Using GSM/UMTS for Single Sign-On

Andreas Pashalidis* and Chris Mitchell
Information Security Group
Royal Holloway University of London
E-mail: {A.Pashalidis,C.Mitchell}@rhul.ac.uk

Abstract

At present, network users have to remember a username and a corresponding password for every service with which they are registered. Single Sign-On (SSO) has been proposed as a solution to the usability, security and management implications of this situation. Under SSO, users authenticate themselves only once to an entity termed the 'Authentication Service Provider' (ASP) and subsequently use disparate Service Providers (SPs) without re-authenticating. The information about the user's authentication status is handled between the ASP and the desired SP in a manner transparent to the user. In this paper we propose a SSO protocol where a GSM or UMTS operator plays the role of the ASP and by which its subscribers can be authenticated to SPs without any user interaction and in a way that preserves the user's privacy and mobility. The protocol only requires minimal changes to the deployed GSM infrastructure.

Keywords: single sign-on, authentication, GSM, UMTS

1 Introduction

Network users have to remember a username and password for every Service Provider[1] (SP) they are registered with. If they could remember different, ideally 'secure' [4] passwords (and corresponding user names) for every such SP, they might have done everything they could from their perspective with respect to security. Unfortunately, this has proven to be very difficult, and thus users either record their passwords (on paper or in a computer file), or, more commonly, use the same password with every SP. This has serious security implications, most obviously that an SP can impersonate a user to all other SPs with whom the same password has been used. Thus, say, the administrator of a (personalisable) news website might be able to access a portal visitor's bank account, credit card numbers, emails or health records.

Single sign-on (SSO) is a technique where users authenticate themselves only once to a trusted *Authentication Service Provider* (ASP), and are automatically logged into the SPs they subsequently use, without necessarily having to re-authenticate each time. Under SSO, SPs need some form of notification from the ASP that indicates the user's authentication status. These notifications are termed *authentication assertions*. SPs assess the provided authentication assertions and decide whether or not to grant access to a protected resource to the specified user. SSO both increases the usability of the network and eliminates the security implications mentioned above (but introduces its own).

The Liberty Alliance[2], a consortium of over 160 companies, has developed a set of open specifications that, among other things, provide web-based SSO. In the Liberty architecture, a trusted third party acts as the ASP which authenticates users and subsequently provides authentication assertions to relying SPs. Relationships between ASP

*The author is sponsored by the State's Scholarship Foundation of Greece.

[1] In the context of this paper a service provider is any entity that provides some kind of service or content to a user. Examples of SPs include messenger services, FTP sites, web sites and streaming media providers.

[2] www.projectliberty.org

and SPs are based on contractual agreements outside the scope of the specifications.

The actual authentication method used by the ASP is not specified by Liberty. Authentication assertions do, however, include descriptions of the initial user identification procedure at the ASP, physical, operational and technical protection procedures, and the authentication method used, [10].

In this paper we propose a SSO protocol where a Global System for Mobile communications (GSM) operator acts as the ASP. The associated authentication method is similar to authentication of subscribers in a GSM network, in that it uses a secret shared between network operator and subscriber. The proposed scheme requires only minimal changes to the deployed GSM infrastructure and can potentially be accommodated by the Liberty specifications.

The rest of this paper is organised as follows. Section 2 reviews the GSM data confidentiality service. Section 3 describes the proposed protocol for user authentication and SSO using GSM, while section 4 analyses the associated security threats. Section 5 discusses the advantages and disadvantages of the protocol. In section 6 the protocol is extended to cover the Universal Mobile Telephone System (UMTS). Finally, sections 7 and 8 give an overview of related work and conclusions.

2 The GSM data confidentiality service

Subscriber data confidentiality for the GSM air interface is based on a long-term secret key K_i, shared between the Subscriber's Identity Module (SIM) and the Authentication Center (AuC) of the home network operator. A secret session key is generated by both parties during subscriber authentication, as described below. (Only the steps relevant to this paper are shown; comprehensive descriptions of GSM air interface security can be found in [1, 5, 6, 17, 18]).

1. The home network generates a random challenge (RAND), and uses the shared secret key K_i to compute a secret session key K_c as follows:

   $$K_c = \mathrm{A8}_{K_i}(\mathrm{RAND})$$

   where A8 is a network-specific key derivation function. The values of RAND and K_c are then passed to the visited network.

2. The visited network sends the value RAND to the Mobile Equipment (ME).

3. The ME passes the RAND to the SIM which recomputes K_c using its stored value of K_i. The key K_c is then output by the SIM to the ME.

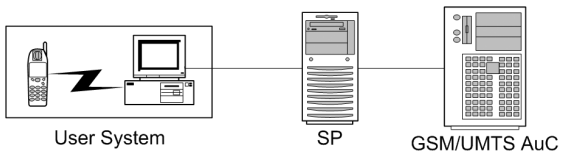

Figure 1: PC as network access device.

The secret session key K_c is used for data and signalling encryption in GSM. It is important to note that the shared secret K_i *never* leaves the subscriber's SIM card or the home network operator's AuC.

3 Using GSM for SSO

This section describes the proposed authentication method and the associated SSO protocol as well as the involved entities.

3.1 System entities

The main components are the User System, the SP and the GSM operator's AuC.

3.1.1 User System

The User System (US) consists of a network access device, a SIM card and a SIM card reader. The device might be a PC, with a GSM Mobile Equipment (ME), e.g. a GSM phone, as the SIM

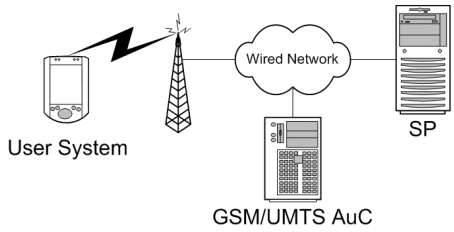

Figure 2: Combined mobile equipment and network access device

card reader. The PC and ME need to be interconnected, e.g. using a cable, infrared, Bluetooth (`www.bluetooth.com`) or a Wireless Local Area Network (WLAN) [7]. Regardless of the method used, it is assumed that the link is protected against eavesdropping. This configuration is shown in Figure 1. Alternatively, the access device and SIM card reader could be combined, e.g. in a Wireless Application Protocol (`www.wapforum.org`) enabled device — see Figure 2.

The US is the entity through which the end user authenticates to the AuC and subsequently achieves SSO at different SPs. It should be noted that no direct communication between US and AuC takes place in the proposed protocol; messages are 'routed' through the SP. However, the US and the AuC need to agree on the use of a Message Authentication Code (MAC) function [14].

3.1.2 Service Provider and Authentication Center

It is assumed that the relationship between SPs and GSM operators is regulated by contractual agreements which are beyond the scope of this paper. To avoid the need for large numbers of such agreements, the SP/GSM relationship could be established via one of a relatively small number of third parties providing a 'broker' service. Such a third party could be established specifically to support this SSO scheme. It is also assumed that the SPs and the AuC have the means to establish an authenticated and integrity protected communications channel (see section 4.5). This might involve routing all communications via the trusted third party that is brokering relationships between SPs and GSM operators.

In the scheme described here, the GSM operator's AuC acts as the ASP for a number of relying SPs (in Liberty terms, the GSM operator and associated SPs form a *circle of trust* [9]). As in any SSO scheme, the AuC needs to be trusted for the purposes of authentication by both the end-users and the SPs.

3.2 The authentication and SSO protocol

The proposed protocol starts whenever the SP wishes to authenticate the user, e.g. when the user requests a protected resource from the SP; it is also used whenever the SP decides that the user has to be re-authenticated. The protocol consists of a series of four messages that are exchanged as follows:

1. SP → US: RAND
2. US → SP: IMSI, MAC_{K_c}(SPID)
3. SP → AuC: IMSI, MAC_{K_c}(SPID), RAND
4. AuC → SP: Authentication Assertion or Failure Notification

In message 1, the SP sends a random challenge (RAND) to the US. The US forwards the RAND to the SIM which then computes a secret session key K_c as described in section 2. The key is returned to the US. The US must also extract the International Mobile Subscriber Identity (IMSI) from the SIM at some point — this can be done with a standard 'call' to the SIM. The IMSI uniquely identifies the home network and the user's subscription.

The US then computes, using key K_c, a MAC on the unique identifier (SPID) of the SP that wishes to authenticate the user. It should be noted that K_c, as in [8], is used here for MAC computation, whilst it was designed for data encryption. If this breach of the key separation principle is a concern then K_c could be passed through a one-way hash function (such as SHA-1 [14]) before being used for the MAC computation [8]. The US constructs message 2 which consists of the IMSI and the MAC and sends it to the SP.

The SP looks at the Mobile Country Code (MCC) and Mobile Network Code (MNC) fields in

the received IMSI to determine the address of the AuC of the user's GSM operator. In message 3 the SP simply forwards the IMSI and the MAC it received from the US to the AuC, and appends the RAND from message 1. In Liberty terms [13] message 3 corresponds to an *authentication request*.

The AuC finds the secret key K_i corresponding to the IMSI of message 3 and derives K_c using the given RAND. It then computes a MAC using K_c (or its hash value) on the SPID of the SP from which it received message 3. If the resulting value matches the MAC received in message 3 the user is deemed authenticated at the AuC. In that case, the AuC sends an authentication assertion to the SP in message 4. Otherwise, message 4 is a failure notification. In Liberty terms this message corresponds to an *authentication response*.

SPs differentiate between users based on their IMSIs. Thus, the protocol can also be used for initial user registration; the SP creates an account for a newly encountered IMSI. Whether or not individual SPs keep more information about users in their accounts and how this is mapped to their IMSIs is outside the scope of this paper.

In the US, the protocol might be implemented as a continuously running process (also known as a 'service' or 'dæmon') or as part of the client software that is used to access the service offered by the SP. In the SP, the protocol would have to be implemented by the software that offers the service (the 'server software'). A File Transfer Protocol (FTP) SP, for example, would have to support the protocol at the FTP server software level. The GSM operator would most likely implement the protocol as a process that runs continuously on a dedicated 'SSO server'.

4 Threat analysis

This section considers threats to the scheme and corresponding countermeasures. Each potential attack is considered separately.

4.1 Stolen SIM attack

An attacker with a stolen SIM is potentially able to make fraudulent phone calls at the owner's expense, at least until the SIM is reported stolen and blocked, assuming that the SIM is not PIN-protected. Typically, SIM cards are stolen with the ME, and the thief also thus gains access to any personal information in the SIM and ME. In the context of the scheme described above, a SIM thief can also impersonate the user to all SPs that support SSO using the proposed protocol. For low or medium value services (such as news portals or email) this risk might be acceptable, compared to the costs associated with SIM card/ME theft. However, for high value services (e.g. online banking, e-commerce or business email) the SSO mechanism may need to be combined with another authentication method, e.g. username and password. In such a case an attacker will need both the user SIM and password to impersonate the user to an SP. Also, once the user realises that the SIM has been stolen (typically soon after the incident) and reports this to the GSM operator, the SIM will be blacklisted and impersonation prevented.

4.2 SIM cloning attack

A SIM card can be cloned (or emulated through software) if the secret key K_i can be extracted. An attack exploiting a weakness in the COMP128v.1 algorithm, which was used by some GSM operators in early SIMs, enables key extraction; see, for example, [19]. Although this is not an attack against the protocol proposed in this paper, cloned SIMs enable the attacker to impersonate subscribers at the GSM network, and therefore at SPs. In any case, as mentioned in section 4.1, high value services should be protected by an additional authentication mechanism.

4.3 Compromise of privacy

Unlike the protocol described in [8], the scheme does not involve the user's Mobile Subscriber Integrated Service Digital Network number or any identifying information other than the IMSI. As the mapping between the IMSI and other user data is kept at the trusted AuC, SPs and eavesdroppers only learn the user's home GSM operator (this information is needed anyway — see section 3.2). Given that each GSM operator has many sub-

scribers, learning a user's operator is not likely to be a major breach of privacy.

If dishonest SPs collude they can correlate information about common users without their consent, using their IMSIs. To address this privacy threat, Liberty specifies the use of different, opaque user handles for each ASP/SP association [11]. However, complete prevention of the 'SP collusion attack' is hard as SPs can still correlate users based on other profile information they may maintain, e.g. names or phone numbers. As stated in the Liberty specifications [12, p.71], 'The only protection is for Principals [users] to be cautious when they choose service providers and understand their privacy policies'.

4.4 Reflection attack

An attacker could forward message 1 received from an SP as part of the authentication process to a victim user, while masquerading as the SP to the user (maybe by spoofing the SP's network address and/or interface). Forwarding the user's valid response (message 2) to the SP might result in successful impersonation.

It is thus important for the user to authenticate the origin of message 1. This can be achieved using a suitable challenge/response protocol or an SSL/TLS channel with server side certificates[3]. Of course, simply trusting the root certificates that come pre-configured in popular web browsers involves certain risks. Ideally, the user should explicitly enable every root key involved in verifying SP certificates.

It should be noted that the reflection attack is *not* prevented if launched by a dishonest SP. However, as explained in section 4.3, users should be cautious when they choose SPs.

4.5 Attacks on the SP/AuC Link

An attacker able to modify network traffic between an SP and an AuC will be able to defeat the system, since the AuC will not be able to determine which SP requested an assertion (the origin of message 3 could be altered) and, more importantly, relying SPs would not be able to have confidence in authentication assertions (message 4). Origin authentication and integrity protection for messages 3 and 4 is therefore a fundamental requirement. There exist well-established techniques that address this requirement, for example SSL/TLS with both server and client side certificates [16], IPsec tunnelling, or a Virtual Private Network [17].

[3]Since the user requests a protected resource, it is likely that a SSL/TLS connection will be required anyway.

4.6 Replay attack

A network eavesdropper could capture message 2 of a previous protocol run between a US and an SP. The attacker might later replay that message to try to impersonate the user to the SP. The attack will only succeed if the SP challenges the attacker with the same RAND (in message 1) as it used previously. It is therefore important for SPs to use numbers with good randomness properties, such that the probability of challenging a given user with the same number twice is negligible. An SP that does not follow this policy, however, only renders itself (and not other SPs) vulnerable to this kind of attack.

4.7 Attacks against the Authentication Center

As for any ASP in an SSO scheme, the AuC is a central point of failure and is therefore a component highly susceptible to service denial attacks. The AuC, at the same time, is the only entity trusted by both end-users and SPs. It is assumed that the GSM network operator will not abuse this trust and that the AuC will be well-protected against service denial and illegal access. This is likely to be true in any event, as AuC failure would also bring down the entire GSM network.

5 Advantages and disadvantages

This section briefly discusses the advantages and disadvantages of the proposed authentication method and SSO protocol.

Advantages resulting from the synergy of combining GSM with SSO include the following.

- No user interaction is required. This yields transparent user authentication at the SP. One envisioned scenario is that users, without taking their mobile phones out of their pockets, approach (public) network access devices and are transparently logged into the SPs they subsequently use.
- As the authentication method is transparent, it can be repeated whenever appropriate. An online banking SP, for example, can, at any time during a session, request user reauthentication for every sensitive resource requested. This increases the achieved level of security without usability implications.
- As the user's SIM card is used for every (re)authentication, there is a simple single logout mechanism. Once the user leaves, reauthentication will fail and the SP can log the user out without user interaction.
- No sensitive information (such as usernames, passwords or cryptographic keys) is sent over the network. This protects the user's privacy. Furthermore, no risks of information exposure arise at the SP.
- The protocol itself does not impose major computational overheads on any of the involved parties, although there may be a cost associated with the countermeasures against the attacks described in section 4.
- Changes in the deployed GSM infrastructure are minimal, in particular, the SIM does not need to be modified.
- The GSM operator could provide the SPs with user profile information (with the user's consent). This would enable SPs to offer location based services, or automatic form completion for e-commerce transactions.
- As GSM operator fraud management systems use the IMSI to refer to 'suspect' subscriptions, fraud detection can easily be extended to cover SPs.

The following disadvantages arise from use of the proposed scheme.

- A GSM operator can only provide authentication for its subscribers. A SP can therefore only offer the SSO service to subscribers of GSM operators with whom a contractual agreement exists.
- SPs may be charged by GSM operators for authentication and SSO service provision. While this is not a disadvantage for GSM operators, SPs must weigh the cost against the benefits gained from the proposed scheme.

6 Using UMTS/3GPP for SSO

A variant of the SSO protocol is now described, using the Universal Mobile Telecommunications System (UMTS) of the Third Generation Partnership Project (`www.3gpp.org`). Like GSM, UMTS authentication uses a secret key (K_i) shared between a subscriber and the home network operator. The main difference is that the network sends an 'authentication token' (AUTN) as well as the RAND to the subscriber's ME [1]. The AUTN can only be produced by the home network AuC. The variant SSO protocol operates as follows.

1. US → SP: IMSI
2. SP → AuC: IMSI
3. AuC → SP: RAND, AUTN
4. SP → US: RAND, AUTN
5. US → SP: MAC_{IK}(SPID)
6. SP → AuC: MAC_{IK}(SPID), IMSI, RAND
7. AuC → SP: Authentication Assertion or Failure Notification

The US sends its IMSI to the SP (message 1) which forwards it to the AuC (message 2). The AuC generates a random challenge (RAND) and computes the AUTN as a function of RAND and the secret key K_i that corresponds to the given

IMSI. The RAND and AUTN values are sent back to the SP (message 3) and forwarded to the US (message 4).

The US's SIM checks the validity of AUTN and, if valid, generates an 'integrity key' IK as a function of RAND and its secret key K_i, just as it does during normal UMTS authentication [1]. The US uses this key (rather than the encryption key K_c) to compute a MAC on the SP's unique identifier (SPID).

The rest of the protocol is similar to the GSM version. In message 5 the US sends the MAC to the SP which forwards it to the AuC in message 6, while appending the US's IMSI and the RAND of message 3 (message 6 corresponds to a Liberty *authentication request* [13]). The AuC derives the integrity key IK from the RAND and the secret K_i corresponding to the given IMSI. Using it, the AuC computes a MAC on the SPID of the SP from which it received message 6. If it matches the MAC of message 6 then the authentication process is deemed successful. If that it the case, the last message, which corresponds to a Liberty authentication response, contains an assertion; otherwise it contains a failure notification.

As in the GSM version of the protocol (section 3.2), the US must authenticate the SP prior to use of the protocol, and communications between SP and AuC must be mutually authenticated and integrity protected (see section 4.4).

Some of the 'side effects' when using the protocol described in this section, compared to its GSM version described in section 3.2, are the following:

- As the RAND is generated by the AuC instead of the SP, the risk of poor random number generation (see section 4.6) is taken away from individual SPs. Associated computational costs, however, are centralised at the AuC.
- The number of messages has almost doubled. This could lead to increased response times.
- Validation of the AUTN of message 4 by the US provides for AuC authentication, integrity and freshness assurance [1]. As a result, previous (RAND, AUTN) pairs cannot be reused. The process is thus in this case no longer independent of subscriber authentication over the UMTS air interface.

7 Related Work

An overview of SSO architectures and related technologies is provided by De Clercq [3]. Specifications that relate to SSO among disparate security domains include SAML [15] and the Liberty Alliance [11].

Claessens et al. [2] propose a GSM-based authentication method for the World Wide Web. The proposed scheme, however, makes extensive use of SMS messaging. There are several commercial SSO solutions supporting authentication methods involving GSM SMS messaging. Examples include Ubisecure's (`www.ubisecure.com`) Ubilogin™ and Entrust's (`www.entrust.com`) Truepass™ SSO products.

The e-commerce user authentication protocol proposed in [8] relies on the fact that GSM subscribers and network operators share a secret key K_i. The same fact is exploited by the protocol proposed in this paper.

8 Conclusion

We have proposed SSO schemes where a GSM or UMTS operator provides authentication assertions to relying SPs. The schemes can potentially be specified as Liberty profiles, thereby conforming to this open specification.

The protocol yields a seamless user experience as no user interactions are needed, allowing several transparent re-authentications to occur within a user/SP session. This leads to an equally seamless, secure and simple single logout mechanism. The protocol preserves user privacy and mobility. To protect against SIM theft or cloning, the scheme can be complemented by an additional mechanism such as username and password. Impersonation will then only succeed if the attacker has access to both the user's SIM and password.

The required changes to the existing GSM or UMTS infrastructure are minimal. The scheme uses existing SIMs and appropriately equipped MEs. The GSM or UMTS operator's AuC, the

SPs and the US need only support the protocol at a software level. Computational overheads are negligible.

References

[1] C. W. Blanchard. Wireless security. In R. Temple and J. Regnault, editors, *Internet and wireless security*, chapter 8, pages 147–162. IEE, 2002.

[2] Joris Claessens, Bart Preneel, and Joos Vandewalle. Combining World Wide Web and wireless security. *Informatica*, 26(2):123–132, 2002.

[3] Jan De Clercq. Single sign-on architectures. In George I. Davida, Yair Frankel, and Owen Rees, editors, *Infrastructure Security, International Conference, InfraSec 2002 Bristol, UK, October 1-3, 2002, Proceedings*, volume 2437 of *Lecture Notes in Computer Science*, pages 40–58. Springer-Verlag, 2002.

[4] Computer Security Center of the Department of Defense, Meade, Fort George G. *Department of Defense Password Management Guideline*, April 1985. CSC-STD-002-85.

[5] European Telecommunications Standards Institution (ETSI). *Digital cellular telecommunications system (Phase 2+); Security aspects (GSM 02.09 version 8.0.1)*, June 2001.

[6] European Telecommunications Standards Institution (ETSI). *Digital cellular telecommunications system (Phase 2+); Security related network functions (GSM 03.20 version 8.1.0)*, July 2001.

[7] IEEE. *Standard 802.11b-1999/Cor 1-2001(Corrigendum to IEEE Std 802.11b-1999)*, 1999-2001.

[8] V. Khu-smith and C.J. Mitchell. Using GSM to enhance e-commerce security. In *Proceedings of the Second ACM International Workshop on Mobile Commerce (WMC '02)*, pages 75–81, New York, 2002. ACM Press.

[9] Liberty Alliance. *Liberty Architecture Glossary Draft Version 1.2-04*, April 2003.

[10] Liberty Alliance. *Liberty Authentication Context Specification Draft Version 1.2-05*, April 2003.

[11] Liberty Alliance. *Liberty ID-FF Architecture Overview DRAFT Version 1.2-03*, April 2003.

[12] Liberty Alliance. *Liberty ID-FF Bindings and Profiles Specification*, April 2003.

[13] Liberty Alliance. *Liberty ID-FF Protocols and Schema Specification Version 1.2 08*, April 2003.

[14] Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1996.

[15] OASIS, http://www.oasis-open.org/committees/security/. *Security Services Technical Commitee Homepage*.

[16] Eric Rescorla. *SSL and TLS*. Addison-Wesley, Reading, Massachusetts, 2001.

[17] Roger J. Sutton. *Secure Communications: Applications and Management*. John Wiley & Sons, 2002.

[18] Klaus Vedder. GSM: Security, services, and the SIM. In Bart Preneel and Vincent Rijmen, editors, *State of the Art in Applied Cryptography*, volume 1528 of *Lecture Notes in Computer Science*, pages 224–240. Springer-Verlag, 1997.

[19] M. Walker and T. Wright. Security. In F. Hillebrand, editor, *GSM and UMTS: The creation of global mobile communication*, chapter 14, pages 385–406. John Wiley & Sons, 2002.