Computer Science’2004, Technical University of Sofia, Bulgaria

Mobility Management Approaches for 4G Networks

Tobias Küfner[1], Dirk Hofmann[2], Christian Vogt[1]
[1] *Institute of Telematics, University of Karlsruhe, Germany*
[2] *Alcatel SEL AG, Corporate Research Center Stuttgart, Germany*

***Abstract**: The fourth generation (4G) of mobile communication networks will provide a multitude of real-time and non-real-time services on top of an IP-based platform. Subscribers can access these services over arbitrary radio-access technologies, including existing WPAN, WLAN, WMAN, and WWAN technologies as well as future air interfaces for broadband wireless access. Thereby, 4G networks pose high demands onto mobility management. This paper[1] provides a comprehensive overview on existing mobility management approaches, identifying, classifying, and comparing the features of the solution space.*

***Keywords:** Mobility Management, 4G Networks.*

1. INTRODUCTION

In the fourth generation (4G) of mobile communication networks, subscribers can access a multitude of IP-based services over arbitrary radio-access technologies, anytime and anywhere. The heterogeneity of these networks and the essentiality of quality of service (QoS), however, pose high demands onto the mobility management technology. Mobility management comprises two functions: location management and handover management. The former locates and tracks a mobile user for service delivery. The latter redirects an ongoing session from one access point to another. Both functions can be provided globally or with local scope. Global mobility management, also known as macro-mobility management, must enable a mobile node to switch between arbitrary access technologies, access networks and service providers. In contrast, local mobility management, or micro-mobility management, is limited to a certain area, a certain radio technology, or a certain service provider. Applications which are sensitive to service interruption or packet loss require seamless and lossless, intra-technology and inter-technology handovers. Thereby, existing WPAN, WLAN, WMAN, and WWAN standards will be integrated with future technologies for broadband wireless access. The remainder of this paper is structured as follows. Section 2 gives an overview on existing mobility-management approaches for 4G networks. In section 3, we compare the significant features of these approaches based on a classification of the solution space. We conclude in section 4.

2. OVERVIEW ON MOBILITY MANAGEMENT APPROACHES

Two hierarchical levels can be distinguished in 4G networks. The first level is formed by layer-2 access networks. Each layer-2 access network is connected to an IP-based layer-3 access network via one or more access routers. In cellular systems this second level of the hierarchy is usually referred to as the core network. The layer-3 access network is connected to the Internet backbone via one or more gateways. Most existing radio-access technologies provide support for layer-2 mobility, i.e. mobility within

---

[1] Parts of this work were funded by the Alcatel SEL AG, Corporate Research Centre Stuttgart, and the Federal Ministry of Education and Research of the Federal Republic of Germany.

a layer-2 access network invisible at layer 3. Exemplarily, we analyse layer-2 mobility in IEEE 802.11 wireless local area networks (WLAN) [4] and UMTS terrestrial radio-access network (UTRAN) [1]. In addition, some cellular systems allow for layer-3 handovers between different layer-2 access networks. Exemplarily, we analyse the GPRS Mobility Management (GMM) [2] in UMTS.

An 802.11 network consists of several access points interconnected by means of a distribution system – typically an Ethernet. When the estimated signal-to-noise ratio drops below a certain threshold, the mobile node scans for the best available access point in the layer-2 network and re-associates with it. Thereafter, a layer-2 update frame is broadcasted in order to register the mobile node's current location with all bridges and switches in the distribution system.

A UTRAN comprises one or more Radio Network Controllers (RNCs) which operate as access routers to the packet-switched domain (the layer-3 access network). Each RNC controls a group of Node Bs, which in turn operate as access points for several radio cells. Just like WLAN, UMTS supports hard handovers. If the previous and the new cell belong to different RNCs, the layer-2 handover is combined with a Serving Radio-Network Subsystem (SRNS) relocation (the layer-3 handover). More sophisticatedly, UMTS supports soft handovers during which a mobile node is simultaneously connected to multiple cells. The traffic to and from the mobile node is synchronised between all connections by the serving RNC.

The GPRS Mobility Management (GMM) handles layer-3 mobility within the UMTS core network on two levels by means of the GPRS Tunnelling Protocol (GTP). The first tunnel connects the serving RNC and the Serving GPRS Support Node (SGSN). The second tunnel connects the SGSN and the Gateway GPRS Support Node (GGSN). Each mobile node is assigned a global IP prefix which belongs to a GGSN. Location management is provided by the Home Location Register (HLR) and by the SGSN. The HLR keeps track of a mobile node's current SGSN, and the SGSN keeps track of the serving RNC as long as the mobile node is in active state. After a layer-2 handover, the serving RNC may no longer be the most suitable in terms of route efficiency. In this case, a SRNS relocation is executed, i.e. the GPRS tunnel endpoints are moved.

UMTS enables mobility within a layer-3 access network, and even inter-technology handovers between a UTRAN and a GSM/EDGE radio-access network (GERAN) are supported. Unfortunately, there is no comfortable way to apply GMM to arbitrary layer-2 access technologies, i.e. tight coupling. For this reason, IETF protocols are expected to enable global IP mobility in future 4G networks. Mobile IPv6 (MIP6) [5], the Stream Control Transmission Protocol (SCTP) [11] and the Session Initiation Protocol (SIP) [8] are the major candidates to manage mobility at layer 3, 4 and 5. These protocols allow for smooth integration with the existing Internet. In contrast, some other approaches such as the Host Identity Protocol (HIP) [7] propose more radical changes in order to enable mobility. It is interesting to note how the following approaches handle the ambiguous role of an IP address. On the one hand, an IP address – more precisely, its prefix – determines the current location of a node in the network topology. On the other hand, the address is used by the upper layers to identify the communication endpoint.

Mobility IPv6 (MIP6) uses two separate IPv6 addresses for these functions: the home address (HoA) as identifier and the care-of address (CoA) as locator. On its home link, the mobile node uses its HoA just like a stationary node. When the mobile node moves to a foreign link, it configures a CoA with the foreign prefix. The home agent, a dedicated router on the home link, maintains a binding between the HoA and the CoA. MIP6 offers two communications modes. In bidirectional-tunnelling mode, the home agent tunnels packets destined to the HoA to the CoA and vice versa. In route-optimisation mode, the correspondent node maintains its own bindings and is thereby able to communicate on the direct path with the mobile node.

SCTP is a reliable transport protocol which improves and enhances the service of TCP. With respect to mobility, the important advantage over TCP is that an SCTP end-point can be associated with multiple IP addresses. Thereby, multi-homed nodes gain fault tolerance. The SCTP enhancement defined in [10] allows mobile nodes to dynamically add or remove an IP address – a locator – to or from an endpoint identifier. This facilitates end-to-end handover management for SCTP sessions.

SIP provides a set of sophisticated location-management functions at session layer, including registration, invitation, conditional redirection, and call forking. SIP unified resource identifiers (URIs) are used for both purposes – as identifiers and locators. A SIP user registers his personal SIP URI together with his current location(s) at a SIP registrar. By aid of SIP proxies a SIP user can invite other registered users without knowing their current location. SIP can also be used for handover management during an ongoing session though it is not primarily designed for this purpose. Thereto, a mobile node sends a re-invite message with its new address to its correspondents.

HIP introduces the Host Identity (HI) namespace in an attempt to solve the identifier/locator problem. A HI serves as a layer-3 endpoint identifier and a public key of an asymmetric key pair at the same time. Public HIs are stored in the DNS together with the fully qualified domain name of a node. IPv6 applications use a 128-bit long cryptographic hash on the HI, the so-called Host Identity Tag (HIT), instead of an IPv6 address. Since each HIT can be mapped dynamically to multiple IP addresses, HIP enables mobility and multi-homing. At the start of a session, the correspondents run the HIP Base Exchange, a four-way handshake for mutual authentication and key exchange. The derived symmetric key is used to protect the actual session with IPSec Encrypted Security Payload (ESP). Handover management is carried out by means of direct peer notifications. Location management is provided by rendezvous servers.

The signalling traffic produced by purely global mobility-management approaches might result in unacceptable long handover latency and increased network load. In light of this background, local mobility-management solutions have been developed within the IETF as complements to global mobility management. As representatives, we analyse Cellular IP (CIP) [3], Hierarchical Mobile IPv6 (HMIP6) [9] and Fast Handover for Mobile IPv6 (FMIP6) [6].

Cellular IP (CIP) builds upon a tree-like layer-3 access network, in which all routers are mobility-aware. The root node acts as a gateway to the Internet. It hides the mobility inside its domain. After a mobile node has moved to a new access router, it sends a route-update message towards the gateway. This message is processed hop by hop until it reaches the crossover router, which is the first router already maintaining a route to the mobile node. Upon receipt of the route update, each router inserts or updates a forwarding entry for the mobile node. In addition to a hard handoff, CIP provides a semi-soft handoff. Thereto, the mobile node switches to the new link prior to the actual layer-2 handover and sends a semi-soft message. On receipt of this message, the crossover router starts to temporarily bi-cast all packets destined to the mobile on the previous and the new route.

Hierarchical Mobile IPv6 (HMIP6) introduces the Mobility Anchor Point (MAP) which acts as a local home agent to the mobile nodes inside an assigned HMIP6 domain. All access routers within this domain announce the MAP's presence in router advertisements. On receipt of such an advertisement, a mobile node can auto-configure a regional CoA (RCoA) from the MAP's prefix in addition to the on-link CoA (LCoA). Thereby, the mobile node is able to install a local binding between the RCoA and the LCoA at the MAP. In order to benefit from HMIP6, the mobile node uses its RCoA in MIP6 home and correspondent registrations. Then, packets destined to the mobile node’s RCoA are routed to the MAP, which tunnels them to the LCoA and vice versa.

Fast Handover for Mobile IPv6 (FMIP6) provides local mobility-management functionality immediately before and after a layer-2 handover in an effort to minimise handover latency. FMIP6 assumes that the link layer provides triggers to the mobility management indicating that a layer-2 handover to a certain access point is likely to occur in near future. Fast Handover enables the mobile node to discover its new access router and to configure a new care-of address prior to the layer-2 handover. When the mobile node switches to the new link, the previous access router tunnels all packets to the new care-of address. Thereby, the mobile node is able to update its location and receive packets in parallel. If the trigger is received sufficiently long enough before the connection to the previous link fails, FMIP6 facilitates a lossless handover.

3. CLASSIFICATION AND COMPARISON

Along with the overview given in the previous section, three axes for categorisation have already been identified: location management vs. handover management, global vs. local scope, and the layer at which mobility is handled. In this section, we further classify the solution space and compare the properties of the approaches.

In IPv6 networks, mobility is tightly coupled to addressing and routing. If a mobile node moves to a new subnet, either a new, topologically correct IP address must be configured or a new route to this node's IP address must be set up. The latter approach requires routers to maintain location information for all mobile nodes which cannot be reached by longest prefix match. This is called host-specific routing. At layer 2, host-specific routing is applied by almost every radio-access technology. CIP is a representative of host-specific routing at layer 3. All other approaches – including GMM, MIP6, SCTP, SIP, HIP, HMIP6, and FMIP6 – rely on address reconfiguration. In the IETF approaches, the mobile node configures and propagates a new care-of address in order to stay reachable and to keep ongoing sessions alive. Contrarily, in UMTS the mobile node is only aware of its fixed IP address which belongs to the GGSN. In fact, the mobile node’s current location in the layer-3 access network corresponds to serving RNC's IP address, but that IP address is only known to the SGSN.

Location management always requires a mobility-aware middle box, such as a HLR, which facilitates initial contact to a mobile node. However, handover management can be carried out by pure end-to-end approaches such as SIP, SCTP, HIP, and MIP6 in route-optimisation mode. This class of solutions directly notifies the correspondents about a new IP address. In contrast, correspondent-transparent approaches hide this information in a mobility-aware middle box such as a GGSN or home agent. The second class of solutions is further split into tunnelling approaches and host-specific routing approaches, depending on how traffic is forwarded to a mobile node.

The following end-to-end arguments apply to all mobility-management solutions which do not rely on a middle box. Firstly, end-to-end approaches scale well in the number of mobile nodes because there is no shared forwarding agent which could otherwise become a bottleneck. Secondly, end-to-end approaches are more robust against network failures than correspondent-transparent solutions because there is no single point of failure. An exception is MIP6 in route-optimisation mode because some signalling of a correspondent registration still goes through the home agent. Thirdly, end-to-end solutions can be installed and modified in a flexible manner by communication peers without the support of an ISP. Fourthly, end-to-end approaches keep the cost for developing and maintaining access-network routers low. From the viewpoint of the Internet design, end-to-end approaches are the natural way to support IP mobility. Contrariwise, host-specific routing approaches shall be deemed to be the worst choice with respect to scalability, robustness, flexibility and cost.

With respect to delay-sensitive applications, route efficiency is another important argument. When handover management is based on end-to-end notifications, the route from a mobile node to its correspondent node is the most efficient with respect to the underlying routing protocol. In correspondent-transparent approaches, the route efficiency depends on the location of the mobility-aware forwarding agent. Hierarchical approaches such as CIP and HMIP6 typically assume a tree-like topology. In this case, the middle box is roughly on-path and the route is more or less direct. HMIP6 enables a mobile node to always register with the most efficient MAP. In UMTS, the SRNS relocation serves the same purpose. In contrast, very inefficient routes might be taken when MIP6 is applied in bidirectional-tunnelling mode. MIP6 route optimisation solves this problem. Fast handover will in general lead to some route inefficiency as well, but, due to its short-time nature, this is less significant.

The end-to-end arguments mentioned above and the built-in route efficiency are two advantages of end-to-end solutions. Nevertheless, there are also arguments in favour of correspondent-transparent approaches. Firstly, the correspondent does not need to implement any mobility-related protocol. Mobility-enabling software must be installed only at the mobile nodes and the involved middle boxes. Secondly, correspondent-transparent solutions provide for a limited degree of location privacy, in that the correspondent node is not able to determine or track the current geographic location from the mobile node's current on-link prefix. Thirdly, if the number of simultaneous correspondents is large, the high number of location updates may strain the resources of the mobile node and, possibly, the mobile node's access network. This is of special importance for mobile servers and peer-to-peer applications.

Especially with break-before-make handovers, the layer-3 handover latency is crucial. It may result in unacceptable service interruptions, if the signalling path is long, or handovers occur frequently. The shortest latency can be achieved by means of host-specific routing updates. Moreover, bi-casting as applied in the CIP semi-soft handoff can minimise packet loss. The layer-3 handover latency achieved by address reconfiguration approaches consists of two parts: the address-configuration latency and the location-update latency. HMIP6 can minimise the location-update latency. FMIP6 can move the location-update latency out of the critical period of time and eliminate the address-configuration latency.

Finally, the requirements and restrictions that mobility management puts on mobile applications play an important role. All approaches at layer 2 and 3 are transparent to the upper layers, i.e. they enable to continue any kind of application after a handover. An exception is HIP, which partially requires modifications to the way applications deal with IPv6 addresses and HITs, respectively. Solutions at layer 4 and 5 restrict the set of supported applications in general. SCTP-based handover management requires the application to use SCTP at transport layer. SIP must be explicitly integrated with the application. Since it is unlikely that this will happen to all IP applications, solutions at layer 4 and 5 would benefit only selected applications. Anyhow, SCTP and SIP can be incrementally deployed.

4. CONCLUSIONS

In this paper, we classified and compared mobility-management approaches for 4G networks. We exemplarily analysed link-layer mobility support in WLAN and UTRAN as well as the GPRS mobility management as representatives of approaches in existing systems. We evaluated mobility support in IPv6, SCTP, and SIP, as well as the Host Identity Protocol as candidates for global IP mobility management. Finally, we looked at Cellular IP, Hierarchical Mobile IPv6 and Fast Handover as representatives of local mobility management. On the one hand, end-to-end approaches turned out to be superior in terms of scalability, robustness, flexibility, infrastructural cost, and route efficiency. On the other hand, correspondent-transparent approaches significantly reduce the signalling load, enable location privacy, and facilitate communications with mobility-agnostic correspondents. The advantages of both classes can be joined in a single solution such as Mobile IPv6 or in a compound approach. Since quality of service is a mandatory feature of 4G networks, local mobility management has to perform seamless and lossless handovers. Nevertheless, mobility support at the network-layer rather than above is required in order to continue any IP application after moving to a new subnet. We conclude that multiple approaches at different layers should be combined in a hierarchical and complementary manner in order to fulfil the requirements of all IP-based applications in future 4G mobile communication networks.

5. REFERENCES

[1] 3rd Generation Partnership Project (2004) *UTRAN overall description (Release 6)*. 3GPP TS 25.401 V6.4.0.

[2] 3rd Generation Partnership Project (2004) *General Packet Radio Service (GPRS); Service description; Stage 2 (Release 6).* 3GPP TS 23.060 V6.5.0.

[3] Campbell, A., Gomeza, J., Wan, C.-Y., Kim, S., Turanyi, Z. Valko, A. (2000) *Cellular IP.* IETF Internet-Draft. Work In Progress.

[4] Institute of Electrical and Electronics Engineers (1999) *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. ANSI/IEEE Std 802.11.

[5] Johnson, D., Perkins, C., Arkko, J. (2004) *Mobility Support in IPv6*. IETF RFC 3775. Standards Track.

[6] Koodli, R. (2004) *Fast Handovers for Mobile IPv6.* IETF Internet-Draft. Work In Progress.

[7] Moskowitz, R., Nikander, P., Jokela, P., Henderson, T. (2004) *Host Identity Protocol.* IETF Internet-Draft. Work In Progress.

[8] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., Schooler, E. (2002) *SIP: Session Initiation Protocol*. IETF RFC 3261. Standards Track.

[9] Soliman, H., Catelluccia, C., El Malki, K., Bellier, L. (2004) *Hierarchical Mobile IPv6 mobility management (HMIPv6)*. IETF Internet-Draft. Work In Progress.

[10] Stewart, R., Ramalhoand, M., Xie, Q., Tuexen, M., Conrad, C. (2004) *Stream Control Transmission Protocol (SCTP) Dynamic Address Reconfiguration.* IETF Internet-Draft. Work In Progress.

[11] Stewart, R., Xie, Q., Morneault, K., Sharp, C., Schwarzbauer, H., Taylor, T., Rytina, I., Kalla, M., Zhang, L., Paxson, V. (2000) *Stream Control Transmission Protocol.* IETF RFC 2960. Standards Track.