the method itself, i.e., how to facilitate cross-layer interactions with efficiency and flexibility. The specific interactions are discussed in the next subsection.

3.3 MM Architecture with Cross-Layer Interactions

The proposed MM architecture is illustrated in Fig. 5 (not all the interactions are shown), enabled by a CLASS-based optimal combination of the cross-layer signalling methods. Considering all the layers' contributions, we have identified the following interactions between layers.

1) CLASS-Based Interactions:

For active messaging in either direction, light-weighted internal messages proposed in CLASS should be used.

Interactions between the network layer and the application layer:

The major application of CLASS is the direct co-ordination between these two layers without bothering the transport layer. There are intrinsic connections between terminal mobility and the high-level mobility, e.g., a live session is kept by both terminal and session MM schemes when a mobile user has a personal area network consisting of various personal communication devices. Thus, the two layers should perform in a co-operative way.

In a simultaneous terminal and session handoff scenario, an integrated approach would improve the management efficiency. In the case of SIP and Mobile IP running in the two layers at the same time, reduced overheads over the air could be expected by using the internal cross-layer signalling between the two protocols. For instance, to obtain a new IP address for a handoff, if without co-ordination, both SIP and Mobile IP would turn to a certain network service, e.g., DHCP (Dynamic Host Configuration Protocol) [20]. Since it is more convenient for Mobile IP to deal with this network-layer matter, we can configure Mobile IP to communicate with the DHCP server only. Anyway, SIP has difficulties to detect the change of the IP address even if it is allowed to contact the DHCP server itself. In [1], polling is used for such detection, however, polling is not optimal for this time-sensitive event and it would invoke many internal overheads. Thus, an active notification is desired only when this event actually happened. CLASS is the right solution to this problem since it can send this event from the network layer to the application layer in a timely and efficient way.

Interactions between the physical/link layers and the network layer:

The link-layer handoff notifications and extra system-specific information to the network layer can accelerate the Layer-3 terminal handoffs in the case of Mobile IP over 802.11b wireless LAN (local area network) [14]. Similar mechanisms could be exploited for other access networks where such Layer-2 information is attainable. In 3G systems, radio parameters measured by the physical/link layers are widely available (e.g., [21]). Useful parameters can be selected from these measurements, such as RSS (received signal strength) and SNR (signal to noise ratio).

Moreover, particular Layer-2 handoff mechanisms enabled by specific MAC techniques could also benefit Layer-3 handoffs. For instance, the CDMA (Code Division Multiple Access) cellular systems support Layer-2 soft handoffs that could lead to seamless handoffs in the network layer.

CLASS serves as a generic interface to enable the above interactions. In addition to improving the performance of intra-system handoffs, this link-aware interface could also help to smooth an inter-system handoff.

Interactions between the physical/link/network layers and the transport/application layers:

Our multi-layer MM architecture also facilitates the QoS adaptation of applications and transport protocols to contexts such as mobility, heterogeneous networks and the time-varying radio channel conditions. The utilisation of the mentioned 3G measurements can also be extended for MM with built-in QoS adaptation, with additional parameters abstracted from related layers. Meantime, different QoS requirements from different applications could also be mapped into controllable parameters of corresponding layers. All the parameters can be coded to CLASS messages for interactions across the protocol stack.

Particularly, notifications of the start and the end of a handoff are generated from the link or network layer and are sent to the transport layer. These messages are targeted to adjust the behaviours of transport protocols (esp., TCP) to mobility. TCP may suspend transmission and maintain the connection states during the handoff.

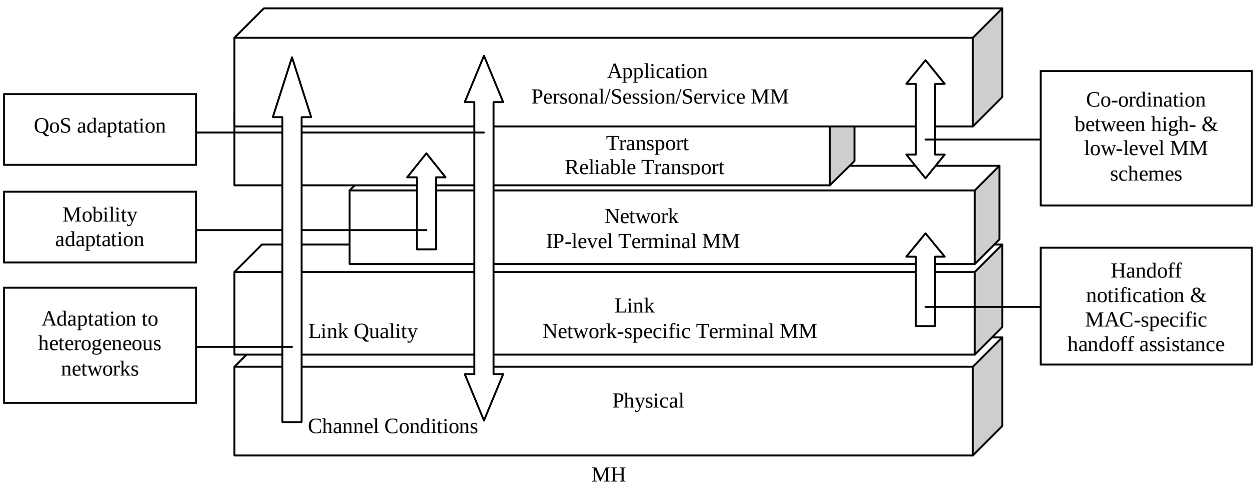

Figure 5: The CLASS-based multi-layer MM architecture

2) Other Interactions:

In some contexts, other methods can also be introduced.

For example, if multimedia traffic is processed on a packet-by-packet basis, extended headers can be applied to carry extra traffic-specific information including QoS requirements from the top of stack to the bottom (Method 1). This "pipe" is acceptable since a data packet has to travel through all the layers anyway.

A possible application of local profiles (Method 3) could be for information related to location updates.

For necessary messaging between a MH and its access networks, new network services can be considered (Method 4). For example, the existing link-layer hints [13] are neither reliable nor widely available, but can be enhanced and enriched through a link-layer agent in the network side. This agent is not necessarily as complex as the dedicated WCI server is. It can be collocated in border base stations (BSs) or the 3G Gateway Location Registers (GLRs) [22] located between two systems, e.g., a wireless LAN and a cellular network. It can monitor and provide an overall physical- and link-layer information. An approaching MH can then be informed of the new system's characteristics and capabilities in a heterogeneous environment, and thus it can prepare for an inter-system handoff. Within a system, the MH itself can observe and adapt to its contexts via the CLASS-based internal cross-layer interactions. The incurred overheads in the network can thus be minimised.

4. Conclusion

The advanced NG mobility management entails all the layers' participation in a highly co-operative way. In this paper, we identified the contributions from individual layers and put all the pieces together in the TCP/IP protocol stack. By applying cross-layer design, we constructed a multi-layer architecture for advanced mobility support. Possible support includes co-ordinated MM of different levels for various mobility types, fast/seamless handoffs, and QoS adaptation to the mobility-incurred context changes, such as heterogeneous networks. Active cross-layer interactions play a crucial role to enable and facilitate such extended functions and improved performance.

Future work is under way to simulate the CLASS-based multi-layer MM architecture and to evaluate its performance compared with single-layer schemes.

References

[1] H. Schulzrinne and E. Wedlund, "Application-Layer Mobility using SIP", Mobile Computing and Communications Review, Vol. 4, No. 3, July 2000, pp. 47-57.

[2] C Perkins, "IP Mobility Support", RFC 2002, 1996.

[3] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, "SIP: Session Initiation Protocol", RFC 2543, 1999.

[4] T. T. Kwon, M. Gerla, S. Das, and S. Das, "Mobility Management for VoIP Service: Mobile IP vs. SIP", IEEE Wireless Communications, Vol.9 No.5, Oct. 2002, pp. 2-11.

[5] Z. H. Haas, "Design Methodologies for Adaptive and Multimedia Networks", Guest Editorial, IEEE Communications Magazine, Vol. 39, No. 11, Nov. 2001, pp. 106-107.

[6] Q. Wang and M. A. Abu-Rgheff, "Cross-Layer Signalling for Next-Generation Wireless Systems", in Proceedings of IEEE WCNC'03, in press.

[7] G. Wu, Y. Bai, J. Lai and A. Ogielski, "Interactions between TCP and RLP in Wireless Internet", in Proceedings of IEEE GLOBECOM'99, Rio de Janeiro, Brazil, Dec. 1999.

[8] P. Sudame and B. R. Badrinath, "On Providing Support for Protocol Adaptation in Mobile Wireless Networks", Mobile Networks and Applications, Vol. 6, No. 1, Jan. 2001, pp. 43-55.

[9] K. Chen, S. H. Shan and K. Nahrstedt, "Cross-Layer Design for Data Accessibility in Mobile Ad Hoc Networks", Wireless Personal Communications, Vol. 21, No. 1, Apr. 2002, pp. 49-76.

[10] A. Dutta, R. Jain, D. Wong, J. Burns, K. Young and H. Schulzrinne, "Multilayered Mobility Management for Survivable Network", in Proceedings of IEEE MILCOM'01, Virginia, USA, Oct. 2001.

[11] T. Kanter, "An Open Service Architecture for Adaptive Personal Mobile Communication", IEEE Communications Magazine, Vol. 8, No. 6, Dec. 2001, pp. 8-17.

[12] Z. R. Turanyi, C. Szabo, E. Kail and A. G. Valko, "Global Internet Roaming with ROAMIP", Mobile Computing and Communications Review, Vol. 4, No. 3, July 2000, pp. 58-68.

[13] S. Seshan, "Low-Latency Handoff for Cellular Data Networks", Ph.D. Dissertation, The University of California at Berkeley, USA, 1995.

[14] N. A. Fikouras, A. J. Könsgen, and C. Görg, "Accelerating Mobile IP Handoffs through Link-Layer Information", in Proceedings of the International Multiconference on Measurement, Modelling, and Evaluation of Computer-Communication Systems (MMB'01), Aachen, Germany, Sept. 2001.

[15] J. Postel, "Internet Control Message Protocol", RFC 792, 1981.

[16] B-J "J" Kim, "A Network Service Providing Wireless Channel Information for Adaptive Mobile Applications: Part I: Proposal", in Proceedings of ICC'01, Helsinki, Finland, June 2001.

[17] Q. Wang and M. A. Abu-Rgheff, "Towards a Complete Solution to Mobility Management for Next-Generation Wireless System", in Proceedings of London Communications Symposium 2002 (LCS'02), London, UK, Sept. 2002.

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

[19] B. H. Walke, "Mobile Radio Networks: Networking, Protocols and Traffic Performance", 2nd Ed., John Wiley & Sons, Chichester, UK, 2002.

[20] R. Droms, "Dynamic Host Configuration Protocol", RFC 2131, 1997.

[21] 3GPP TS 25.215 V5.1.0, "Physical Layer - Measurements (FDD)", Sept. 2002.

[22] 3GPP TS 23.119 V5.0.0, "Gateway Location Register (GLR) - Stage 2", June 2002.