Security, Privacy, and Trust in Mobile Systems

Marco Cremonini
Università di Milano, Italy

Ernesto Damiani
Università di Milano, Italy

Sabrina De Capitani di Vimercati
Università di Milano, Italy

Pierangela Samarati
Università di Milano, Italy

INTRODUCTION

Access to general purpose information and communication technology (ICT) is not equally distributed on our planet: developed countries represent about 70% of all Internet users, while its percentage of Internet hosts has raised from 90% in 2000 to about 99% in 2002.

Things change dramatically if we look at mobile and wireless technology: developing countries already represent about 40% of mobile connections in 2000, with a foreseen growth rate that is faster in developing countries than in developed ones in the period 2000-2005 (mainly due to India and the People's Republic of China). This trend is driven by the new perspectives offered by mobile electronic technology applications that provide an alternative to poor telecommunication infrastructures still common in many developing countries. The technological evolution in wireless data communications is introducing a rich landscape of new services relying on three main technologies:

- proximity (or personal) area networks (PANs), composed of personal and wearable devices capable of automatically setting up transient communication environments (also known as *ad hoc* networks);
- wireless local area network technologies (WLANs); and
- a third generation of mobile telecommunications (3G), gradually replacing General Packet Radio Service (GPRS) and the related set of technologies collectively called "2.5 Generation" (2.5G).

PAN is a new technology bringing the "always connected" principle to the personal space. On the other hand, 3G systems and WLANs have coexisted for a while; what is new is their interconnection, aimed at decoupling terminals and applications from the access method. 3G mobile networks already provide video-capable bandwidth, global roaming for voice and data, and access to Internet-rich online content.

Thanks to their increasing integration, PANs, WLANs, and 3G networks will extend the user's connectivity in a complementary and hierarchical manner; in the fullness of time, they will provide all the functionalities of an *Integrated Services Multimedia Network* (ISMN), enabling a whole set of new business models and applications.

The fusion of these technologies will eventually result in an ultimate ubiquitous wireless system that will be operated from anywhere, including homes, business locations, vehicles, and even commercial aircrafts.

However, although wireless communications provide great flexibility and mobility, they often come at the expense of security. Indeed, wireless communications rely on open and public transmission media that expose new vulnerabilities in addition to the security threats found in wired networks. A number of specific open issues and even inherent dangers, some of which had been already identified and described in the early stages of wireless technology adoption, are yet to be solved (Howard, 2000). For instance, with wireless communications, important and vital information is often placed on a mobile device that is vulnerable to theft and loss. In addition, information is transmitted over the unprotected airwaves, and finally, 3G networks are getting smaller and more numerous, causing opportunities for hackers and other abusers to increase.

BACKGROUND

2G and 2.5G Mobile Authentication

GSM 2G systems introduced the *Subscriber Identity Module* (SIM) cards containing the user's identity and an authentication key (i.e., a shared secret key) supposed to last for the entire duration of the subscription. SIM-based authentication does not require any user action, other than entering the familiar four-digit *Personal Identification Number* (PIN) into the terminal. With GSM, a temporary user identity is allocated by the area operator where the user is located and is reassigned to another user as soon as the original requestor leaves the area. With the advent of 2.5G systems, enhanced by the *General Packet Radio Service* (GPRS), overlaying, certificates-based authentication became possible (Smith, 2002).

3G Authentication and On-the-Air Confidentiality

In the design of 3G systems like UMTS, a new security architecture was introduced (Blanchard, 2000). The new approach maintained backward compatibility with GSM, while trying to overcome some perceived weaknesses of 2G systems. Like in 2G systems, 3G systems identify users by means of the identity stored in the SIM. Differently from 2G systems, 3G authentication was designed with the following features:

- **Mutual Authentication:** Both the user and the network operator are identified in the authentication exchange.
- **Key Freshness:** Assurance that authentication information and keys are not being reused.
- **Integrity of Signaling:** Protection of service messages, for example, during the encryption algorithm negotiation.
- **Strong Encryption:** Strong cryptography, obtained via a combination of key length and algorithm design, is performed inside the core network rather than at the periphery.

Early Identity Management Systems

Starting from the late '80s, many examples of *Identity Management (IM)* systems have been proposed. In 1985, Chaum (1985) considered a device that helps the user with payment transactions and upholds the user's privacy. Clark (1993) proposed the *digital individual*, the individual's data shadow in the computer system which can be compared to user's identity.

Digital security and, more generally, digital identity management have grown fast in recent years, especially in mobile scenarios where personal communication and new computing devices will generate new security and integrity requirements for users and service information (Jendrike et al., 2002; Roussos & Patel, 2002).

MOBILE IDENTITY MANAGEMENT

Personal Identity Management in 3G Mobile Systems

Privacy and security issues related to mobile systems have been often described in terms of traditional security functionalities (e.g., access control, integrity, authentication, non repudiation, availability, and confidentiality). However, recent developments in ICT-based business models revealed the necessity to approach the concept of privacy and security more broadly, embracing not only the technical aspects, but also socioeconomic issues (Kagal, Parker, Chen, Joshi, & Finin, 2003). The ongoing transition from monolithic and localized systems, mainly based on single technology and weakly opened to integration, towards multi-application, multi-access, multiplayer, distributed, and heterogeneous scenarios, is generating a context in which mobile applications and systems could play a strategic role. In other words, technology and business must be strongly Internet-worked with users' social dynamics, standards, policy, and regulation to create a digital identity management framework where digital identity is conceived as "an electronic representation of individuals' or organizations' sensitive information" (Damiani, De Capitani di Vimercati, & Samarati, 2003). Support offered by this framework is crucial for building and maintaining trust relationships in today's globally interconnected society because it:

- offers adequate security and availability;
- permits the presentation of a different subset of the users' identity depending on the ongoing and perceived application and communication context;
- guarantees that identity, personal data, and user profile (including location-based information) are safeguarded and no thefts will happen.

A *Digital Identity Management Framework* is realized by taking into consideration both the architecture of the framework, and those external elements that may influence an identity manager (e.g., regulations, standards, and so on). In particular, with respect to the framework's architecture, the following main elements can be recognized.

User

The service requestor associated with a profile. The digital identity management framework should let the user keep her desired level of privacy depending on the situ-