Securing access to mobile networks beyond 3G

Christian Gehrmann

Ericsson Mobile Communications AB

Günther Horn

Siemens AG

Nigel Jefferies

Vodafone Group R&D

Chris Mitchell

Royal Holloway, University of London

The way forward for mobile networks is indicated by a number of themes including the move to IP-based networks, roaming across a variety of access networks, an increasingly user-focussed approach to services and applications, and terminals that are increasingly distributed, multifunctional and personalized devices. In the light of this, there is a need to determine how security technology can be used to protect the interests of all participants. This paper identifies the pertinent research issues, and describes how the IST SHAMAN project is dealing with them.

I. IP-based mobile networks with heterogeneous access networks

Some trends are already discernible that are likely to characterise post-3G network architectures:

- The use of IP will be the key to global connectivity. It will not be limited to the core network, but will extend far into the radio access network. The protocols will be largely specified by the IETF.
- There will be a gradual evolution rather than revolution from 3G systems.
- Post-3G systems will be characterised by the combined use of several different access network technologies to provide the user with the most appropriate support for his services [Comb01].

An objective of IST project SHAMAN is to specify a security architecture for such post-3G mobile systems. This can only be developed when a reasonably detailed specification of such a mobile system is available. One of the first tasks for the project was the evaluation of other ongoing post-3G activities to select one or a small number of network architectures for post-3G systems that could serve as a reference. Three promising sources for network reference architectures of post-3G systems have been identified:

- the Mobile Wireless Internet Forum [MWIF];
- IST project BRAIN (with its successor MIND) [Brain];
- ETSI activities on the (loose or tight) coupling of HIPERLAN with UMTS [Hiper].

The first two sources consistently follow the “all-IP” approach, based on the use of IETF-defined techniques, whereas the third is a more short-term activity using existing protocols. The number of reference architectures may be reduced further, but, given that the future direction of post-3G systems is still uncertain, it is considered wise not to focus on a particular solution too early.

The essential functional layers in an “all-IP” mobile system that need to be secured have been identified as mobility management, quality of service and session control.

These layers are seen as being largely independent in the sense that each one can be provided in a system without the other two and therefore constitutes a valid object of investigation on its own. In particular, it is possible to study security solutions for each of these layers separately. This approach has already been taken by the IETF. For example, solutions for mobility management include Mobile IP (v4 and v6), Hierarchical Mobile IP, Regional Registration and various other micro-mobility protocols, each with their own security solutions. Solutions for quality of service include the use of RSVP, for which quite different security solutions are under consideration at the IETF. Finally, a popular protocol for session control is the Session Initiation Protocol (SIP), again with its own security solution proposed by the IETF.

In contrast to the IETF approach, SHAMAN (in common with the MWIF and BRAIN activities that serve as references for SHAMAN) aims at systems which incorporate more than one of these layers. Therefore, the investigation of security solutions for individual layers can only be an intermediate step to a single security architecture. This is necessary to avoid high complexity and low performance of the system. In other words, security has to be integrated across all the functional layers. In this respect, security is comparable to management functions like Operation & Maintenance and Accounting.

Our priority is securing access-to-network services. Other security objectives such as securing end-to-end communication between users of conversational services, securing commercial transactions or the download of objects from application servers are studied with secondary priority. Securing access-to-network services involves the following tasks.

I.I Entity authentication

The user and the network have to be mutually authenticated. Authentication is typically coupled with session key establishment for use with the confidentiality and integrity mechanisms (see below).

Some radio access networks have their own mechanisms for authenticating mobile nodes. However, these mechanisms are commonly performed at Layer 2 and are of local significance only. For global roaming, authentication mechanisms based at Layer 3 or higher must be used possibly involving contact with an entity in the home network. It may then be possible to disable Layer 2 authentication mechanisms (such as Bluetooth or WLAN link authentication).

Authentication can be divided into initial authentication (which may be done with the help of the home network) and subsequent authentication (or re-authentication) required if the lifetime of the initial registration expires or if an access point is changed. The subsequent authentication will not usually require involvement of the home network if appropriate keys for re-authentication have been distributed in the initial authentication.

(On-line) involvement of the home network may also be unnecessary when authentication and authorisation are based on certificates. The choice between secret-key and public-key based authentication methods may lead to quite different security architectures.

I.II Key Establishment

Key establishment comprises the generation and distribution of session keys required for the encryption and integrity protection of the traffic on the wireless link or for subsequent (local) authentication between mobile node and access network. Key agreement must rely on some pre-shared secret between the network access points and the mobile client, or on the use of certificates. Some steps may be performed off line (such as management of the Ki in GSM) or the pre-shared secret may be reduced to a minimal human-memorable passkey. Solutions for the latter case are being developed, for example, using IEEE 802.1X with EAP for access control and authentication, and using a passkey-authenticated key agreement protocol.

Furthermore, there may be a hierarchy of keys, e.g. a key agreement key that is used to derive a link authentication key (as e.g. in Bluetooth) which in turn is used to derive link encryption and integrity keys.

I.III Authorisation

In an IP-based system, AAA servers and protocols typically take care of entity authentication and authorisation. It is likely that SHAMAN will also follow this approach.

I.IV Confidentiality and integrity of user and signalling data

The traditional approach in mobile systems has been to protect only the most vulnerable part of the network, namely the radio access. This, however, leads to solutions that are particular to the radio access technology. The definition of a global mechanism for session key establishment (or, more generally, security association establishment) which is compatible with the different confidentiality and integrity mechanisms is required. It is worth considering moving confidentiality and integrity one layer up from the radio link layer to the network layer where IP is available as a unifying glue, but this approach may leave radio access network specific messages unprotected. A decision cannot be taken without a careful security analysis.

I.V Support for global roaming

Global roaming is a key requirement on any post-3G mobile system, and has a strong impact on security. A difficulty especially for key establishment arises from the fact that, in general, a roaming user will have had no previous contact with the entities of the network into which he roams. This means that the visited network entity responsible for controlling access and the user have to find a common point of trust to help them complete their security tasks. In addition, the user has to have a means to communicate with the point of control in the visited network without being allowed to access other resources of that network before authentication and access control have been successfully performed. We refer to this latter requirement as "secure initial access".

In second and third generation mobile systems, the common point of trust is the home service provider of the user. Secure initial access is realised by establishing a dedicated signalling channel between user and VLR or SGSN. In other systems, other common points of trust may be used, such as Trusted Third Parties, which issue certificates to users and network entities and which need not be contacted on-line. The use of certificates that require the use of so-called public key techniques would be a novelty for mobile systems and presents challenges that are addressed in the final section of this paper. For a while, it seemed that Mobile IPv6 with its mandatory support for IPSec would provide the general solution. But, at least as long as no global Public Key Infrastructure for mobile users is available and as long as the performance problems associated with the use of the Internet Key Exchange protocol in a mobile environment persist, this hope does not seem justified. This is also underlined by the fact that the key management problems associated with Mobile IPv6 are currently considered unsolved, and that secret-key based solutions have been discussed for Mobile IP [Kink].

I.VI A unified access procedure

An important factor in the success of GSM (and, so we expect, UMTS) is the use of a single security token, the (U)SIM-card, by which the user can obtain global access. We need to understand whether and how security procedures for access to networks services and applications can be based on the use of a single security token, such as a single smart card, with possibly multiple security applications, and whether and how multiple smart cards providing different functions will co-operate.

II. A security architecture for future terminals and applications

In addition to common trends in their individual development, computers, phones, and networks also show a pattern of convergence. Computers with the right accessories are able to carry voice calls, powerful mobile phones are clear competitors to handheld computers, and local networks (e.g. laptop, Personal Digital Assistants (PDA), phone) have multiple access options to other networks. The boundaries between the computer, the phone, and the (local) network are becoming more and more blurred, resulting in a personal, networked communication and computing environment that has been referred to both as a *distributed terminal* and as a *Personal Area Network* (PAN). Cable connections are cumbersome to use. Short-range infrared communication, IrDA [IrDA] is less awkward, but the line-of-sight requirement limits its usefulness. In the future, we expect *short-range* wireless communication like Bluetooth [BSIG] to be widely used and available in most devices.

Another important technical aspect of mobile terminals is that the user has the possibility to configure the terminal in several different ways. This requires dynamic software upgrades to the terminals. Currently it is possible for the user to download his favourite applications for personal computing devices like laptops and PDAs and this is about to be possible also for mobile terminals. In addition, the manufacturer and mobile operators would like to have the possibility to upgrade and reconfigure the terminals. Hence, we foresee a situation where future terminals will be *dynamically reconfigurable* and *distributed.*

A large number of wireless network types can be covered within the PAN concept. In order to limit the scope of our work we have developed a simple PAN reference model. The model has the following main attributes:

- We are only considering wireless communication in a licence-free spectrum. One example is the Bluetooth wireless technology.
- Wireless *ad hoc networks* in general are considered, but only relatively small ad hoc networks (in the order of 10 devices) without any advanced routing capabilities.
- We use a simplified PAN device model with three basic entities: a device that supports one or more local communication bearers, a device with a global network interface, and a (possibly standalone) subscription module.
- PAN devices with limited resources and computing capabilities are considered.
- We do not assume that a supporting infrastructure (for key management or administration) is available to the PAN. However, we consider the usage scenario where at least one of the PAN devices has a global network connection through a "global" or "local" interface.

We consider a distributed terminal concept where many core functions in mobile terminals are accessible over the PAN. This includes core terminal functions such as access to the SIM/USIM card and connection management functions. An example is the Association of Radio Industries and Business (ARIB) interface description [MTTA]. One usage scenario we are considering is the so-called *car scenario.* In this scenario, the car's built-in interface does not have its own subscription module; thus, it can only be operated if another unit with a subscription module is available. We have made a threat analysis of the car scenario as well as of other real-world scenarios. For the car scenario, the identified threats are:

- Malicious content downloaded into the PAN devices.
- Access from an untrusted PAN device to core functionality in another PAN device (core functionality includes access to the SIM/USIM).
- Passive eavesdropping of the inter-device PAN communication.
- Impersonation of a PAN device.

II.1 Security requirements

Our first step in developing a new terminal and application security architecture has been to identify the security requirements that our architecture should satisfy. The requirements are listed using a new role model that has been included into our PAN reference model. The identified roles are: PAN component user, PAN component owner, Application service provider, Communications service provider, Authorisation authority, and PAN component manufacturer, PAN manager and the Intruder.

II.II Trust model

One major problem in ad hoc network is the lack of supporting infrastructure and pre-defined trust relations between the different PAN components. Short-term and long-term trust relations need to be created dynamically. Trust can be on user, service, or device level. To perform efficient access control and define communication security policies, we need to develop a trust model applicable to our PAN reference model.

II.III Control and configuration

The configuration and control of the PAN involves some kind of user interaction during its entire life cycle from the set-up of the network to its termination. The PAN user may be required to have the capabilities of the PAN manager or PAN component administrator. The control and configuration application needs to provide sufficient intelligence to inform/warn the user or even to prevent the user from certain configuration and control actions once it has sufficient evidence to do so. In order to achieve this the user needs to be able to request basic information about the device, e.g. the trust status and access rights. Our security architecture includes a special PAN component database that will handle this information. The architecture is not limited to a model where a human user performs configuration and hence the role of the PAN manager and the role of PAN component manager might be given to an application.

II.IV Internal communication security

There is a central security manager for each PAN in the SHAMAN security architecture. This functional entity is able to handle all communication security settings. Whenever a PAN is configured or reconfigured to allow terminal components to join the PAN, some initial access control is performed. For this to happen the new terminal component must be identified and keys to be used to secure the communication must be agreed. There are several possibilities such as pre-distributed symmetric keys, password-based key exchange methods, and PKI.

II.V Secure execution

The ability to reconfigure and software upgrade a device means that malicious software is a serious threat to the whole system. On the other hand, dynamic software upgrading gives greater flexibility and better user convenience. A security architecture for safe software downloading is defined within the 3GPP MExE work [MExE]. We have analysed the MExE work and studied how it can be applied for dynamic configurable distributed terminals. A research project [TeSSA] provides a solution for secure execution of mobile code on a set of distributed hosts. SPKI (Simple Public Key Infrastructure) certificates are used to carry permissions for a given resource. MExE combined with the TESSA architecture could be the basis for secure execution in our architecture.

III Public Key Infrastructures

Work is required to develop public key infrastructures (PKIs) to support the security architectures discussed above. This will require the identification of PKI requirements for the networks and terminal discussed above, and gaining an understanding of the facilities offered by existing PKI schemes, including ITU-T X.509, PKIX, ISO/IEC 15945, SPKI and WTLS.

Based on these ongoing studies, we have derived an understanding of where existing work falls short of meeting the needs of SHAMAN. In some areas, the need for PKI support is clear, and the existing work is known to be of limited value. These areas include the following.

III.I Interoperability

PKI interoperability issues will clearly arise when a user moves from one environment to another and 'picks up' new terminal components. We are studying possible solutions to the interoperability problems, including the definition of methods for 'translating' certificates from one domain to another. This translation will need to include mappings of policy statements and, where necessary, translation of certificate formats (see [BM00]).

III.II Authorisation issues

The emerging X.509-based PKI schemes use policy identifiers in a range of ways. This includes indicating the policy under which a certificate has been issued, and limiting the ways in which certification paths can be constructed. However, relatively little work has so far been performed on automatic processing of policy statements themselves. With an ever-growing number of mobile components, possessing certificates issued by large numbers of different CAs, such automatic processing is likely to become an important requirement to enable devices to interact freely and securely.

IV Conclusion

IST SHAMAN [SHAMAN] is tackling security issues arising from identified features of mobile systems beyond 3G. These include global roaming, heterogeneous access networks, the distributed personal environment, and the need for a supporting public-key infrastructure. This two-year project, which started in December 2000, has the following participants: Vodafone Group R&D, Royal Holloway, University of London, Siemens Atea n.v., Nokia Corporation, Ericsson Radio Systems, T-Nova Deutsche Telekom Innovationsgesellschaft mbH, Giesecke & Devrient, Siemens Aktiengesellschaft.

References

[Comb01] P. Combelles: "What do we call 4G?", WWRF kick-off meeting, WG4, Munich, 6-7 March, 2001, http://www.ist-wsi.org

[MWIF] MWIF Technical Report MTR-004, http://www.mwif.org

[Brain] http://www.ist-brain.org/

[Hiper] http://www.etsi.org/technicalactiv/hiperlan2.htm

[Kink] http://www.vpnc.org/ietf-kink/

[BSIG] Bluetooth Special Interest Group (SIG), http://www.bluetooth.com/

[MTTA] "MT-TA Interface Description", Association of Radio Industries and Businesses (ARIB) , http://www.arib.or.jp/IMT-2000/V130Jan01/T12/0_T12coverV130.html.

[MExE] 3GPP TS 23.057 "Mobile Station Application Execution Environment (MExE)".

[TeSSA] Telecommunications Software Security Architecture (TeSSA), http://www.tml.hut.fi/Research/TeSSA/

[BM00] N. Borselius and C.J. Mitchell, 'Certificate translation', in: *Proceedings of NORDSEC 2000 - 5th Nordic Workshop on Secure IT Systems*, Reykjavik, Iceland, 12/13 October 2000, pp.289-300.

[SHAMAN] http://www.ist-shaman.org