Additional Issues in Mobile Security

Andrej Šoštarič

Povzetek — V sestavku opisujemo temeljne varnostne modele v mobilnih komunikacijah. Predstavljamo njihove lastnosti in slabosti. Poleg tega opozarjamo tudi na druge varnostne vidike in pomisleke pri mobilnem komuniciranju in ne le izključno na varnost komunikacijskih kanalov. Pri tem se dotikamo področij, kot sta enoznačni avtentikacija in avtorizacija mobilnih uporabnikov, ki pri uporabi različnih komunikacijskih poti, kot so WAP, WEB, SMS in govor, predstavljata izziv podobnih razsežnosti kot npr. kodiranje oz. šifriranje govora in sporočil.

Ključne besede — varnostni modeli, GSM, GPRS, WAP, PKI, avtentikacija, avtorizacija, večdostopni portal

Abstract — In this article fundamental security models together with their properties and weaknesses are described. Additionally we do not address only security of communication channels, but we also consider some other important issues, such as authentication and authorization. These represent the same if not even bigger challenge when customers use different communication means, such as WAP, WEB, SMS and /or voice.

Keywords — security models, GSM, GPRS, WAP, PKI, authentication, authorization, multi-access portal

I. INTRODUCTION

Generally, it is true – being mobile in the heads of users and service providers is concerned as not too reliable and trusted way of accessing some sensitive information. On the other hand, one of the most important things that underlies all the rosy predictions about mobile commerce and ubiquitous wireless data networks is the assumption that data can be stored securely on devices, encrypted successfully over the air and handled equally securely on the server. Unfortunately, rarely a month goes by when a researcher doesn't find a new security hole in a wireless technology being touted as the next great wireless economic engine. Because of the processing, memory and battery-life requirements of mobile devices, traditional encryption techniques (which sometimes rely on powerful processors and large amounts of memory) also fall flat on wireless platforms, creating the need for wireless-specific security technologies.

In the article we would like to show to what extent being mobile actually means being secure. First, we will give an overview of the mobile world and point out those parts of the whole system, which might be especially dependent on security standards and measures used. We will address technologies like GSM, GPRS, and WAP.

Next, we will show to what extent a certain technology can be used in order to meet different security demands, including authentication and authorization, and at the end we will try to present a prediction of future development in the area of mobile technologies and corresponding security.

II. GSM SECURITY MODEL

The GSM Security Model is based on a shared secret between the subscriber's home, network's HLR and the subscriber's SIM. The shared secret, called Ki, is a 128-bit key used to generate a 32-bit signed response (SRES) to a random challenge (RAND), made by the MSC, and a 64-bit session key (Kc), used for the encryption of the over-the-air channel.

There are several algorithms involved – A3 for calculating the SRES, A8 for calculating the session key Kc, and A5, which is used for generation of keystream. Detailed procedure is explained in [1] and presented in Figure 1.

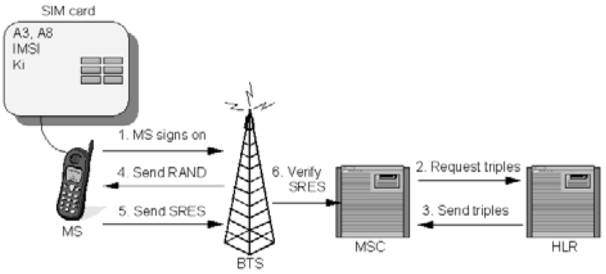

Figure 1: Mobile station authentication

A call can be decrypted if the attacker knows the Kc (and some other implicit information, such as frame number). The same Kc is used as long as the MSC does not authenticate the MS again, in which case a new Kc is generated. In practice, the same Kc may be in use for days. The MS authentication is an optional procedure in the beginning of a call, but it is usually not performed. Thus, the Kc is not changed during calls. Only the over-the-air traffic is encrypted in a GSM network. Once the frames have been received by the BTS, it decrypts them and sends them in plaintext to the operator's backbone network.

III. POSSIBLE INTERCEPTION ATTACKS IN GSM NETWORKS

The interesting question about the GSM security model is whether a call can be eavesdropped, now that at least one of the algorithms it depends on has been proven faulty.

Scientists around the world seem to be unanimous that the over-the-air interception and real time decoding of a call are still impossible regardless of the reduced key space. But there seem to be other ways of attacking the system that are feasible and seem to be very real threats. There are also many attacks that are realistic, yet do not abuse any of the faults in the security algorithms. Let us list only few of them:

- *Brute-force attack against A5* – a real-time brute-force attack against the GSM security system is not feasible, as stated above since the time complexity is far too big, but with the distributed computer systems we can drastically reduce the time required.
- *Divide-and-conquer attack against A5* – a divide-and-conquer attack is based on a known-plain-text attack and can dramatically reduce the complexity (up to 2^9 – 2^14) [1].
- *Accessing the operator's signaling network* – the airwaves between the MS and the BTS are not the only vulnerable point in the GSM system. The transmissions are encrypted only between the MS and the BTS. After the BTS, the traffic is transmitted in plain text within the operators network. If the attacker can access the operator's signaling network, he will be able to listen to everything that is transmitted, including the actual phone call as well as the RAND, SRES and Kc. The SS7 signaling network used in the operator's GSM network is completely insecure if the attacker gains direct access to it.
- *Retrieving the key from the SIM* – the security of the whole GSM security model is based on the secret Ki. If this key is compromised the whole account is compromised. Once the attacker is able to retrieve the Ki, he can not only listen to the subscribers calls, but also place calls billed to the original subscriber's account, because he can now impersonate the legitimate subscriber.

As we can see, the GSM security model is devided into many levels and is thus vulnerable to numerous attacks targeted to different parts of an operator's network. Assuming that the security algorithms were not broken, the GSM architecture would still be vulnerable to attacks targeting the operator's backbone network or HLR and to various social engineering scenarios in which the attacker bribes an employee of the operator, etc. Further more, the secretly designed security algorithms incorporated into the GSM system have been proven faulty.

All this means that if somebody wants to intercept a GSM call, he can do so. It cannot be assumed that the GSM security model provides any kind of security against a dedicated attacker. The required resources depend on the attack chosen. Thus, one should not rely solely on the GSM security model when transferring confidential data over the GSM network.

However, the reality is that although the GSM standard was supposed to prevent the problems of phone fraud and call interception found in the analog mobile phone systems by using strong crypto for MS authentication and over-the-air traffic encryption, these promises were not kept. The current GSM standard and implementation enable both, subscriber identity cloning and call interception. Although the implementation of cloning or call interception is a little bit more difficult, due to the digital technology that is used, compared to the analog counterparts, the threat is still very real, especially in cases where the transmitted data is valuable.

IV. GPRS SECURITY VS. GSM SECURITY

In the GPRS system, the frames are transmitted as cipher text from the MS to the SGSN (Serving GPRS Support Node). This is done because the GPRS system uses multiple time slots in parallel in order to achieve a greater transmission rate. To one GPRS phone multiple time slots can be allocated by the network, thus increasing the transmission rate of that MS. The frames can be sent in 'parallel' time slots to the same BTS or to two different BTSs if the MS is handed over from one BTS to another.

To a BTS the use of one time slot is seen as a separate call. Thus, the BTS is unable to put the frames from different timeslots together. This means that there has to be a network component that is able to receive the frames from one MS, defragment them and send them onwards to the actual destination. The BTSs are also unable to decrypt the frames, because consecutive frames on one channel don't have consecutive frame numbers (Figure 2). To simplify the implementation, the frames are decrypted at the SGSN where all of the frames end up and it is thus easy to keep track of frame numbers. The solution is based on the ease of implementation and has not been implemented in order to increase system security. As a side effect, the GPRS system effectively prevents eavesdropping on the backbone between the BTS and SGSN, because the frames are still encrypted at this point. In GPRS, the triples from the HLR are transmitted to the SGSN and not to the MSC. Thus, security of GPRS depends largely on the placement and security of the SGSNs and is in general far higher than in the GSM world.

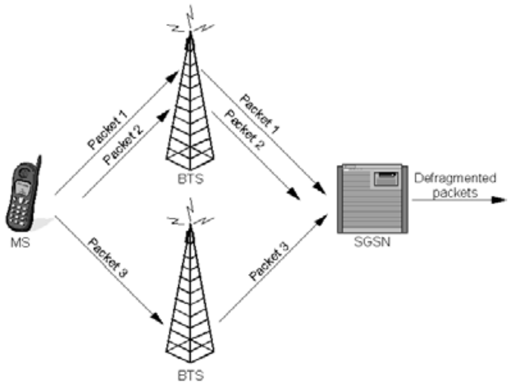

Figure 2: GPRS architecture

The GPRS system uses a new A5 implementation as well, which is not known publicly. This and the fact that the frames are not decrypted at the BTS but at the SGSN rules out a couple of attacks. First, it is very hard to attack the A5 implementation when it is not known. Secondly, the Kc is not transmitted to the BTSs and the transmission channel between the BTS and the SGSN is encrypted making it thus useless to monitor the backbone between the BTS and the SGSN. This does not mean that the GPRS security model would somehow be more secure than the GSM-only security model. It means that identical attacks that work with a GSM-only network do not work with GPRS. As soon as the A5 implementation used in GPRS leaks out, the GPRS security model is vulnerable to new attacks. And the implementation will leak out eventually or the design is successfully reverse-engineered. As stated above, the security of a crypto system should be based solely on the key. However the majority of the attacks against the GSM-only system are applicable to GPRS as well. E.g. the SIM-cloning attack. Additionally, the GPRS model introduces another security threat through the use of SGSNs, which know the triples from the HLR. This means that the security of the GPRS network depends largely on the positions of the SGSNs in the network architecture and the security of the SGSNs. If the SGSNs are vulnerable to an attack, then the triples are vulnerable as well.

V. WAP Security Model

The Wireless Application Protocol (WAP) is the most popular wireless data technology in use today. It has its own security mechanism, named Wireless Transport Layer Security (WTLS). WTLS is a wireless relative of the more common SSL mechanism used by all major Web browsers. WTLS resembles SSL in that that they both rely on certificates on the client and server to verify the identity of the participants involved. While SSL implementations generally rely on RSA encryption, WTLS supports RSA, Diffie-Hellman, and Elliptic Curve encryption, but in practice most vendors are focusing support on RSA because of its widespread use.

WTLS is all about adding security to low CPU-powered wireless devices by making the cryptography efficient. Because PDA and cell phone CPUs are typically slow, using SSL end to end can take anywhere from 30 seconds to several minutes, depending on the key size used to negotiate an SSL connection. WTLS can use familiar public key exchange algorithms, such as RSA or Diffie-Hellman, but these algorithms are resource-intensive and, therefore, slow. Elliptic Curve (EC) cryptography promises to require far fewer resources and should find wide deployment for CPU-starved PDAs and cell phones.

WTLS's key exchange protocol is also uniquely suited for wireless applications. Vendors can implement any of three classes of authentication types:

- *Anonymous authentication (class 1)* has limited use -- mainly for testing purposes -- because end users have no way of determining to whom they are talking. The client forms an encrypted connection with an unknown server.
- *Server authentication (class 2)* will probably be the most common model used. As with SSL, once clients are assured they are talking securely to the correct server, they can authenticate using alternative means such as user name/password. Bear in mind that WTLS certificates are not the same as X.509 certificates, and they can't be used interchangeably.
- *Server- and client-authentication (class 3)* is possibly the strongest class, as the server and the client authenticate each other's WTLS certificate. Client certificates required for Class 3 authentication pose special management problems. Not only must the key pairs be generated on the mobile device (or generated in bulk and securely loaded onto the mobile devices), but the client certificate has to be safeguarded and managed until the certificate expires. Client certificates need not be retained on the handheld device. Preferably, during negotiation, the client may refer the WTLS gateway to a directory to retrieve the client certificate from a directory. That saves the bandwidth needed to send the client certificate over the air and may improve negotiation performance; however, the WAP gateway needs to trust the directory the client refers to in order to have any assurance of authentication. The directory that holds user certificates must also be available at all times, or it won't be able to retrieve the certificate when requested. The key pair associated with the client certificate resides only on the client.

WTLS also doesn't provide for end-to-end security due to WAP's current architecture and limitations of server-side SSL. While WAP clients can securely exchange data with a WAP gateway using WTLS, the gateway must open an SSL session with a back-end server in order to complete the transaction. It is at the WTLS gateway where the potential problem exists. Between the time the data is decrypted and "decapsulated" from WTLS and WAP and re-encapsulated and re-encrypted in SSL, the protected data is exposed – albeit for only about a few hundred milliseconds. For most applications and users, this shouldn't be a big deal. If someone breaks into your WAP/WTLS gateway, you have bigger problems to deal with.

Due to this requirement, WAP 1.x suffered a serious security setback after it was revealed that data could be accessed, unencrypted, for a brief moment at the point where the WAP gateway passed data off to the back-end server. The WAP Forum has addressed this issue in WAP 2.0, offering end-to-end security for the first time to WAP developers. One of the concerns with cryptography regards export of certain key lengths to other countries. The WAP Forum is sensitive to this issue, too, and the WTLS draft supports various key lengths used with the bulk encryption algorithms, so that the security parameters can be negotiated according to geographic need rather than server support.

At the moment there are actually three options that are available for end-to-end WTLS security. The first is WTLS tunneling, which tunnels WTLS traffic through a service provider's network to a remote WAP gateway. WTLS proxy, meanwhile, conveys WTLS connections through the carrier's WAP gateway. Neither solution is widely deployed and each will require partnerships with carriers and phone manufacturers to implement. The third option, wireless PKI, is the most promising of all, therefore, it attracts special attention.

VI. WIRELESS PKI

It becomes clear that the success of mobile services, such as mobile commerce, banking, payment, etc., depends on security infrastructure that does not rely on cryptography that has proven to be inherently weak and that provides end-to-end security. Only Wireless Public Key Infrastructure (PKI) meets these requirements. This is why PKI will emerge as the de facto mobile commerce security standard [4]. It is also why organizations such as Baltimore Technologies have been working with WAP Forum to define a definitive set of WAP industry standards and protocols to ensure a trusted environment in which mobile business can flourish.

Wireless PKI provides such a secure and trusted trading environment by meeting the four key requirements of electronic security using cryptography, digital signatures and digital certificates. These four key requirements are [4]:

- *Confidentiality* – assurance that nobody can eavesdrop.
- *Authentication* – assurance that the parties you are doing business with are who they claim to be.
- *Integrity* – assurance that information you send or receive is not tampered with on its journey.
- *Non-repudiation* – assurance that agreements are legally binding.

There are many types of mobile transactions, some already taking place, some predicted. Wireless PKI-based technology enables secure mobile business across all of these transactions and across all wireless platforms. Good news is, wireless PKI is not about building a completely new security infrastructure. It is about extending the wired trust model to the wireless applications. PKI-centric solution is supposed to be the most important enabler of the end-to-end security.

Compatibility with today's mobile access devices as well as with evolving technology, standards and protocols and your existing security infrastructure, or any future infrastructure you may adopt, is equally important. There are many approaches to PKI implementation and just as many incompatible solutions available from different vendors, but many organizations have already signed an interoperability pact for their PKI products.

VII. CONSIDERATIONS WHILE INTRODUCING NEW MOBILE APPLICATIONS

If we concentrate on WAP Internet-access that is becoming more and more used channel of accessing data today, we see that its architecture is three-tiered (comprising a mobile access device, a WAP Gateway and a Web Server), whereas PC access to the Web is two-tiered (comprising a PC and the Web server). An extra layer is opened to additional security breaches. But not everything is as bad as it looks, of course, since there are many wireless services offered and many more are still to come. Different technologies described above can already today be successfully applied and used already today for different purposes and services. We just must have in mind the technology based security constraints and be aware of trade-off between the investment into the secure infrastructure and possible security flaws.

Currently, most applications and services, even if they are being developed by independent service providers, are introduced by mobile operators, since they own the wireless infrastructure. On the other hand, in order to enable mobile channels, such as SMS, WAP and voice, mobile operators offer perhaps the most advanced types of multi-access portals. A structure of such portal is anything but simple, therefore special middleware platforms are being developed, which unify access to the mobile operator's commonly used infrastructure, authentication, authorization, accounting, and enable different access methods and open infrastructure to the external world. Structure of a multi-access portal is shown in Figure 3.

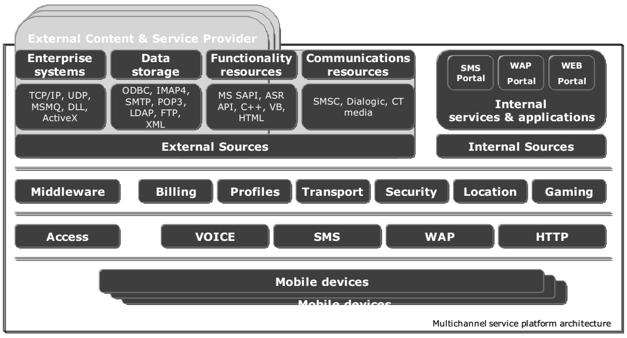

Figure 3: Multi-access/channel service platform architecture

In the same figure, security is only one block in the middleware layer, but we can say that it is far the most important enabler for success of the next generation services and networks. Here we do not address only data encryption but also secure way of contacting the mobile operator's infrastructure. If mobile operators want to attract new services from the external world, they must open connections to their internal user databases as well to some other sensitive equipment and communication resources, such as SMS Centers, which mediate the flow of short messages, prepaid and postpaid billing systems, location-based servers, etc.

To simplify connection to the internal mobile operator's resources several initiatives took place in the past. The most promising one is a Parlay initiative [5], which is actually a consortium of several important infrastructure manufacturers and operators in order to provide a higher-level and secure way of addressing infrastructure mentioned above.

To illustrate the complexity of introduction of the simplest premium-rate service (game, application, etc.) to the SMS and WAP channel, let's point out some of the problems which we should be aware of:

- *Authentication* – while for the SMS-based channel the service provider knows exactly from which phone numbers the request are coming in, many WAP gateways are not capable of providing such information. In the latter case, we must introduce another way of user authentication (i.e. username/password-based). If so, we must be aware, that the same users should use the same username/password combination whenever they address other mobile operator's portals (Voice or WEB portal). They might use some other authentication methods – digital certificate/ password combination, PIN numbers, etc. – as well. Therefore, every mobile operator needs a reliable, secure, and simple authentication system.
- *Billing* – the capability to charge for the services is a unique value that mobile operators have in the new e-conomy. At the beginning it was clear why mobile operators were trying to introduce all the information services, games, etc. – all by themselves. Today they realized that if they want to be attractive in the battle for customers they will have to let external service and content providers to offer new services in order to keep the pace with or be in front of their competition. The rate of premium-rate services is becoming higher and higher. Mobile operators have to offer a way of connecting to their internal prepaid and postpaid billing systems. The connection types may vary a lot from operator to operator. The ratio between prepaid and postpaid users is also country dependent. Well-developed countries have more postpaid users, while some other countries from the Eastern Europe have almost only prepaid users. Every operator has to provide a set of secure connectors into their billing system. The majority of problems lies on the prepaid side, since the billing should be done in real-time. On the other hand, most of the prepaid users are anonymous, so user authentication is almost impossible thus meaning that not all services should be introduced in such networks.
- *User databases* – in some cases (external) service providers have capabilities to retrieve and store some service-specific information in the operator's databases (i.e. games scores, nicknames, number of games played, etc.). There's no need to say, how sensitive this information is and how well protected these data are. Whenever mobile operators talk to the external service providers they have to authenticate these providers as well. This is done in most cases by means of digital certificates. It is not unusual for mobile operators to act as certificate authorities and to issue digital certificates by themselves.
- *Communication infrastructure* – to route the short messages to the external service provider connection to the communication infrastructure is needed. This is done usually by using standard communication protocols which are supposed to be secure enough (HTTP, TCP/IP, SMPP, etc.) but some threats still remain. The highest risk today is a spam traffic. Even though someone unauthorized may not break into the

communication system of the mobile operator he/she may still cause a lot of damage by sending a lot of spam messages. Therefore, special interest should be paid to the placement of anti-spam filters.

VIII. What applications to offer today?

Taking into account all the security issues what applications can still be offered to the users using current mobile technologies?

Most information services do not need any kind of special security measures. As long as the services offer only general information, such as news, weather, horoscop, games, etc. they might be exposed to the vast number of the simplest mobile phones that have the possibility of sending short messages. Some services are being charged. Most often, the payment amounts are small; therefore authentication is based on the IMSI (phone number) found on the SIM.

Next step represent those SMS services that offer more sensitive information. If we want to check the bank account status or even use mobile phone as a payment tool we need reliable authentication mechanism and end-to-end security as well. For this purpose more advanced SIM/WIM cards, known as smartcards, should be used. These cards contain besides memory also CPU. Memory is not used only for storing phone numbers but also for storing encryption keys. Accordingly, CPU is used for message cyphering. Currently, most cards come with 32 kb or 64 kb of available memory. Both, symmetric and asymmetric keys are being used. The latter require more processing power and only in the last few months PKI-based encryption and authentication became possible. Unfortunately, different card suppliers have different understanding of standardization guidelines. Even Java-based smartcard implementations lack software portability.

IX. Conclusion

Despite some proponents' claims to the contrary, wireless data technologies still possess a level of insecurity, particularly if custom security measures (such as encryption) are not put in place by the enterprise or application developer. WAP 2.0 hopes to solve WAP's primary security problems, but the all-important vendor implementations of the standard will decide whether the public accepts the level of security offered. In the mean time, the next generation mobile phones and PDAs will leverage from technologies like HDML, XHTML, UMTS, Internet V6 to name just a few of them. Finally, mobile world will most probably become a logical extension to the wired Internet. Mobile devices will become more powerful and capable of performing stronger encryption methods. But the risk of eavesdropping will remain.

It is also very important to understand the complexity of the mobile operators' infrastructure. Eavesdropping is not the only problem they face. In order to survive on the market they have to open their systems to the external service providers. Within this process new and different security holes arise. Hand in hand new middleware technologies try to keep the pace with the demands of a modern mobile world.

As you can see from the issues raised in this brief discussion, a long road lies ahead for mobile security vendors seeking to gain the public's trust. Only when these products and technologies are proven to be secure from end to end will mobile commerce begin to take off.

References

[1] Lauri Pesonen, *GSM Interception*, Internet article, Department of Computer Science and Engineering, Helsinki University of Technology, November 1999.

[2] Mike Fratto, *Tutorial on Wireless Security*, Internet article, Network Computing, January 2001.

[3] Bryan Morgan, *Thinking about wireless security*, Internet article, Radichio, November 2001.

[4] Guy Singh, *Securing the mobile E-Conomy*, Internet article, Baltimore Technologies, August 2000.

[5] The Parlay Group, http://www.parlay.org, 2001

Andrej Šoštarič (andrej.sostaric@hermes.si) reached his PhD degree at Ecole Central de Nantes, France and at University of Maribor. After eight years of being assistant at the Faculty of Electrical Engineering and Computer Science at University of Maribor, where he still teaches as a lecturer, he joined the HERMES SoftLab company, where he's been working for last four years. Currenlty he has a position of Senior Project Manager and Technical Sales Consultant. His area of work covers mostly mobile telecommunications.