SEINIT Security for Heterogeneous Mobile Network Services

Jimmy McGibney, Miguel Ponce de Leon, John Ronan,
Telecommunications Software & Systems Group, Waterford Institute of Technology, Cork Road, Waterford, Ireland
Tel: +353 51 302900, Fax: + 353 51 302901,
{jmcgibney, miguelpdl, jronan}@tssg.org

Abstract: This paper presents a model for securing mobile services using heterogeneous access networks, and implementing sample solutions using this framework. This is a project that is defining new security models and policies to address the new issues of the pervasive computing world. The security models and policies are implemented over IPv6 infrastructures to cover various business cases and assessed against real life scenarios. SEINIT is developing a trusted and dependable security framework with the end-user as the focus.

1. Introduction

With applications, such as public information services that allow users to request message (i.e. SMS, email) reminders about hospital services, national car test examinations and driving test examinations, being extended onto the mobile platform, there are new challenges to securing these services. The reliability of operations and reduction in the vulnerability of large and critical infrastructures, such as the communication methods, and information and communication systems pertaining to these infrastructures, is critical. This is especially true considering individuals' freedom and the protection required for their computerised identity and privacy and for the public body providing these services.

In the provision of mobile services, wireless access and fixed wired network technologies will make up the critical infrastructure, and in this environment there will be and are many types of threats to the service information, such as viruses, credit card fraud, wiretapping, infringement of private life, economic espionage, hacking and big brother monitoring, which can be propagated to the user.

It is clear that this mobile digital space will be made up of persistent and volatile digital assets and will often be in an indefinite geographical space.

SEINIT (Security Expert Initiative) is an EU FP6 IST project, developing a trusted and dependable security framework with a suitable, consistent, yet customisable level of trust and security in mobile, heterogeneous networks.

In order to tackle the various heterogeneous entities (Figure 1) that take part in a mobile communication exchange, and to design appropriate security mechanisms for this federated framework, SEINIT sees the data and information related to the mobile service as being organisation independent and centred around the end-user, linked more to the individual, the organisation and the state that own the data, rather than to devices or infrastructures over which it travels.

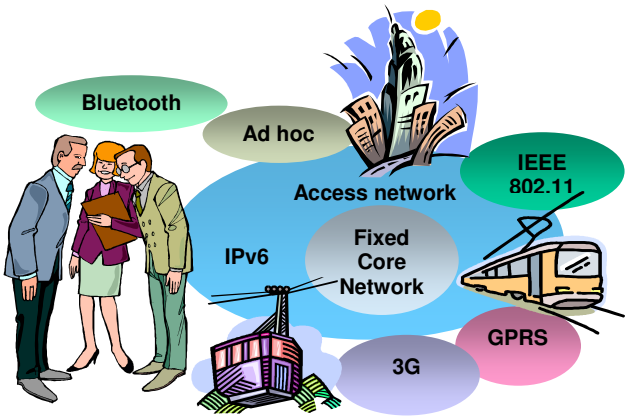

Figure 1: Typical heterogeneous environment[1]

This paper examines the scope of the various mobile service domains and presents an overview of emerging technologies and issues in securing end-to-end mobile services. It also looks at new challenges and threats that are faced by our mobile cyberspace, including the building and maintaining of trust in digital infrastructures for mobile services, which include physical mobility and decentralised mobile infrastructures.

This paper is organised as follows. The following section provides a discussion of threats and vulnerabilities for mobile services. Section 3 then examines security mechanisms and technologies that can help to address these threats. Next, section 4 presents the SEINIT approach, based on the concept of an Infosphere. Finally, section 5 discusses implementation experience and section 6 our conclusions.

2. Threats and Vulnerabilities for mobile services

Computers and communication networks have become an integral part of our daily lives and lie at the basis of our economic, social and institutional environment. Their pervasive use mobile shows how more and more this form of communication is becoming indispensable; however their vulnerability is a major problem[2].

The digital age presents two contradictory aspects:

- On one hand, digital information is vulnerable. This immaterial information can be destroyed, amputated, falsified, confiscated, plagiarised and modified in an infinite number of ways.
- On the other hand, digital information is volatile. This information can be adjusted and personalised.

No week passes without us being reminded of the requirements for information security: a web site has been vandalised, a new virus attack was carried out and has proliferated into, and through networks, or another business has experienced disastrous loss of data.

With the combination of the Internet with mobile communication technologies in the provision of mobile services, the types of threats to digital information grows every day, threats like viruses, spam, illicit content, credit card fraud, wiretapping, infringement of private life, economic espionage, cyber-warfare, hacking and big brother monitoring.

However threats on networks can also be accidental (i.e. human errors, hardware or software breakdowns) or intentional (i.e. attacks by an entity which should not take part in the operation or the exchange, or by an unauthorised entity which takes part in it).

Threats also appear in the form of intrusions, unauthorised accesses, theft of identity or information, denial of services, attacks on mails, attacks on network protocols, operating systems and the security devices. These threats can be carried out with the complicity or the

naivety of users, but often these attacks are undetectable and leave the victim in ignorance. In relation to the provision of mobile services these attacks can aim at a determined target with a precise objective and, as a side effect, impede, or otherwise hinder a broad spectrum of potential human targets.

Cyber attacks are expected to increase in the future and this will of course, affect mobile services:

- Cellular communications are protected by their centralised structures in the form of administration by a telecom operator, and by the notion of virtual circuit that is still present in GSM technology. The tendency towards "computerised" telecommunications (voice over IP, GPRS, UMTS) will break the trust in this sector.
- The issue of security in a mobile universe within the ubiquitous presence of information technology is a great challenge. How do you impart trust to a digital world? What kind of digital governance should be enforced to restore users' confidence?

Alongside the technological development, new threats will emerge. Some of these can already be predicted:

- Threats to private life communications, following the advent of Voice over IP (VoIP) replacing the traditional telephone lines. The telephone will operate as successions of information packets as in a computing network, and consequently, telephone calls will become susceptible to the same attacks as the traditional computer attacks (stolen identity, caller's anonymity, etc.).
- Smart labels will soon replace national ID cards and passports. While this improvement will "eliminate" the tedious job of an administrator, it will also make it possible to tamper with the smart labels and to physically track individuals. With huge implications for personal privacy.

GPRS and UMTS are significant improvements of daily life and especially of the regular daily routine of mobile workers (those that are required to travel). It will simply be necessary to be careful and avoid disagreeable surprises while seeking differentiation and diversity in order to prevent or restrict the effects of individual errors.

Decentralised mobile architectures include wireless connected peers using ad-hoc networks. Threats to these systems are typically divided into passive and active classes. These two broad classes are then subdivided into other types of threats.

2.1 Passive Attack

An attack in which an unauthorised party gains access to an asset and does not modify its content (i.e., eavesdropping). Passive attacks can be either eavesdropping or traffic analysis (sometimes called traffic flow analysis). These two passive attacks are described below:

- Eavesdropping: The attacker monitors transmissions for message content. An example of this attack is a person listening into the transmissions on a LAN between two workstations or tuning into transmissions between a wireless handset and a base station.
- Traffic analysis: The attacker, in a more subtle way, gains intelligence by monitoring the transmissions for patterns of communication. A considerable amount of information is contained in the flow of messages between communicating parties.

2.2 Active Attack

An attack whereby an unauthorised party makes modifications to a message, data stream, or file. It is possible to detect this type of attack but it may not be preventable. Active attacks may take the form of one of four types (or combination thereof): masquerading, replay, message modification, and denial-of-service (DoS). These attacks are defined below:

- Masquerading: The attacker impersonates an authorised user and thereby gains certain unauthorised privileges.
- Replay: The attacker monitors transmissions (passive attack) and retransmits messages as the legitimate user.
- Message modification: The attacker alters a legitimate message by deleting, adding to, changing, or reordering it.
- Denial of service: The attacker prevents or prohibits the normal use or management of communications facilities.

The most widely used mobile and wireless networks, namely GSM/GPRS/UMTS and wireless local area networks (WLANs) based on the IEEE 802.11 series of standards, are subject to both passive and active attacks in some or all of the categories listed above.

2.3 GSM

GSM security is based having a secret key stored on the user's SIM card and another copy of it in the network's Authentication Centre. When the user attempts to connect, the network issues a random number to the user. To be correctly authenticated, the SIM must correctly encrypt that number with the secret key and return it to the network. This also provides the basis for the SIM and the network to agree temporary encryption keys that protect the confidentiality of communications.

Thus GSM provide encryption as well as authentication of the user by the network. Several risks remain, however:

- There is no authentication of the network by the user – i.e. someone could set up a bogus network base station and accept user connections.
- The generated temporary encryption keys used in GSM are too short and the encryption scheme can be broken in a short time (as short as one minute, depending on processing power).
- The operator is free to choose authentication and encryption algorithms. Some widely used algorithms (especially COMP-128 have well-documented weaknesses.
- Encryption is just done on the radio interface – i.e. between the mobile device and the base station, and not in the rest of the network.
- There is no built-in protection against denial of service attacks.

2.4 UMTS

UMTS provides some enhancements. Now, we have mutual authentication – i.e. both the user and the network authenticate each other. Also, longer encryption keys are used and algorithms are made publicly available for maximum scrutiny and testing.

The risks associated with IEEE 802.11 WLANs [3] are similar in some ways. Again, there are risks of loss of proprietary information, legal and recovery costs, tarnished image, and loss of network service.

WLANs risk loss of confidentiality following an active attack as well. Sniffing software can obtain user names and passwords (as well as any other data traversing the network) as they are sent over a wireless connection. An adversary may be able to masquerade as a legitimate user and gain access to the wired network from an access point. The malicious eavesdropper then uses the user name, password, and IP address information to gain access to network resources and sensitive corporate data.

Data integrity issues in wireless networks are similar to those in wired networks. Because organisations frequently implement wireless and wired communications without adequate authentication, in conjunction with cryptographic protection of data, integrity can be difficult to achieve. An attacker, for example, can compromise data integrity by deleting or modifying the data in an e-mail from an account on the wireless system. This can be detrimental to an organisation if important e-mail is widely distributed among e-mail recipients. Message modification attacks are possible when cryptographic checking mechanisms such as message authentication codes and hashes are not used.

A denial of network availability involves some form of DoS attack, such as jamming. Jamming occurs when a malicious user deliberately emanates a signal from a wireless device in order to overwhelm legitimate wireless signals. Non-malicious users can also cause a DoS. A user, for instance, may unintentionally monopolise a wireless signal by downloading large files, effectively denying other users access to the network. As a result, agency security policies should limit the types and amounts of data that users are able to download on wireless networks. More details on DoS attacks for WLANs is provided in [4].

3. Emerging Technologies for securing mobile services

In implementing the components and models for a trusted and dependable security framework, the SEINIT project has identified, investigated and selected emerging security technologies for different wired and wireless networks that are threatened and vulnerable when enabling mobile services.

In the area of Mobile IPv6, and in particular in the mutual authentication of the Binding Updates exchange between the Mobile Node (MN) and any Corresponding Node (CN), the mechanism of Return Routability is being used.

This increases the time for the binding registration by roughly the round trip time between MN and CN, and prevents against the majority of know attack scenarios.

With IPv6 privacy extensions for address auto configuration [5], there is a mechanism to mitigate privacy concerns, which might arise from the use of static IPv6 address derived from IEEE identifiers such as MAC addresses.

Cryptographically Generated Addresses (CGAs) allow for a secure association of an IPv6 address, the CGA, with a public key. While this kind of association is mainly done using certificates, and therefore requires the deployment of Public Key Infrastructures (PKIs), the CGA approach does not require any further infrastructure.

Extensible Authentication Protocol (EAP) is an authentication framework which supports multiple authentication methods. EAP typically runs directly over data link layers such as PPP or IEEE 802, without requiring IP. EAP is used to select a specific authentication mechanism and permits the use of a backend authentication server, which may implement some or all authentication methods.

In comparison the Protocol for carrying Authentication for Network Access (PANA) is used to provide a link layer agnostic transport mechanism for carrying EAP based network authentication information. This has been achieved by running PANA on top of UDP/IP.

Not long after its development, Wired Equivalent Privacy's (WEP) cryptographic weakness began to be exposed, and breaches in WLAN security were exposed. WEP does provide some margin of security compared with no security at all and remains useful for casual use in deflecting would-be eavesdroppers. For large enterprise users, WEP's native security can be strengthened by deploying other security technologies such as VPN or 802.1x authentication with dynamic WEP keys.

As an effect of WEP shortcomings, WiFi Protected Access (WPA) has emerged as a wireless network security technology that improves the authentication and encryption features of WEP. WPA has:

- Enhanced Data Encryption through Temporal Key Integrity (TKIP)
- Enterprise-level User Authentication via 802.1x and EAP

802.11i is the name of the IEEE Task group dedicated to standardising WLAN security. The 802.11i security framework is based on Robust Security Mechanisms (RSN). RSN has two parts:

- The Security Association: looks at RSN negotiation procedures, IEEE 802.1x authentication and IEEE 802.1x key management
- Data Privacy and Management: looks at TKIP, which is a software patch to WEP to provide a minimal adequate level of data privacy and Advanced Encryption Standard (AES) which is a more robust data privacy scheme.

4. SEINIT Infosphere

In order to mitigate the risks of threats and vulnerabilities to mobile Services, the SEINIT security paradigm is oriented towards defining new security models, more fitting to the reality of information systems, specifying new security policies that are more effective, adaptable to the surrounding ambience and implemented on the new digital infrastructure systems.

Before describing the SEINIT Infosphere concept we must first look to the idea of security equalling privacy. Digital privacy is defined in [6] as a set of security functions (anonymity, pseudonymity, unlinkability, unobservability) used to counteract often-intentional threats consisting of intrusion into the intimate secrets of individuals. The notion of computer privacy has been defined in [6] as having the following properties:

- Anonymity – guaranteeing that a user can use a certain resource or service without revealing his identity.
- Pseudonymity, assumption of aliases – guaranteeing that a user can use a certain resource or service without revealing his identity but remaining responsible for his actions.
- Unlinkability – representing the impossibility of other users to determine a connection between the different operations performed by a single user.
- Unobservability – guaranteeing that a user can use a resource or a service without other users being able to determine whether a certain resource or service is being used.

4.1 The three Infospheres

SEINIT has defined an infosphere as a digital space made up of a persistent and a volatile asset in an often indefinite geographical space, which is linked more to the individual, the organisation and the state rather than to devices or infrastructures.

SEINIT maps all information concerning an entity to an Infosphere. The information can reside in a Security Domain controlled by the Infosphere or in another domain. Either way the Infosphere Supervisor will have influence over the information. From this initial

work the project has identified three subcategories of an infosphere. As illustrated in Figure 2, an infosphere may span several security domains. For example, a user's personal infosphere might be defined as all the sensitive personal information that exists on that person. Some of that information will be under the user's direct control (on own devices) while other parts will be elsewhere, such as in the databases of his/her local hospital, tax office or university.

The individual infosphere (security of a Personal Area Network, a distributed terminal, computers within a user's reach or field of vision) is a domain where security is an essential factor, yet relatively unexplored except as an extension of network security. Its features include lightweight methods for identification, authentication, protection, and management. The goal of this modelling is to examine more generally and abstractly security in a personal environment.

In the open collective infosphere (MAN, WLAN, and LAN security, network security in an enterprise, a campus, or a public space), security has been inherited from the Internet and from LANs (client-server security). This sub category explores security in an anonymous collective environment and more generally analyses the security of heterogeneous networks. End-to-end security through heterogeneous access networks (the Internet, UTRAN, WLAN, Bluetooth, etc.) remains an open problem, especially when we include nomadic users (generalised mobility) and the mobility of network nodes that are moveable (via reconfiguration) or mobile (via the handover of applications and their status).

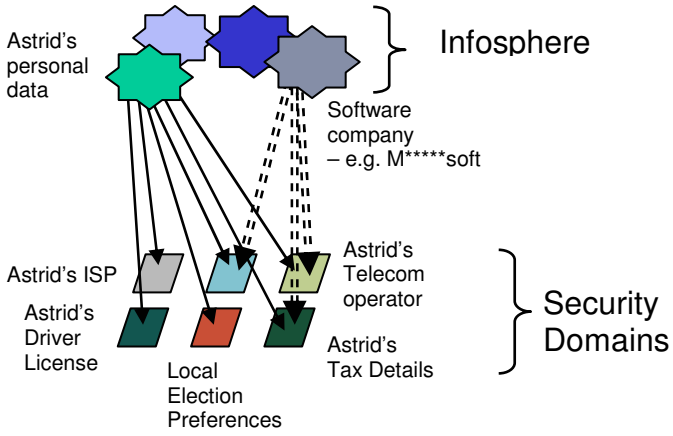

Figure 2: How infospheres map onto security domains [1]

The undefined infosphere (security of WANs, reduction of vulnerabilities in critical infrastructures) has been poorly understood because it is a very new field (the effect of cascades and avalanches in cyber attacks, etc.). This subcategory analyses security in an open, large-scale environment that handles very large systems (the electrical power grid, the telephone network or the information network for a region or a country). This work consists of designing and simulating security for large, critical infrastructures in the telecommunications field in order to warn of faults, recover after faults (scarring), to predict problems, and to avoid the proliferation of faults.

5. Implementation Experience

In order to perform any meaningful critique of the ideas proposed, a set of User scenarios were defined([7]§5). The project then matched the scenarios with technology in order to best meet the needs of the scenarios ([7]§6), all the while adhering to the "infosphere"

concept. The project then sent about implementing these scenarios using a combination of available technological components[8] coupled with the SEINIT Security Virtual Machine(SSVM).

5.1 The SEINIT Security Virtual Machine

The SEINIT SVM brings together the SEINIT high-level virtual concepts and demonstrates how these may be implemented in a concrete way, enhancing the security of both new and legacy applications, by gluing together different security technologies.

The goal of the SSVM is to give the ability to use a set of high-level security tools, to make them adaptable to heterogeneous security domains, and use the trust management functionalities, which will be provided by the SEINIT infrastructure. Whether SEINIT-enabled or not, any application will benefit from the various functionalities provided by the middleware.

The main part of the SEINIT SSVM has been developed in three main building blocks: Information, Decision, and Action:

- The Information module gathers and stores all information regarding the current context within which the local device is currently operating, the local device's configuration, and trust constraints that may be placed upon communications with remote devices. This module also stores the security policies that may be defined in terms of high and low levels.
- All information that is gathered and stored within the Information module is processed within the Decision module, in order to generate and establish the necessary security policies ready for enforcement.
- Once policies are ready for enforcement, they are pushed to the Policy Enforcement Point within the Action module. The Action module is responsible for loading the appropriate components (e.g. IPsec sub-system for VPN or encryption algorithms) in order to enforce the relevant policies within the relevant device sub-systems.

The SSVM has been realized in JAVA code and therefore can be installed on different platforms. Its core relies on management classes dedicated to information communication throughout the whole chain of processes.

The following scenario is one that was implemented in order to test the infosphere concept as well as user scenario.

1. Bootstrap – A user switches on his/her laptop computer (device) and the SSVM must listen for announcement message pointing to where the initial security policy of the current domain can be retrieved from. This is the SDIP (Service Discovery and Initial Policy) Module.
2. Initial Policy Request – The SDIP requests the initial security policy from the Policy Information Point (PIP)
3. PIP server Response – The PIP server responds to the SDIP with the required information.
4. Policy Analysis – The SDIP analyses the policy information received in step 3 what technology and/or techniques (if any) should be used to authenticate the users computer. Using this information the SDIP may request a Technology Abstraction Layer (TechnoWrapper) to utilise the required technology with the correct options.

5. Network access – If the policy returned in step 3 requires it. Then the TechnoWrapper module will utilise an authentication component to authenticate the users computer to the network.
6. Authentication and Authorisation – The AAA server validates the users credentials and instructs the Access Router to permit traffic to pass for the users' computer. The presence of an authenticated user is also communicated to the SSVM.
7. User uses the network connection – Once authenticated, the user can now use the network connection.

The SSVM continuously monitors the threat level using trus management (IDS/Honepot) components and communicates any changes in this threat level back to the Middleware running on the user's computer. This in turn allows for the SSVM to adjust the perceived threat level and hence instruct the TechnoWrappers to increase or decrease the level of security (encryption) depending on the threat level.

The main advantage of this closed-loop feedback is that in the event of malevolent activity being detected within the vicinity of the user, the middleware is informed, and depending on the users personal preferences, the security level can automatically be re-negotiated, on-the-fly, in order to encrypt all traffic to and from the network access layer (Wireless Access Point). This can also be triggered on user request, or by a SEINIT aware application.

These steps highlighted above hide a large amount of complexity in this domain. At each step, we have adopted industry *best practice*[9] for security and deployed the relevant technology. This includes PANA & DIAMETER for authentication and authorisation, IPsec for network layer encryption and DNS/DNSSEC for an authenticated source of Public Key Certificates for the establishment of secure channels and for storing information on how to locate Security Level Agreements.

5.2 Open Research Issues

While continuing to refine our current implementation for the duration of the project we also are looking at two specific areas that we feel have are not adequately addressed.

Secure Bootstrapping

Secure bootstrapping is required each time users power up (or connect) their devices in an unknown, and perhaps untrusted environment. The term secure bootstrapping describes the ability of a device to discover step by steps its local policies, the environment together with the offered services, and the enforced policies in this domain. The policies of the domain to which the device wants to connect to must be matched against the local policies of the device and its user. If the policies are compatible, the device can make use the offered services to gain secure access to available resources (e.g. file servers, databases, or simply Internet connectivity).

A slightly different kind of bootstrapping is needed when the device moves from one point of attachment of a security domain to another. In the new security domain, it has to again locate the available services and local domain policy required to access these services. Once the local device / user policy has been compared to the domain policy and (hopefully) has been found to match, the device has to initiate a new security negotiation, i.e. it has to authenticate again at the local domain and built up again its secure connections to other domains to continue its work.

Tightly bound to secure bootstrapping is the necessity for booting devices or users just newly arriving at a security domain to include a service and policy discovery mechanism.

This means, for service discovery that the devices have to automatically discover the availability of SEINIT services. The requirements for policy discovery would then be that the device needs to automatically detect the appropriate policy server, download the policy description of the security domain, match it against its local policy settings and to select then the appropriate information, e.g. how they is required to gain access and whether they have the appropriate local services / credentials available.

Enhanced Mobility Support

To support the mobility of mobile devices and users roaming between security domains, a context transfer method could be used. This method would enable to mobile device to transfer security context information (e.g. that it has already successfully authenticated at a security domain and its established secure connections to other security domains) from the current security domain to the next security domain it plans to attach to. If the security domains have trust established with each other and are able to transfer security context information between them, the mobile device would just need to prove that it is the same device which has already authenticated at the old security domain.

This authentication would only take place locally in the new security domain and not require any authentication lookups at other domains. In parallel, the new security domain could already initiate the transfer of security associations. This would greatly reduce the time required by the mobile device to move from one security to another and re-establishing its communication.

Several techniques exist to integrate Mobile IP(v6) within the SEINIT framework, each of them having specific advantages and disadvantages. Which of them is the most suitable, could be investigated in the further work of SEINIT.

6. Conclusions

In summary, the mobile and wireless networks are subjected to significant threats, not least by the increased ease of eavesdropping by outsiders as well as the risk of attackers setting up bogus base stations or operating with bogus terminal equipment. The use of public radio spectrum also increases the opportunities for denial of service attacks. A wide variety of wireless technologies and services exist, and various security approaches are taken, some of which have significant flaws. This paper outlined the SEINIT approach to unifying these approaches and simplifying the provision of security in complex heterogeneous environments.

The principal value of SEINIT is in the seamless integration of security into an environment where varied and diverse services are delivered to users on a disappearing communication infrastructure. The idea of a disappearing communication infrastructure is that its features become increasingly transparent to users, service providers and services.

The designs of future ambient systems – that is, IT systems intimately integrated with everyday environments – will have to be based on radically new architectures comprising of an unbounded set of "building blocks", where these blocks may be embedded in everyday objects, be it stand-alone objects or software entities. A secure service framework is a crucial part of this.

References

[1] SEINIT Consortium, D1.1, (2004), “Trust, Security and Policy Framework Specification”, SEINIT Deliverable 1.1, available at http://www.seinit.org

[2] SEINIT Consortium, D1.2, (2004), “Assessment of threats & vulnerabilities in networks”, SEINIT Deliverable 1.2, available at http://www.seinit.org
[3] IEEE/ANSI Standard 802.11 (also ISO/IEC 8802-11), (1999), Wireless LAN Medium Access Control and Physical Layer Specification.
[4] Bellardo, J. & Savage, S., (2003), 802.11 Denial of Service Attacks: Real Vulnerabilities and Practical Solutions, 12th USENIX Security Symposium
[5] Narten, T. & Draves, (2001), R., Privacy Extensions for Stateless Address Autoconfiguration in IPv6, RFC 3041, Internet Engineering Task Force
[6] ISO/IEC 15408, (1999), Information technology - Security techniques - Evaluation criteria for IT security, ISO/IEC
[7] SEINIT Consortium, D6.1, (2005), “Description of trial scenarios and security test-bed”, SEINIT Deliverable 6.1, available at http://www.seinit.org
[8] SEINIT Consortium, D4.2, (2004), “Early Integrated platform components & tools”, SEINIT Deliverable 4.2, available at http://www.seinit.org
[9] SEINIT Consortium, D2.4, (2005), “Security best practice manual”, SEINIT Deliverable 2.4, available at http://www.seinit.org