Secure mobile commerce

by S. Schwiderski-Grosche and H. Knospe

M-commerce, or mobile commerce, is a major application domain for mobile devices, enabling users to perform commercial transactions wherever they go. However, these applications require a high level of security. In this paper, the special characteristics of m-commerce are identified and some important security issues considered.

1 Introduction

The term *e-commerce* (electronic commerce) denotes business processes on the Internet, such as the buying and selling of goods. There is a distinction between business-to-business (B2B) and business-to-consumer (B2C) markets. In the first case, the business processes are carried out between businesses; in the latter case, they are carried out between businesses and end consumers. This general definition of e-commerce does not say anything about the kind of device that the end user employs to gain access to the Internet. The underlying technology could be wireline (e.g. using a home PC as the end user device) or wireless (e.g. using a mobile phone as the end user device).

The term *m-commerce* (mobile commerce) is all about wireless e-commerce, that is where mobile devices are used to do business on the Internet, either in the B2B or B2C market. As such, m-commerce is a subset of e-commerce.

With the omnipresent availability of mobile phones (and other mobile devices), m-commerce services have a promising future, especially in the B2C market. Future applications include buying over the phone, purchase and redemption of ticket and reward schemes, travel and weather information, and writing contracts on the move. However, the success of m-commerce very much depends on the security of the underlying technologies. For example, today the chargeback rate for credit card transactions on the Internet is 15%, versus 1% for point-of-sale (POS) credit card transactions. Chargeback rates grow to 30% when digital products are sold[1]. For m-commerce to take off, fraud rates have to be reduced to an acceptable level. As such, security can be regarded as an enabling factor for the success of m-commerce applications. In this paper, we discuss two main areas of m-commerce that are relevant to security, namely:

- *network technology*. In m-commerce, all data is transmitted via a mobile telecommunication network. Here, we consider existing network and service technologies for second generation (2G), third generation (3G) and other wireless systems.
- *m-payment* (mobile payment). Doing business on the Internet requires payment for goods and services. M-payment systems have different requirements and characteristics from e-payment systems. Here, we give an overview of current payment technology.

In the next section, we provide background information on m-commerce. Section 3 is devoted to network technologies for m-commerce. Different approaches to m-payment will be discussed in Section 4.

2 M-commerce and its security challenges

Definition of m-commerce

There are many definitions of the term m-commerce[1,2]. Common to all definitions is that a terminal or mobile device is employed to communicate over a mobile telecommunication network. There are different views as to the purpose of this communication. Some definitions restrict m-commerce to transactions involving a monetary value, whereas other definitions generalise the term to services that involve communication, information, transaction, and entertainment. Here we define m-commerce as using a mobile device for business transactions performed over a mobile telecommunication network, possibly involving the transfer of monetary values.

Mobile devices

M-commerce is not just about using mobile phones as end user devices. The following list gives an overview of different kinds of mobile devices:

- mobile phone
- PDA (Personal Digital Assistant)
- smart phone (the smart phone combines mobile phone and PDA technology into one device)
- laptop computer
- earpiece (as part of a personal area network).

Each mobile device has certain characteristics that influence its usability, such as:

- size and colour of display
- input device, availability of keyboard and mouse
- memory and CPU (Central Processing Unit) processing power
- network connectivity, bandwidth capacity
- supported operating systems (e.g. PalmOS, Microsoft Pocket PC)

- availability of internal smartcard reader (e.g. for a SIM [Subscriber Identity Module] card in mobile phones).

Depending on these factors, the services that the end user can receive differ considerably. Moreover, depending on the network technology used for transmission, the bandwidth capacity varies and influences the kind of services that the end user is able to receive.

In mobile phones, there exist three solutions to internal smartcards: *single SIM*, *dual chip*, and *dual slot*. Single SIM, in which all confidential user information is stored on one smartcard, is the solution that is most widely available today. Dual chip means that there are two smartcards in the mobile phone, one for user authentication to the network operator and one for value-added services such as m-payment or digital signature. A dual-slot mobile phone has a SIM card and a card slot for a full-sized external smartcard. With this solution different cards can be used one after the other. Moreover, the cards can also be used in traditional POS and ATM (Automated Teller Machine) terminals.

Differences to e-commerce

In comparison to e-commerce, m-commerce has both advantages and disadvantages. The following list summarises the advantages of m-commerce[1]:

- Ubiquity—the end user device is mobile, so that the user can access m-commerce applications in real time at any place.
- Accessibility—the end user is accessible anywhere, at any time. Accessibility is related to ubiquity. It is probably the greatest advantage over e-commerce applications involving a wired end user device.
- Security—depending on the specific end user device, the device offers a certain level of inherent security. For example, the SIM card commonly employed in mobile phones is a smartcard that stores confidential user information, such as the user's secret authentication key. As such, the mobile phone can be regarded as a smartcard reader with smartcard.
- Localisation—a network operator can localise registered users by using a positioning system, such as GPS (Global Positioning System), or via GSM (Global System for Mobile) or UMTS (Universal Mobile Telecommunication System) network technology, and offer location-dependent services. These services include providing local information about hotels, restaurants, and amenities, travel information, emergency calls, and mobile office facilities.
- Convenience—the size and weight of mobile devices and their ubiquity and accessibility make them an ideal tool for performing personal tasks.
- Personalisation—mobile devices are usually not shared between users. This makes it possible to adjust a mobile device to the user's needs and wishes (starting with the mobile phone housing and ring tones). On the other hand, a mobile operator can offer personalised services to its users, depending on specified user characteristics (e.g. a user may prefer Italian food) and the user's location (see above).

The following list summarises the main disadvantages of m-commerce:

- Mobile devices offer limited capabilities (such as limited display). Between mobile devices these capabilities vary so much that end user services will need to be customised accordingly.
- The heterogeneity of devices, operating systems, and network technologies is a challenge for a uniform end user platform. For this reason, standardisation bodies consisting of telecommunication companies, device manufacturers, and value-added service providers integrate their work (see Section 4). For example, many current mobile devices implement an IP (Internet Protocol) stack to provide standard network connectivity. At the application level, the Java 2 Micro Edition (J2ME) offers a standardised application platform for heterogeneous devices.
- Mobile devices are more prone to theft and destruction. According to a government report, more than 700 000 mobile phones are stolen in the UK each year[3]. Since mobile phones are highly personalised and contain confidential user information, they need to be protected according to the highest security standards.
- Communication over the air interface between the mobile device and the network introduces additional security threats (e.g. eavesdropping, see Section 3).

Security challenges

As mentioned earlier, m-commerce is not possible without a secure environment, especially for those transactions involving monetary value. Depending on the points of view of the different participants in an m-commerce scenario, there are different security challenges. These challenges relate to:

- the mobile device—confidential user data on the mobile device as well as the device itself should be protected from unauthorised use. The security mechanisms employed here include user authentication (e.g. PIN [Personal Identification Number] or password authentication), secure storage of confidential data (e.g. the SIM card in mobile phones) and security of the operating system.
- the radio interface—access to a telecommunication network requires the protection of transmitted data in terms of confidentiality, integrity, and authenticity. In particular, the user's personal data should be protected from eavesdropping. Different security mechanisms for different mobile network technologies (i.e. in 2G, 3G, and other systems) will be presented in Section 3.
- the network operator infrastructure—security mechanisms for the end user often terminate in the access network. This raises questions regarding the security of the user's data within and beyond the access network. Moreover, the user receives certain services for which he/she has to pay. This often involves the network operator, who will want to be assured about correct charging and billing.
- the kind of m-commerce application—m-commerce

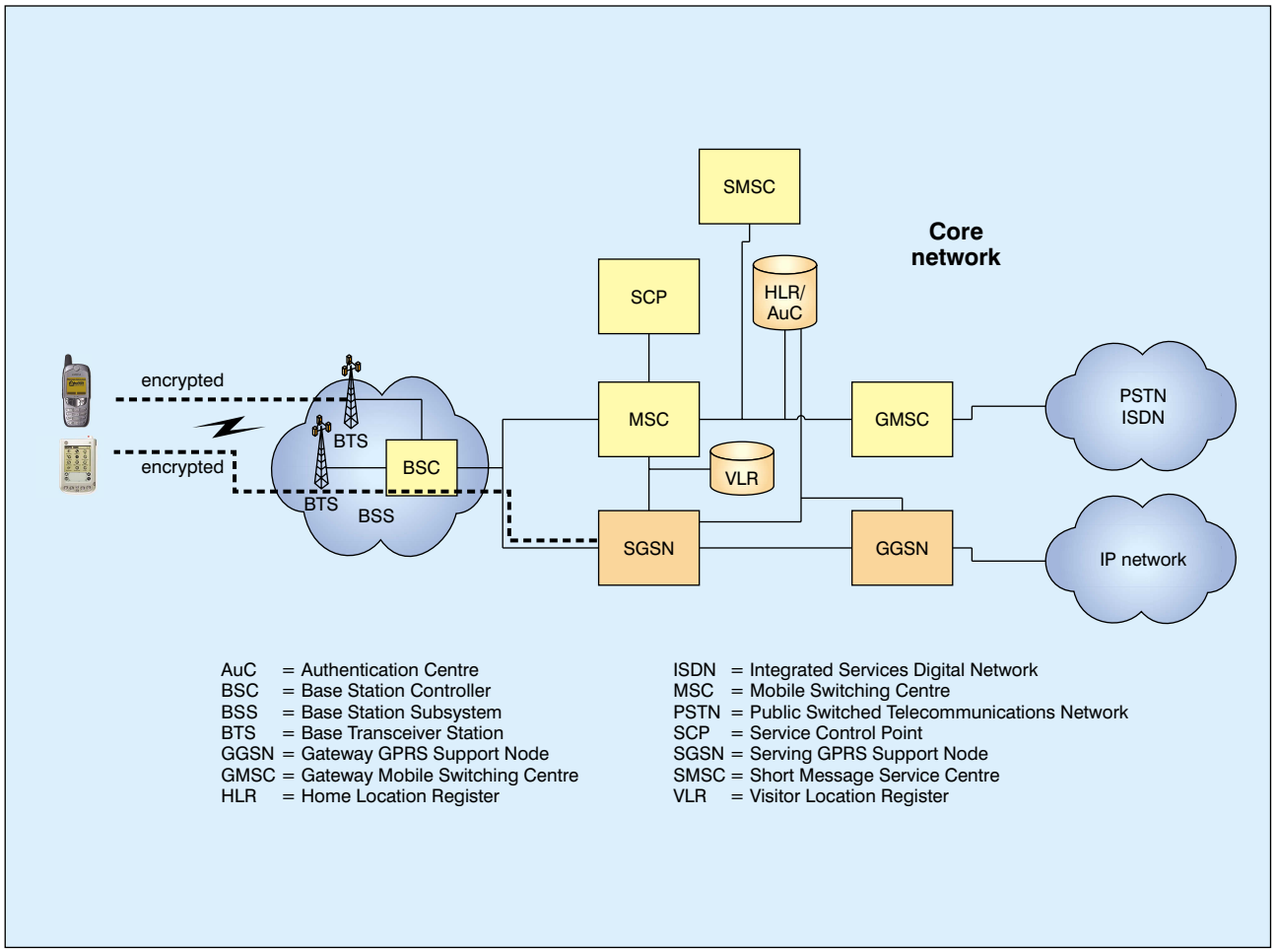

Fig. 1 GSM architecture

applications, especially those involving payment, need to be secured to assure customers, merchants, and network operators. For example, in a payment scenario both sides will want to authenticate each other before committing to a payment. Also, the customer will want assurance about the delivery of goods or services. In addition to the authenticity, confidentiality and integrity of sent payment information, non-repudiation is important. Section 4 will review security mechanisms for m-payment.

3 Security technologies relevant for m-commerce

In this section, we give an overview of the technologies that are relevant to secure m-commerce transactions. We focus on those network and service technologies that are specific to mobile devices. The security architecture of current and potential future mobile systems has been studied in the IST SHAMAN project*.

Security of network technologies

We first discuss the security of network technologies used for mobile commerce.

*SHAMAN—Security for Heterogeneous Access in Mobile Applications and Networks—is project 2000-25350 of the European Commission's IST (Information Society Technologies) programme. See http://www.ist-shaman.org.

GSM

GSM (Global System for Mobile Communication) is the current European standard for mobile communications. Since GSM handsets are popular and widespread, they have to be considered as the major device for mobile commerce at the moment. In the first years of GSM (beginning of the 1990s), the devices were very limited with respect to their capabilities other than telephony. Dial-in data sessions over circuit-switched connections were possible but relatively slow (9·6 kbit/s) and required a separate device (computer), which reduces mobility. As the GSM core network was extended with more and more data service elements, cellular phones also became more powerful. A number of data services were established:

- SMS (Short Message Service) allows the exchange of 160-character short messages over the signalling channel.
- WAP (Wireless Application Protocol) permits access to Internet content formatted in WML (Wireless Mark-up Language). At first, WAP used only circuit-switched connections.
- HSCSD (High Speed Circuit Switched Data) provides higher data rates by channel bundling.
- GPRS (General Packet Radio Service) extends GSM with packet-oriented services. With GPRS, the mobile node can stay 'always on' without blocking a connection

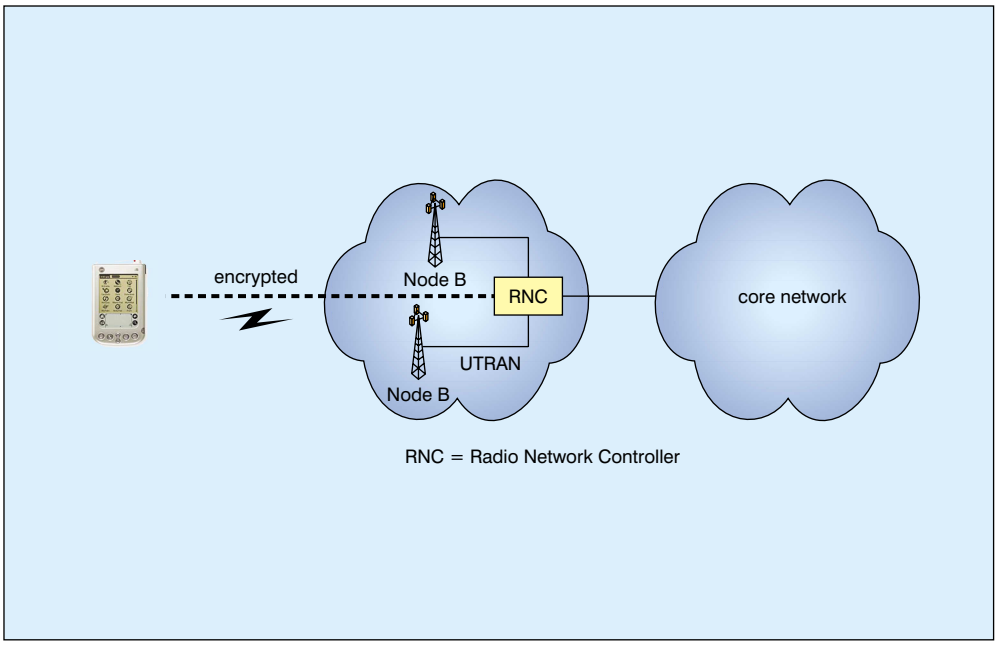

Fig. 2 UTRAN system

timeslot with the base station. GPRS can also be used as a bearer service for WAP and SMS.

The basic architecture of GSM including GPRS, IN (Intelligent Network) and SMS components is depicted in Fig. 1. The mobile station communicates over the wireless interface with a base transceiver station (BTS) which is part of a base station subsystem (BSS). The base station controller (BSC) is connected with a mobile switching centre (MSC) and a serving GPRS support node (SGSN). The latter two are the central switching components for circuit- and packet-switched data.

When a customer subscribes, the GSM home network assigns the mobile station a unique identifier, the international mobile subscriber identity (IMSI), and an authentication key K_i. The IMSI and the secret authentication key K_i of the mobile station (MS) are stored in the subscriber identity module (SIM), which is assumed to be tamper proof. On the network side, the IMSI, K_i and other information are stored in the home location register (HLR) and authentication centre (AuC).

GSM provides the following security features for the link between the mobile station and the network[4,5]:

- IMSI confidentiality
- IMSI authentication
- user data confidentiality on physical connections
- connectionless user data confidentiality
- signalling information element confidentiality.

GSM provides the basic security mechanisms for m-commerce transactions. In particular, the *mobile customer authenticates* towards the network with a challenge/response protocol based on the secret key K_i. Furthermore, the *wireless link between the mobile station and the BTS is encrypted* with a symmetric key, which is also derived from K_i. For GPRS, encryption extends from the mobile to the SGSN. The secret key K_i is never sent over the network. But there are weaknesses[5]: since the network is not authenticated, a false base station can perform a 'man-in-the-middle' attack. The base station can suppress IMSI confidentiality and encryption and this is not even visible to the mobile station. Furthermore, there are known attacks against GSM authentication and encryption algorithms, but they usually require (short-time) physical access to the SIM.

UMTS

UMTS (Universal Mobile Telecommunication System) is the next generation (3G) mobile telecommunication system and a further development of GSM. The major difference to GSM is the radio network (UTRAN, the UMTS Terrestrial Radio Access Network) with its transition to wideband code division multiple access (WCDMA) radio technology. Two new network components, the radio network controller (RNC) and Node B, are introduced in UTRAN. Furthermore, the security protocols have been modified and now the RNC is responsible for de-/ciphering.

The main components of the GSM/GPRS core network with the MSC, SGSN etc. can be reused or evolved to UMTS. This is depicted in Fig. 2.

In general, the security architecture of UMTS is carefully designed to fix the security weaknesses of GSM[5,6]. As described above, the main problems of GSM originate from two facts: authentication is one way (the mobile station does not authenticate the network), and encryption is optional. In UMTS, authentication is mutual, and encryption is mandatory unless the mobile station and the network agree on an unciphered connection. In addition, integrity protection is always mandatory and protects against replay or modification of signalling messages. Sequence numbers in authentication vectors protect against reuse of authentication vectors by network impersonators. Furthermore, UMTS introduces new cipher algorithms and longer encryption keys. Thus, UMTS does not seem to have any obvious security holes. UMTS security is discussed in detail in another paper in this issue[7].

WLAN

The IEEE standard 802.11 specifies families of wireless local-area networks (WLANs) which operate in the unlicensed 2·4 GHz and 5 GHz band. The standards specify the physical layer (PHY) and the medium access control (MAC) layer. For the network layer and above, WLAN employs a classical IP stack. A number of commercial products (even for PDAs) are available, and IEEE 802·11b, offering 11 Mbit/s raw bandwidth, is currently very popular. When operated in the infrastructure mode, the mobile station attaches to an access point (AP) which provides connectivity to fixed net IP networks (e.g. the Internet) or to other mobile stations.

In the default mode, WLAN does not provide any security. This means that a mobile attacker can eavesdrop and manipulate all the wireless traffic with standard tools.

In order to provide a certain level of security, the IEEE defined WEP (Wired Equivalent Privacy). WEP was designed to provide:

- authentication to protect the association to an access point
- integrity protection of MAC frames
- confidentiality of MAC frames.

The protection is based on secret WEP keys of either 40 or 104 bits. Concatenated with a clear text initialisation vector, the secret key serves as input for the RC4 stream cipher. But it has been shown[8] that *authentication and integrity protection is completely insecure and encryption at least partly insecure*. It suffices for an attacker to intercept a single successful authentication exchange between a mobile station and the access point to be able to authenticate without knowing the secret keys. Furthermore, since a CRC (Cyclic Redundancy Check) sum is used for integrity protection, an attacker can modify the data and adapt the checksum accordingly. For example, if the position of commercially sensitive information (e.g. an amount) within a datagram is known, the corresponding bits can be XORed with any value. With a large number of intercepted frames, the WEP keys can even be recovered, breaking the encryption. Furthermore, since the WEP keys are *network keys*, preserving their secrecy is difficult for private networks and impossible for public WLAN hotspots.

In recent work of the IEEE Task Group on Security (TgI), the new security standard[9] IEEE 802.1X has been adopted. 802.1X is a framework for port-based network access control that employs the Extensible Authentication Protocol (EAP[10]) for a variety of authentication and key management mechanisms, e.g. certificate-based transport layer security (TLS). But the weaknesses of WEP cannot be remedied by the new authentication and key management schemes in 802·1X. The IEEE is currently working towards a new standard (WEP2), and a number of proposals are in circulation.

Another approach is to use virtual private network (VPN) technologies, and in particular IPsec, in order to establish *network layer* security. The IPsec protocol (or more specifically the ESP Tunnel protocol) is an Internet standard[11] for the protection of IP packets between two nodes (e.g. a mobile station and a security gateway). This architecture is depicted in Fig. 3. Note that link-layer-specific information (e.g. MAC addresses) is still unprotected.

Bluetooth

Bluetooth is a wireless technology developed by the Bluetooth Special Interest Group[12] and is mainly aiming at *ad hoc* piconets and connections to peripheral devices. Bluetooth is also operating in the unlicensed 2·4 GHz band and can be considered as a *de facto* standard. The Bluetooth specification defines a complete protocol stack, so, unlike WLAN, it is not restricted to IP connectivity. Although raw bandwidth is limited to 1 MBit/s, the Bluetooth technology will probably often be used in the future to connect devices in the personal environment, which makes it relevant for m-commerce.

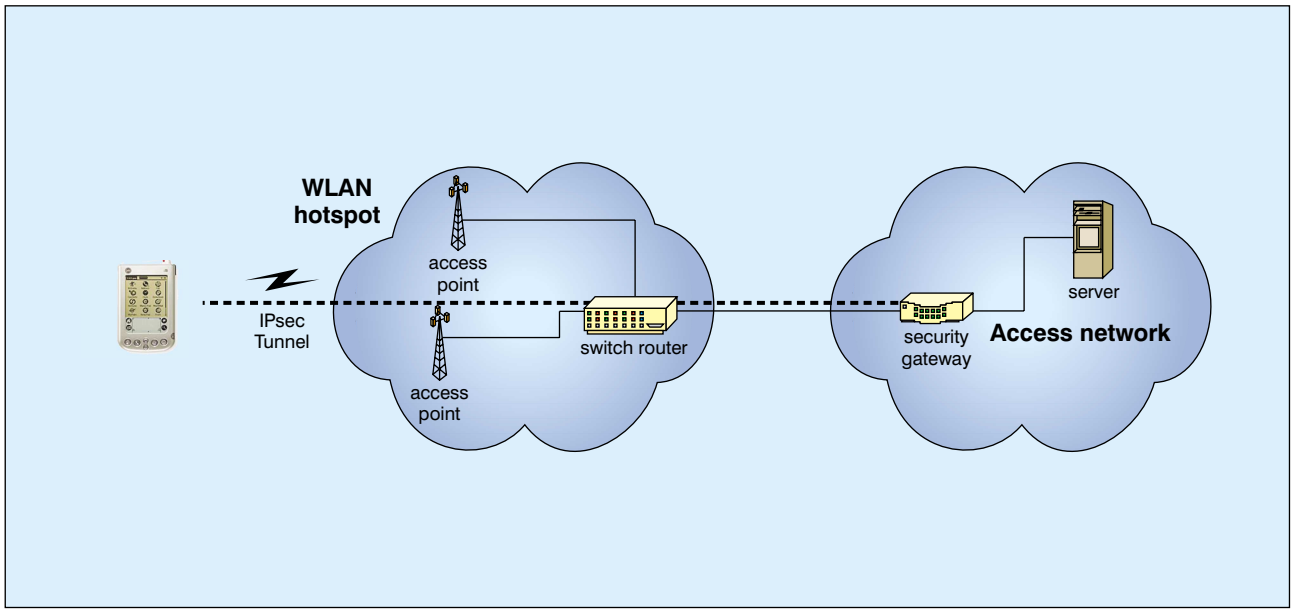

Fig. 3 WLAN security with IPsec

Bluetooth specifies three security modes, including 'no security'. It provides *link layer security* with a challenge–response protocol for authentication and a stream cipher encryption of user and signalling data[13]. When the connecting devices do not share a key in advance, they have to establish an initialisation key in a pairing procedure. This is based on a PIN, which must be entered into both devices (or imported from some application).

Bluetooth can currently be considered secure for small *ad hoc* networks, provided the pairing happens in a safe environment and the PIN is strong enough. The existing attacks are still theoretical in nature. However, *privacy requirements* may not be met since the unique Bluetooth device address allows the tracing of personal devices and hence their owner.

Transport layer security

The above technologies provide (to some extent) security for the wireless link between the mobile customer and the access network or access device. If the access network is considered secure and the m-commerce transaction is completely handled within the network, this may be sufficient. But often, an m-commerce transaction involves parties outside the access network (merchant, payment service provider etc.). In this section, we discuss *end-to-end security* for mobile devices. This protects applications that communicate over an IP port. More information, in particular on certificates and public key infrastructure (PKI), can be found in another paper in this issue[14].

SSL/TLS

The SSL/TLS (Internet Secure Socket Layer[15]) protocol is by far the most widely used Internet security protocol. Its main application is the HTTPS protocol (HTTP over SSL), but it may also be used as a standalone protocol. SSL requires a bidirectional byte stream service (i.e. Transmission Control Protocol, TCP). SUN has implemented a client-side version of SSL for limited devices called KSSL (Kilobyte SSL). KSSL does not offer client-side authentication and only implements certain commonly used cipher suites, but it has a very small footprint and runs on small devices using the J2ME platform.

WTLS

The WAP Forum has standardised a transport layer security protocol (WTLS) as part of the WAP 1 stack[16]. WTLS provides transport security between a WAP device (e.g. a mobile phone) and a WAP gateway that performs the protocol transformation to SSL/TLS. Hence, no real end-to-end security is provided and the WAP gateway needs to be trusted.

Note that the WAP Forum now proposes a WAP 2 stack[17], which is a classical TCP/IP stack on a wireless bearer medium. This permits end-to-end SSL/TLS sessions.

Service security

Here, we discuss the security of network services that can be used for m-commerce transactions.

Intelligent network

With the introduction of intelligent network (IN) technology to GSM networks, additional services could be realised. The IN architecture for GSM (called CAMEL, Customised Application for Mobile Enhanced network Logic[18]) was adapted from the fixed network standard ETSI Core INAP, and was originally designed for circuit-switched calls (CAMEL phase 1 and 2). The IN is triggered during call handling at the mobile switching centre if the home location register entry indicates subscription to an IN service. With CAMEL phase 3[19], the IN services can also be applied to SMS and to packet data services. The IN component service control point (SCP) controls the call or data service via the CAMEL Application Part (CAP) protocol, which runs on top of the SS7 (Signalling System Number 7) protocol.

Prominent examples of IN services are the transformation of dialled numbers (e.g. to realise virtual private nets) and prepaid services. The IN platform provides some flexibility for the generation of m-commerce services. IN handling can, for example, be triggered by a specific called party, a calling party, a USSD (see below) string (requiring CAMEL phase 2), a mobile originating SMS (requiring CAMEL phase 3) or a mobile terminating SMS (requiring CAMEL phase 4).

The security of an IN service depends on the underlying GSM or UMTS network security (see above) and on the specific characteristics of the service application.

Parlay/OSA

Parlay/OSA (Open Service Access) is an initiative of the industry (Parlay group), ETSI and the 3rd Generation Partnership Project and aims at introducing standard interfaces to network services[20]. The IN platform and SS7 based protocols like INAP (Intelligent Network Application Protocol) and CAP are relatively complex and the generation of services is reserved to operators and manufacturers. Now Parlay offers standard application programming interfaces, which allows service provisioning on IT platforms using standard middleware (e.g. CORBA). The Parlay/OSA framework provides gateway functionality between applications and service capability servers with access to the mobile core network. M-commerce applications can then access network functionality, e.g. enquire about status and location of a mobile user, send messages or place calls. Parlay/OSA applications are portable among networks, which is usually not the case for IN services.

Security is an important issue, since Parlay/OSA potentially opens the core network to intruders. The Parlay/OSA framework offers authentication and encryption on the application layer[21], but the security also depends on the underlying network architecture, i.e. firewalls, and strict policies should protect core network components.

SMS

SMS (Short Message Service) is a very popular data service for GSM networks. Although SMS messages are

limited to 160 characters, a considerable number of m-commerce scenarios are based on this service. The sender and receiver of an SMS message are identified by their international mobile subscriber identities, which an attacker cannot forge without breaking the GSM/UMTS security mechanisms (e.g. by cloning a SIM card). Hence SMS messages can be used for authentication (at least towards the network). Furthermore, SMS data is transmitted in the GSM (UMTS) signalling plane, which ensures the confidentiality of messages. However, the protection ends in the GSM or UMTS network, there is no end-to-end security, and the network operator and its infrastructure (e.g. SMSC, Short Message Service Centre) must be trusted (when no other security mechanisms are applied to the SMS message, cf. the section on SIM/USIM applications below).

USSD

The GSM Unstructured Supplementary Service Data (USSD) service allows data communication between a mobile station and either the HLR, VLR (Visiting Location Register), MSC or SCP in a way transparent to the other network entities. Unlike the asynchronous SMS service, a USSD request opens a session that may induce other network operations or a USSD response before releasing the connection. Mobile-originated USSD may be thought of as a trigger for a network operation. USSD works with any mobile phone since the coded commands are entered in the same way as a phone number (e.g. *123#1234567890#).

With USSD, roaming can be offered for prepaid GSM customers before IN services (CAMEL) are implemented in a network. Another USSD application (requiring CAMEL phase 2) is replenishing a prepaid account by incorporating the voucher number in a USSD string. In principle, any transaction, e.g. a payment operation, could be triggered by USSD data.

USSD possesses no separate security properties; instead it relies on the GSM/UMTS signalling plane security mechanisms.

SIM/USIM application toolkit

The SIM and USIM application toolkits (SAT and USAT, respectively) allow operators and other providers to create applications that reside in the SIM/USIM. These applications can, for example, send, receive and interpret SMS or USSD strings. Currently, there exist banking applications using the SIM application toolkit. In Reference 22, a format for secured packets is defined. This permits the sending application (e.g. the one residing on the SIM card) to send protected messages to the receiving application (which, for example, runs at a payment service operator). The required security mechanisms are:

- authentication
- message integrity
- replay detection and sequence integrity
- proof of receipt and proof of execution
- message confidentiality
- indication of the security mechanisms used.

However, it depends on the applications whether these security mechanisms are implemented and whether their cryptographic strength is sufficient.

4 M-payment

This section gives an introduction to payment mechanisms for m-commerce. First, we will motivate the heterogeneity of payment system solutions, before categorising e-payment and m-payment systems, respectively. Then we will give examples of m-payment systems.

Background on payment systems

E-payment systems provide means for payment of goods or services over the Internet. In contrast to conventional payment systems, the customer sends all payment-related data to the merchant over the Internet; no further external interaction between customer and merchant is required (e.g. sending an invoice by mail or confirmation by fax). To date, there exist more than 100 different e-payment systems[23–25].

There exist a large number of payment systems for e- and m-payments. The following distinctive features of payment systems motivate this diversity:

- time of payment
- payment amount
- anonymity issues
- security requirements
- online or offline validation.

Time of payment denotes the relation between the initiation of a payment transaction and the actual payment. In prepaid payment systems, the customer's account is debited before the payment and the amount is stored, for example on smart cards, in specific customer accounts or as electronic cash. In pay-now payment systems the customer's account is debited at the time of payment (for example, ATM card or debit card with PIN), and in post-payment systems payment can be regarded as a 'payment promise' where the merchant's account is credited before the customer's account is debited (for example, credit card systems).

The *payment amount* has an influence on the design of electronic payment protocols. For example, payments in the order of 1 € are only viable if the incurred computational and communications overhead is kept small. Accordingly, there is a distinction between:

- *micropayments* (up to about 1 €)
- *small payments* (about 1 to 10 €)
- *macropayments* (more than about 10 €).

Electronic payment systems often originate with conventional payment systems. As such, cash-like payment systems should provide *anonymity* to the customer. There are different degrees of anonymity: complete anonymity means that the customer remains anonymous to the merchant and the bank. However, in

many payment systems, only partial (or no) anonymity can be provided.

The *security requirements* of electronic payment systems differ. Generally, integrity, authentication, authorisation, confidentiality, availability, and reliability issues need to be considered, depending on the specific requirements of an electronic payment system.

Offline payment validation means that no third party (e.g. a bank or credit card institution) is involved during the payment procedure, whereas *online payment validation* involves some kind of background payment server as a trusted third party. The latter causes an additional communication overhead, but reduces certain risks, e.g. double spending.

The above discussion summarises some distinctive features of payment systems. There are other issues such as:

- overhead imposed on customers and merchants (e.g. installation of software, registration)
- performance (e.g. response times)
- cost incurred per payment transaction
- fulfilment of the ACID (Atomicity, Consistency, Isolation, Durability) principle for payment transactions (i.e. transactions have to be executed all or nothing, leaving the system in a consistent state, and their effect should be durable)
- national or international deployment.

The above list of distinctive features gives an idea of the complexity and variety of payment systems.

Categorisation of e-payment systems

E-payment systems are typically modelled on conventional payment systems. As such, there are the following categories[25]:

- *Direct cash:* In direct-cash-like payment systems, the customer withdraws money from the issuer, that is the third party interacting with the customer (for example, a bank or service provider), and hands payment tokens for the payment amount to the merchant. The merchant deposits the payment tokens with its acquirer, that is the third party interacting with the merchant (for example, a bank or service provider). The issuer and acquirer then settle the payment. This payment scenario is sketched in Fig. 4. Since digital cash is trivial to copy, direct-cash-like payment systems involve either tamper-proof hardware (i.e. smart cards) or online validation by the issuer (i.e. double spending test).
- *Cheque:* In this scenario, the customer hands

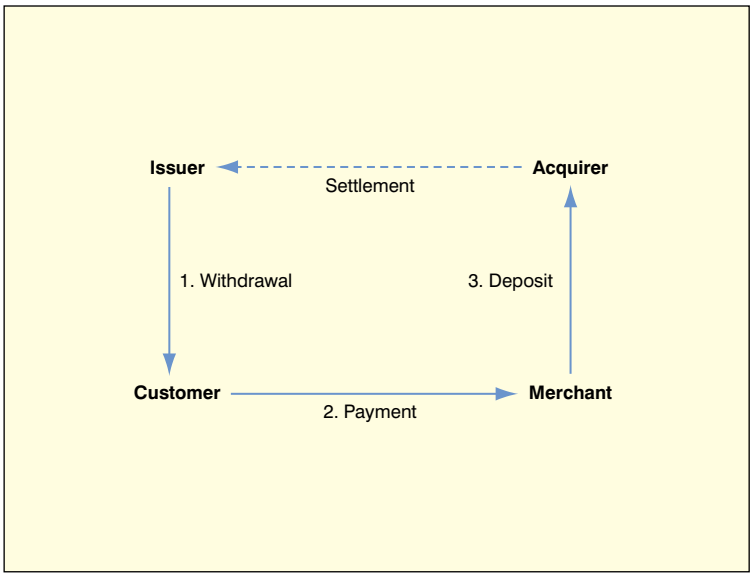

Fig. 4 Direct-cash-like payment system

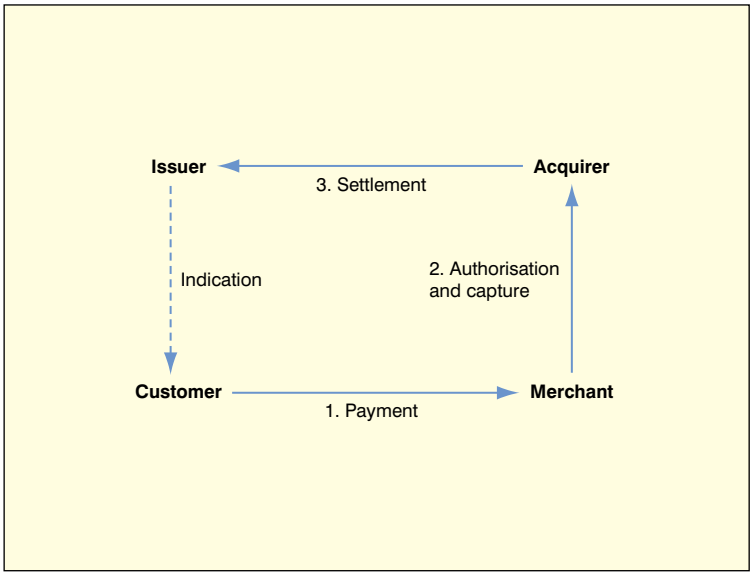

Fig. 5 Cheque-like payment system

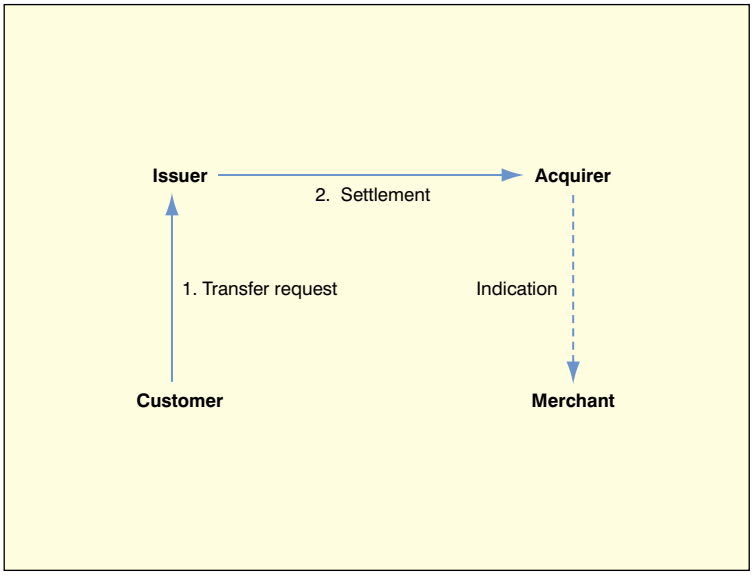

Fig. 6 Bank transfer payment system

a cheque (a payment authorisation) to the merchant. The cheque is presented to the acquirer who redeems it from the issuer. Cheque-like payments are sketched in Fig. 5.

- *Credit card:* In terms of the information flow, credit-card-based payment systems are similar to cheque-like payment systems, with the difference that credit-card-based payment systems use the existing credit card infrastructure for settling the payment.
- *Bank transfer:* The bank transfer model is sketched in Fig. 6. Here, the customer instructs the issuer to transfer money to the merchant's account at the acquirer. The merchant is notified of the incoming payment.
- *Debit advice:* This model describes the opposite case to the bank transfer model. The merchant instructs the acquirer to charge the account at the issuer. The customer is notified of the outgoing payment.

Categorisation of m-payment systems

Most e-payment systems are not suitable for use in a mobile context, that is using a mobile device and communicating over a mobile telecommunication network. This is due to the special characteristics of mobile devices and mobile telecommunications (see Section 2). In the following, we categorise m-payment systems according to the whereabouts of the customer's money.

Software electronic coins: In this case, monetary value is stored on the mobile device and the customer has full control of his/her money wherever he/she goes and whatever he/she does. An electronic coin is represented as a file containing, among other information, a value, a serial number, a validity period, and the signature of the issuing bank. Since software electronic coins are easy to copy, the validity of an electronic coin depends on its uniqueness in terms of its serial number. The customer transfers electronic coins to the merchant, who forwards them to the issuing bank for the 'double spending test'. In this test, it is checked whether the electronic coin has been spent beforehand. If yes, it is rejected. Otherwise, its serial number is entered into the double spending database and the money is credited to the merchant's account*. The generation and storage of electronic coins is an orthogonal problem. Due to the limitations of mobile devices, electronic coins may have to be generated and stored externally, until they are downloaded onto the mobile device.

Hardware electronic coins: In this case, monetary value is stored on a secure hardware token, typically a smartcard, in the mobile device. The presentation of electronic money is not important, as long as it is stored securely on the smartcard. Electronic money could be represented as a simple numeric counter. In order to get to the money, the customer's smartcard and the merchant's payment server authenticate each other and a secure channel is set up between them. Then, electronic money can be transferred from one to the other. This approach is quite attractive because smartcards provide an additional level of mobility. That means that the payment smartcard can also be used in point-of-sale transactions.

Background account: Here, the money is stored remotely on an account at a trusted third party. Depending on the specific payment system, the account could be a credit card account, a bank account, or an account held at the network operator. Common to all scenarios is that, on receipt of an invoice, the customer sends an authentication and authorisation message to the merchant that allows the trusted third party (that holds the account) to identify the customer and to verify the payment authorisation. The accounts can then be settled. There are numerous payment systems that fall into this category. The differences concern the nature of the trusted third party and the procedure to send authentication and authorisation data. For example, in some cases this data is sent in the clear (e.g. a credit card authorisation), not providing any security against eavesdropping, and in some cases this information is encrypted and digitally signed, providing anonymity to the customer (e.g. SET—Secure Electronic Transactions).

Examples of m-payment systems

In the previous section, we gave a categorisation of m-payment systems. In this section, we survey existing payment solutions for m-payment, as well as e-payment solutions that are suitable for mobile use.

Software electronic coins: There are several e-payment systems that are based on electronic coins. As for cash, one main advantage is that the customer can potentially remain completely anonymous to the merchant as well as to the bank, while staying in full control of the money. E-commerce solutions of this type include eCash, NetCash, and MilliCent. Due to the storage and processing constraints of mobile devices, an adaptation of the software is necessary. Moreover, storing electronic coins on the mobile device is problematic. One option is to run the full-fledged payment system on a home PC and download electronic coins when needed.

Hardware electronic coins: There are various e-payment systems that implement an e-purse, that is electronic cash on a smartcard, for example GeldKarte and Mondex. In both cases, electronic money is stored on the card and can be transferred directly from the customer to the merchant. Shadow accounts are held at the bank to log transactions. Currently, these payment schemes are being adapted for m-commerce, where the GeldKarte and Mondex cards can be used in dual-slot mobile phones. One smartcard-based system for m-commerce is already in operation; the system by BarclayCard and Cellnet uses a dual-slot Motorola mobile phone for payments of up to £50.

Background account: Depending on the type of trusted third party, there are various different approaches to a background account solution.

If the background account is held at a network operator, the charged amount is transferred to the existing billing solution and included in the customer bill. Customers pay their bills using traditional systems, such as direct debit, cheque or cash. Examples in this category are the M-Pay

*This description is based on the eCash electronic payment system.

Bill service from Vodafone and Mobilepay by Sonera. These systems work for micro-payments only; accepting higher payments would imply that mobile operators become subject to a host of banking regulations.

If the background account is held at a credit card institution, the payment mechanism is all about secure transmission of credit card data to the credit card company. The first possibility uses a dual-slot mobile phone, whereas the second possibility employs a dual-chip mobile phone. The dual-slot solution (for example, ItiAchat by Mastercard, Oberthur smartcards, France Télécom, Europay, and Motorola) has the advantage that it works with 'normal' credit cards that can also be used in traditional POS terminals. On the other hand, dual-slot mobile phones are quite bulky. This solution is favoured by credit card companies and banks, since it allows them to stay in control of the payment functionality. The dual-chip solution (for example, EMPS—Electronic Mobile Payment System—by MeritaNordbanken, Nokia and Visa) has the advantage that handsets can be kept small. This solution is favoured by mobile operators, because they are in control of the dual chips.

Finally, if the background account is held at a bank, the existing banking infrastructure and technology can be reused. Examples are Paybox and MobiPay by BBVA and Telefónica. Typically, the merchant receives the customer's mobile phone number (or a pseudonym) and passes it on to the payment server, together with the payment details. The customer authorises the payment by providing a PIN.

The interested reader is referred to the ePSO (electronic-Payment Systems Observatory) project for a comprehensive overview of payment systems (including references to their origin) and a selection of background papers[24]. Current approaches to m-payment are discussed in Reference 2.

In References 26 and 27 we have studied the use of software electronic coins and hardware electronic coins for a specific application in m-commerce, namely for *ad hoc* payment of the access of a mobile device to an access network with which no former relationship exists (i.e. the access network is not the home network and does not have a roaming agreement with the home network). Hence, the service for which the user of the mobile device pays is the current use of this device.

Standardisation and forums

One important aspect of m-commerce is standardisation. Due to the heterogeneity of technologies for mobile devices, and the need for transmission and payment over the air interface, it is essential to find common approaches, both at a national and an international level. The following list summarises standardisation bodies and forums dealing with issues relating to m-commerce:

- PayCircle® (http://www.paycircle.org) is a vendor-independent non-profit organisation. Its main focus is to accelerate the use of payment technology and develop or adopt open payment APIs (uniform Application Programming Interfaces) based on XML, SOAP, Java and other Internet languages.
- MoSign (http://www.mosign.de): Banks, technology partners and end-device manufacturers have joined forces in the MoSign (Mobile Signature) project to create a platform for secure, legally binding mobile transactions based on existing standards.
- Mobile Payment Forum (http://www.mobilepaymentforum.org) is a global, cross-industry organisation dedicated to developing a framework for standardised, secure, and authenticated mobile commerce using payment card accounts.
- mSign (www.msign.org): The Mobile Electronic Signature Consortium is an association of companies and organisations from the mobile phone and Internet sectors. The objective is to establish and develop a secure cross-application infrastructure for the deployment of mobile digital signatures.
- mwif (http://www.mwif.org): The Mobile Wireless Internet Forum (MWIF) is an international non-profit industry association. Its mission is to drive acceptance and adoption of a single open mobile wireless and Internet architecture that is independent of the access technology.
- Radicchio (http://www.radicchio.org): As a non-profit organisation, Radicchio brings together market leaders to establish a common foundation for secure m-commerce by reaching a consensus on important interoperability issues.
- Encorus (http://www.encorus.com): Encorus Technologies is focused on building a flexible and open infrastructure and efficient payment processing services to drive the acceptance and usage of mobile payments worldwide.
- Mobile electronic Transactions MeT (http://www.mobiletransaction.org): MeT Limited is a company founded to establish a framework for secure mobile transactions, ensuring that interoperable mobile transaction solutions are developed around the world and enabling consumers to access goods and services seamlessly wherever they may be with consistent user experience, that means independent of device, service and network.

5 Conclusions

There will be no m-commerce without security of the underlying technologies. In this paper we have discussed security issues relating to network and service technologies (see Section 3) and m-payment (see Section 4). Regarding m-payment, some systems are under development or already operational. One of the main future challenges will be to unify payment solutions and provide the highest possible level of security.

Acknowledgment

The work described in this paper has been supported by the European Commission through the IST Programme under Contract IST-2000-25350.

References

1 WEITZEL, T., and KÖNIG, W.: 'Vom E- zum M-Payment' (in German). See http://much-magic.wiwi.uni-frankfurt.de/profs/mobile/infos.htm.
2 Datamonitor: 'Mobile payment systems'. Whitepaper, November 2000. See http://www.sun.co.uk/wireless/resources/pdf/datamonitor.pdf.
3 BBC News: 'Tough penalties for mobile phone theft', 3rd May 2002. See http://news.bbc.co.uk/hi/english/uk/newsid_1966000/1966247.stm
4 'Digital cellular telecommunication system (Phase 2+); Security aspects'. GSM 02.09 version 7.0.1 Release 1998
5 WALKER, M., and WRIGHT, T.: 'Security', *in* HILLEBRAND, F. (Ed.): 'GSM and UMTS: the creation of global mobile communication' (John Wiley, 2001)
6 '3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3G Security; Security Architecture'. 3GPP TS 33.102 3.9.0 Release 1999
7 BOMAN, K., HORN, G., HOWARD, P., and NIEMI, V.: 'UMTS security', *Electron. Commun. Eng. J.*, October 2002, **14,** (5), pp.191–204
8 BORISOV, N., GOLDBERG, I., and WAGNER, D.: 'Intercepting mobile communications: the insecurity of 802.11'. Proc. MOBICOM 2001, see http://citeseer.nj.nec.com/article/borisov01intercepting.html.
9 'Port-based network access control'. IEEE Standard 802.1X-2001
10 BLUNK, L., and VOLLBRECHT, J.: 'PPP Extensible Authentication Protocol (EAP)'. RFC 2284
11 KENT, S., and ATKINSON, R.: 'Security architecture for the Internet Protocol'. RFC 2401
12 Bluetooth Special Interest Group, see http://www.bluetooth.com
13 LÖHLEIN, B.: 'Bluetooth—technology, security and weaknesses'. 6th World Multiconference on Systemics, Cybernetics, and Informatics, Volume IV: Mobile/Wireless Computing and Communication Systems I, Orlando, Florida, USA, July 2002
14 DANKERS, J., GAREFALAKIS, T., SCHAFFELHOFER, R., and WRIGHT, T.: 'Public key infrastructure in mobile systems', *Electron. Commun. Eng. J.*, October 2002, **14,** (5), pp.180–190
15 DIERKS, T., and ALLEN, C.: 'The TLS protocol, Version 1.0'. RFC 2246
16 'Wireless transport layer security, Version 06-Apr-2001'. WAP Forum, see http://www1.wapforum.org/tech/documents/WAP-261-WTLS-20010406-a.pdf
17 'Wireless Application Protocol, WAP 2.0, Technical White Paper'. WAP Forum, see http://www.wapforum.org/what/WAPWhite_Paper1.pdf
18 'Digital cellular telecommunications system (Phase 2+); Customized Applications for Mobile Network Enhanced Logic (CAMEL) Phase 2; Stage 2'. 3GPP TS 03.78 Version 7.7.0 Release 1998
19 'Digital cellular telecommunications system (Phase 2+); Customized Applications for Mobile Network Enhanced Logic (CAMEL) Phase 3 - Stage 2'. 3GPP TS 23.078 version 4.4.0 Release 4
20 'Open Service Access; Application Programming Interface, Part 1: Overview'. ETSI ES 201 915-1 V1.1.1 (2001-12), see http://www.parlay.org/specs/index.asp
21 'Open Service Access; Application Programming Interface, Part 3: Framework'. ETSI ES 201 915-3 V1.1.1 (2001-12), see http://www.parlay.org/specs/index.asp
22 'Digital cellular telecommunication system (Phase 2+); Security Mechanisms for the SIM application toolkit'. GSM 03.48 version 8.3.0 release 1999
23 O'MAHONEY, D., PIERCE, M., and TEWARI, H.: 'Electronic payment systems for e-commerce'. Artech House Computer Security Series, 2nd edn.
24 ePayment Systems Observatory, Database on E-payment Systems, see http://epso.jrc.es/
25 WEBER, R.: 'Chablis—market analysis of digital payment systems'. Technical Report TUM-I9819, TU Munich, 1998, see http://chablis.informatik.tu-muenchen.de/Mstudy/
26 KNOSPE, H., and SCHWIDERSKI-GROSCHE, S.: 'Future mobile networks: ad-hoc access based on online payment with smartcards'. 13th IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications, PIMRC 2002, September 2002, to be published
27 KNOSPE, H., and SCHWIDERSKI-GROSCHE, S.: 'Online payment for access to heterogeneous mobile networks'. IST Mobile & Wireless Telecommunications Summit 2002, Thessaloniki, Greece, June 2002

Scarlet Schwiderski-Grosche gained a Diplom-Informatikerin degree in Computer Science at the Technical University of Braunschweig, Germany in 1992. She was awarded a PhD degree in distributed systems technology (on composite event detection in distributed systems) by Cambridge University in 1996. After a one-year post-doctoral research position in Cambridge, Scarlet worked as a post-doctoral researcher in Darmstadt (Germany) at the GMD—German National Research Centre for Information Technology (now part of Fraunhofer)—on biometrics and wireless communication protocols. In August 2001 she joined the Information Security Group at Royal Holloway to work on the EU SHAMAN project (see http://www.ist-shaman.org/). Her special interests are security in mobile telecommunications systems, e-payment in m-commerce and e-commerce systems, and biometrics.

Address: Information Security Group, Royal Holloway University of London, Egham, Surrey, TW20 0EX, UK
E-mail: scarlet.schwiderski-grosche@rhul.ac.uk

Heiko Knospe received a Diplom degree in Mathematics from the University of Köln, Germany, in 1992 and a PhD degree in Mathematics from the University of Münster, Germany, in 1997. During his stay in Münster from 1995 to 1997 he was a research associate and member of the number theory group. Since 1997 he has worked for Deutsche Telekom and in particular its affiliates T-Mobile and T-Systems, the focus of his work being on billing systems, mobile networks and information security. His research interests include accounting protocols, security of mobile technologies and applications of elliptic curves.

Address: T-Systems Nova GmbH, Technologiezentrum ES21d, D-64307 Darmstadt, Germany
E-mail: heiko.knospe@t-systems.com

First received 28th June and in final form 16th August 2002