Towards a Complete Solution to Mobility Management for Next-Generation Wireless System

Qi Wang and Mosa Ali Abu-Rgheff
The University of Plymouth/Department of Communication and Electronic Eng.
<qwang, mosa>@plymouth.ac.uk

Abstract: While the 3G wireless system is still under deployment world-wide, pilot research has been targeted to the next generation (NG), whose full scope is not yet well defined. One feature of NG should be its heterogeneous communication environment, where different generation cellular systems coexist and various wireless access networks and corresponding services are used. Advanced mobility management (MM) is thus needed to fulfil seamless global roaming in terms of not only the traditional terminal mobility, but the high-level personal, session and service mobility. In this paper we discuss mobility management in previous, current and future mobile generations. We then present our vision of the NG framework and propose an approach leading towards a complete mobility management solution from the protocol stack and network architecture perspectives. We conclude with a series of associated open research problems in this hot research area.

1. Introduction

Wireless systems have been developed in an evolutionary way generation by generation over the last twenty years or so. 1G systems are of diminishing importance. The dominant generations today are 2G and 2.5G with 3G coming into use and are represented in Europe by GSM, GPRS and UMTS, respectively. In a wider context, besides these terrestrial wide area cellular systems there are a number of other wireless systems, such as the global area satellite systems, the wireless local area networks (WLANs), and the personal area networks (PANs) using, e.g., Bluetooth technology. These systems are optimised for different ranges and this gives the potential for them to co-operate in a complementary way. NG wireless system will provide multimedia mobile communications anytime and anywhere. It would be cost-effective to achieve this aim by converging various wireless networks. However, mobility management across such a heterogeneous system is not ready.

Mobility management in 2G includes two tasks: one is location management, and the other is handoff (or handover) management. Accordingly the traditional MM definition [1] can be referred to as **terminal mobility**. In NG, there are three other types of high-level mobility: personal mobility, service mobility and session mobility. For a user roaming across heterogeneous networks, a complete MM scenario can include but maybe not be limited to the following aspects:

- The user can have only one device (**terminal mobility**) or a PAN consisting of several personal devices (**session mobility**) to provide a live session. The former is done in a transparent way, while the latter can also be done manually.
- The user is globally reachable (**pre-session personal mobility**) and can originate or receive a session by access to different terminals (**personal mobility**).
- Subscribed services are personalised by user profiles, and the personalised services are consistently provided with quality of service (QoS) requirements met regardless of the user's locations (**service mobility**).

The remainder of this paper is organised as follows: In Section 2, we discuss the evolution of mobility management, focused on NG. In Section 3, an integrated approach towards a complete MM solution is proposed. Finally, the paper concludes in Section 4.

2. The Evolution of Mobility Management

2G (GSM)

Mobile terminals and users authenticate and identify themselves by reporting their locations to their Home Location Registers (HLRs) through the MAP (Mobile Application Part) messages. The Visited Location Registers (VLRs) store caches of necessary user information locally. A mobile terminal updates its location when crossing location areas (LAs) which are sets of cells.

2.5G (GPRS)

Mobile terminals report their locations to their HLRs through SGSNs (Serving GPRS Support Nodes). The location area unit that GPRS uses is Routing Area (RA), which is typically a subset of one, and only one, GSM LA. This smaller granularity allows for signalling and paging over smaller areas, and thereby achieves a better optimisation of radio resources. GPRS co-operates with the GSM LA-based location management, resulting in a more efficient paging mechanism for mobile terminals that use GSM and GPRS simultaneously. Although IP networks were introduced between SGSNs and GGSNs (Gateway GPRS Support Node), GPRS is only a step preparing for other than using IP mobility by tunnelling IP, and in principle, the mobility management of 2G and 2.5G are both Link-Layer based and for terminal mobility only.

3G (UMTS)

In UMTS, mobile terminals report their locations to their HLRs through combined SGSNs and GGSNs. In a later phase, **Mobile IP** (RFC 2002) would be introduced for IP mobility. By then, SGSNs and GGSNs would have been integrated and the integrated node would act as the home agent of the mobile terminal. In contrast to the monopoly role in GPRS, UMTS SGSN shares mobility management with the UMTS Terrestrial Radio Access Network (UTRAN). For further thinning location management requirements, RAs are in turn partitioned into URAs (UTRAN Registration Areas) to better serve pico-cells. The IN (Intelligence Network) concept was evolved to facilitate service mobility. In UMTS, it is called Customised Applications for Mobile network Enhanced Logic (CAMEL).

In 3G, global roaming becomes more practical with GSM, GPRS and UMTS co-existing to cover a global area. The evolution approach of cellular generations, cumbersome as it is in a sense, facilitates the mobility management of the hybrid system. However, some modifications are required. In [2], a middleman is introduced between the home network and the visited network. It acts as HLR for the visited network and as VLR for the home network to reduce signalling response time and latter stage traffic. In [3], an inter-system handoff scheme was proposed based on the analysis of the boundary cell region between systems. This scheme seems to bring the systems together quite naturally by taking advantage of their existing handoff procedures. However, it is an indirection solution and the extra signalling time introduced has a very large impact on the overall handoff time for pico- and micro-cells. [4] illustrated the UMTS-GSM handoff procedure based on some modifications of GSM to facilitate the discovery of UMTS. This approach leads to a more direct solution.

Beyond 3G (Heterogeneous Systems)

Although NG still lacks a clear definition, some features that have been widely recognised are: all IP-based protocol stack, heterogeneous access networks, various user terminals, and multimedia data traffic (e.g., [5]). From the network point of view, heterogeneity presents the major challenge for mobility management. A user may roam over a series of networks during his global travels. (Figure 1, derived from [6]).

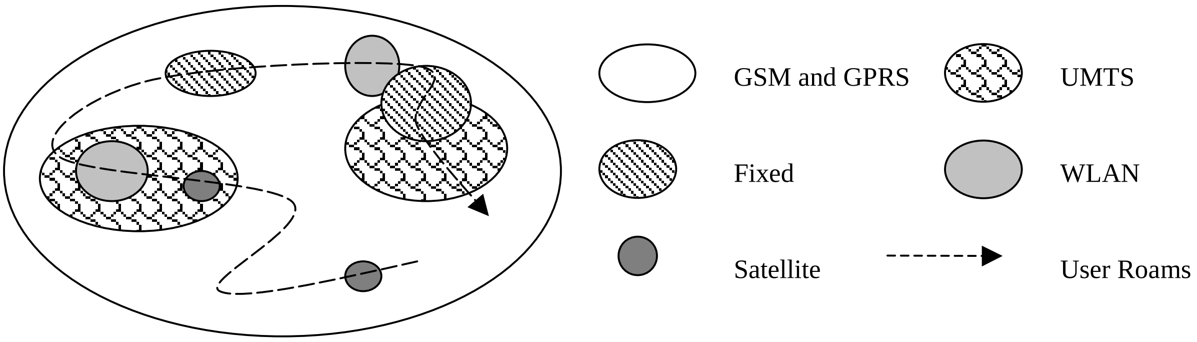

Figure 1. Seamless Roam over Ubiquitous Heterogeneous Networks

Several projects explore a degree of heterogeneity by investigation of somewhat limited combinations of systems. For instance, OWLAN [7] and WINNER [8] deal with roaming over WLAN and 2G/2.5G cellular systems, while SUITED [9] over terrestrial and satellite systems. However, they are basically targeted at terminal mobility only. [10] integrated **SIP** (RFC 2543) for personal and for service mobility, while relying on **Mobile IP** for terminal mobility. This leads to unsuccessful handoff between GSM and WLAN according to the trials. The OnTheMove project [11], using an enhanced Mobile IP, also met with the same problem when emulating UMTS bearers using GSM, WLAN and DECT. This problem mainly resulted from the long session set up time of GSM, and Mobile IP was unable to offer a solution. Since, as the current dominant technology, GSM will be a part of any heterogeneous system for some years to come while NG matures and is adopted, either modifications of GSM or a solution better than Mobile IP is needed. Furthermore, a generic framework that can accommodate all possible existing and emerging systems has not yet been devised.

From a protocol stack perspective, Network Layer is the lowest possible layer where convergence of the heterogeneous wireless systems can be developed since the major difference between them lies in the two lower layers. Besides, the desire to extend the great success of IP in the wired world to wireless leads to an all IP vision. Usually, Network-Layer mobility management protocols, together with traditional Link-Layer schemes, are used to deal with terminal mobility. So far, only **Mobile IP** has reached the RFC standard level and its basic form only suits macro mobility and suffers from the well-known triangular routing problem. Quite a few **micro mobility management proposals** (e.g., [12]) such as Cellular IP, HAWAII and TeleMIP have been proposed but they are still under discussion in the IETF. The primary difference between them lies in the local routing schemes in a domain. For example, host-specific caches, encapsulation, multicast and so on. Most of them, if not all, rely on Mobile IP for inter-domain mobility management. Therefore, a long-term solution is still needed.

SIP has been chosen by the 3GPP for call signalling but also could be used for Application-Layer mobility management [13] although no standardised specifications for this purpose have been worked out. In principle SIP is an Application-Layer multimedia signalling protocol. Personal, session and service mobility are all largely in the application level as well. Therefore, with the help of SIP infrastructure, SIP provides a framework and has the potential capabilities to support advanced high-level mobility management by augmented signalling [14]. To provide personal mobility, the infrastructure can obtain the current device the user is using in the visited network. If the device changes its current IP address during a session due to the user's roaming, theoretically what the device needs to do is simply notify the corresponding host by the versatile SIP INVITE message (notification) so that the session can resume. However, the basic SIP has some difficulties in dealing with mid-session mobility since the IP issue belongs to Network Layer and is beyond SIP. For session mobility, SIP can INVITE another device to join the session and hand it over. But simultaneous handoffs of the two communicating parties have not yet been supported [15]. For service mobility, only initial proposals have been presented [16] and QoS issues have not been addressed. To sum up, SIP is able to facilitate an advanced mobility management scenario but currently it is more of a conceptual model rather than a mature solution, although **extensions of SIP** (or SIP-based solutions) are a promising approach. **Alternative approaches** such as Mobile People [17], ICEBERG [18, 19] and IPMoA [20] prefer to tackle part of this problem from their own points of view and take different angles, which could contribute to building parts of a complete MM solution.

3. Approach towards Complete Mobility Management

From the protocol stack perspective, since terminal mobility and the high-level mobility can be basically dealt with by Network and Application Layer respectively, it is very natural to combine these two layers' schemes for a complete solution. A representative example is the combination of Mobile IP and SIP. However, a naïve implementation of them simultaneously would lead to performance inefficiency and even system disorder when the intrinsic connections between them are considered, e.g., a live session is kept by both terminal and session mobility management. Therefore, an integration or co-ordination of the two schemes is needed (Figure 2).

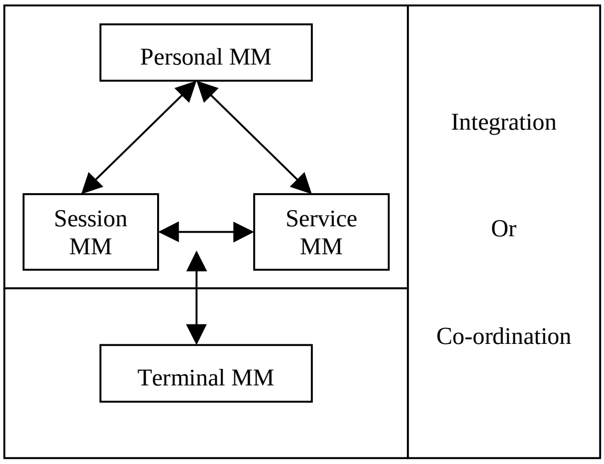

Figure 2. Integration of All Types of MM

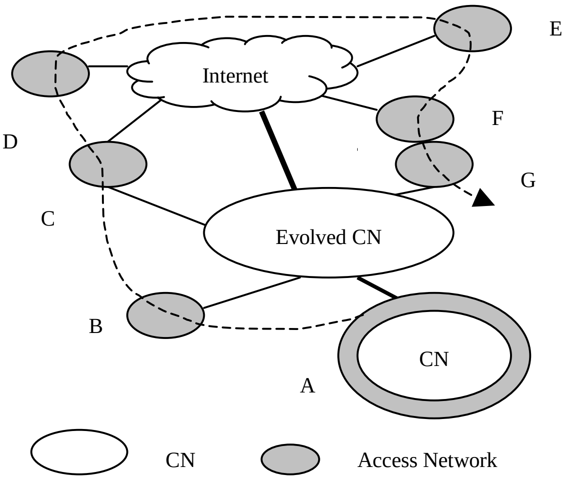

Figure 3. NG Hybrid Heterogeneous System

From the perspective of the network architecture, there are two frameworks. One is from the telecommunications field. Although a common evolved core network is expected, different generations with corresponding core networks are expected to co-exist for quite a long time due to technical or non-technical reasons globally. These independent networks are largely distributed, and thus a framework based on border observations is natural and the network-specific schemes can be light-weighted with fast respond to mobility issues.

The other is from the computer networks field. An Internet-centred uniform system is the long-term vision (e.g., ICEBERG), and thus centralised mobility management architecture that comprises highly shared functional entities is expected accordingly. In this framework, mobility management is a pure core network service. The biggest merit is that the development and deployment of new access points can be simplified. Besides, location databases can be managed easily and updated conveniently. One potential drawback of this architecture is that communications between two access networks have to be via the Internet in all circumstances, even if the caller and the callee are near to each other geographically (or the access networks are even overlapped) but far from the Internet. This can create another type of inefficient routing.

Our vision is that these two frameworks would co-exist and make up a hybrid heterogeneous system. Figure 3 illustrates such a scenario. Various heterogeneous access networks (from B to G) are connected to either core network, the Internet or the evolved cellular CN (or both, for example, C). A is a distributed independent network which has its own CN and access network. In this architecture, although both frameworks can have their own optimised MM schemes which have integrated all kinds of MM and interoperate, a uniform integrated solution is preferred. With this solution, mobility issues are managed in a flexible way, so that, for example two user can maintain a live session via either border-based or centre-based schemes, depending on the specific context, to achieve better overall performance.

4. Conclusion

There is no straightforward solution that takes account of the multiplicity of mobility management requirements and the heterogeneity of next-generation wireless system. Various approaches from multiple angles have been proposed to achieve advanced mobility management towards a complete solution.
Centralised architecture or a distributed one? Network core or edges? Network-Layer based or Application-level based? Mobile IP or SIP? These questions can not be simply answered with YES or NO for each approach has pros and cons and is probably not enough by itself. Therefore, a real open, accommodative, autonomous but still efficient mobility management solution is called for by the next-generation wireless system. We proposed in this paper a framework from both protocol stack and network architecture aspects as a step towards such a target.

References

[1] I. F. Akyildiz, J. McNair, J. Ho, H. Uzunalioglu, and W. Wang, Mobility Management in Current and Future Communications Networks, IEEE Network, July/August 1998
[2] Y-B. Lin and I. Chlamtac, Wireless and Mobile Network Architecture, John Wiley & Sons, Inc., 2001
[3] J. McNair, I.F. Akyildiz, and M. Bender, An Inter-System Handoff Technique for the IMT-2000 System, INFOCOM'00, Tel Aviv, Israel, March 2000
[4] H. Kaaranen, S. Naghian, L. Laitinen, A. Ahtiainen, and V. Niemi, UMTS Networks: Architecture, Mobility and Services, John Wiley & Sons, Ltd, 2001
[5] B. G. Evans and S. McLaughlin, Visions of 4G, IEE Electronics & Communication Engineering Journal, December 2000
[6] H. H. KARI, On Wireless Future, Presentation at MediaPoli seminar, November 1999
[7] J. Ala-Laurila, J. Mikkonen, and J. Rinnemaa, Wireless LAN Access Network Architecture for Mobile Operators, IEEE Communication Magazine, November 2001
[8] M.Ylianttila, M. Pande, J. Mäkelä, and P. Mähönen, Optimisation Scheme for Mobile Users Performing Vertical Handoffs between IEEE 802.11 and GPRS/EDGE Networks, Globecom'01, Texas, USA, November 2001
[9] P. Conforto, C. Tocci, G. Losquadro, A. Spazio, R. E. Sheriffand, P. M. L.Chan, and Y. F. Hu, Ubiquitous Internet in an Integrated Satellite-Terrestrial Environment: The SUITED Solution, IEEE Communication Magazine, January 2002
[10] T. Kanter, An Open Service Architecture for Adaptive Personal Mobile Communication, IEEE Communication Magazine, December 2001
[11] A. Fasbender, F. Reichert, E. Geulen, J. Hjelm, and T. Wierlemann, Any Network, Any Terminal, Anywhere, IEEE Communications Magazine, April 1999
[12] A. T. Campbell, J. Gomez, S. Kim, C-Y. Wan, Z. R. Turanyi, and A. G. Valko, Comparison of IP Micro-Mobility Protocols, IEEE Wireless Communications Magazine, February 2002
[13] H. Schulzrinne and E. Wedlund, Application-Layer Mobility using SIP, ACM Mobile Computing and Communications Review, Vol. 4, No. 3, July 2000
[14] F. Vakil, A. Dutta, J-C. Chen, M. Tauil, S. Baba, N. Nakajima, Y. Shobatake, and H. Schulzrinne, Mobility Management in a SIP Environment: Requirements, Functions and Issues, IETF Draft, December 2000, Work in Progress
[15] H. J. Wang and R. H. Katz, Mobility Support in Unified Communication Networks, Workshop on Wireless Mobile Multimedia, Rome, Italy, July 2001
[16] F. Vakil, A. Dutta, J-C. Chen, M. Tauil, S. Baba, N. Nakajima, Y. Shobatake, and H. Schulzrinne, Supporting Service Mobility with SIP, IETF Draft, December 2000, Work in Progress
[17] P. Maniatis, M. Roussopoulos, E. Swierk, K. Lai, G. Appenzeller, X. Zhao, and M. Baker, The Mobile People Architecture, ACM Mobile Computing and Communications Review, Vol. 1, No. 2, July 1999
[18] H. J. Wang, B. Raman, C-N. Chuah, R. Biswas, R. Gummadi, B. Hohlt, X. Hong, E. Kiciman, Z. Mao, J. S. Shih, L. Subramanian, B. Y. Zhao, A. D. Joseph and R. H. Katz, ICERGERG: An Internet Core Network Architecture for Integrated Communications, IEEE Personal Communications, August 2000
[19] B. Raman, R. H. Katz, and A. D. Joseph, Universal Inbox: Providing Extensible Personal Mobility and Service Mobility in an Integrated Communication Network, Workshop on Mobile Computing Systems and Applications, California, USA, December 2000
[20] B. Thai, R. Wan and A. Seneviratne, Personal Communications in Integrated Personal Mobility Architecture, ICON'01, Bangkok, Thailand, October 2001